

The Evolution of the Moral Sentiments and the Metaphysics of Morals

Fritz Allhoff

Accepted: 26 October 2008 / Published online: 2 December 2008
© Springer Science + Business Media B.V. 2008

Abstract So-called evolutionary error theorists, such as Michael Ruse and Richard Joyce, have argued that naturalistic accounts of the moral sentiments lead us to adopt an error theory approach to morality. Roughly, the argument is that an appreciation of the etiology of those sentiments undermines any reason to think that they track moral truth and, furthermore, undermines any reason to think that moral truth actually exists. I argue that this approach offers us a false dichotomy between error theory and some form of moral realism. While accepting the presuppositions of the evolutionary error theorist, I argue that contract-based approaches to morality can be sensitive to those presuppositions while still vindicating morality. Invoking Stephen Darwall's distinction between contractualism and contractarianism, I go on to offer an evolutionary-based contractarianism.

Keywords Evolutionary ethics · Error theory · Contractualism/contractarianism

1 Introduction

Let us suppose that natural selection, coupled with epigenetic rules, has endowed us with various moral sentiments. We could debate about what these moral sentiments are, but plausible suggestions might include approbations toward justice, fairness, reciprocity, and

Parts of this paper were presented at the 2005 meeting of the International Studies for the History, Philosophy, and Social Studies of Biology (Guelph, Canada); I thank those in attendance, especially Kim Sterelny and Ben Fraser, for helpful discussion. I also thank Matthew Hanser, Richard Joyce, Robert Richards, Michael Ruse, and two anonymous reviewers for their feedback. Some of the writing of this paper was done as a Research Fellow at The Australian National University's Centre for Applied Philosophy and Public Ethics; I thank ANU and CAPPE for their support.

F. Allhoff (✉)

Department of Philosophy, Western Michigan University, Kalamazoo, MI, USA
e-mail: fritz.allhoff@wmich.edu

altruism as well as disapprobations toward their contraries (perhaps as regulated through guilt, shame, revenge, etc.).^{1,2} Similarly, we might also debate whether all humans have these moral sentiments, whether the strength of these moral sentiments varies, whether culture can affect these moral sentiments, etc. To these challenges, I propose that we postulate a central group of moral sentiments that are roughly invariant across populations, and that these sentiments are psychologically efficacious, to some degree or another, in all of their possessors.³

Also, we might question whether evolution by natural selection *directly* endowed humans with these moral sentiments, or whether they derive from some other proximate (or ultimate) mechanism. For example, maybe moral sentiments are wholly the results of cultural evolution. Or maybe they are “spandrels” (i.e., by-products) of something else (e.g., the “big brain”) that was directly selected for by evolution (Gould and Lewontin 1978). These are, of course, empirical possibilities, though they are not the explanations that I favor. I concede that there could be disagreement as to *why* the moral sentiments exist, though I maintain that evolution by natural selection provides the most plausible account. Regardless, for the purpose of this paper, I will *assume* that our moral sentiments were *directly* selected for through the process of evolution by natural selection.

If this is right, what moral mileage might we be able to obtain? In other words, how would a naturalistic account of the moral sentiments affect the field of ethics? Before establishing my positive account, let us start with the account that I deny, namely

¹ There are other important issues that pertain to the characterization of these sentiments. For example, we could ask whether they are activated by judgments or mere perceptions, whether they attach to act-types or act-tokens, etc. While my general account should be compatible with any answers to these questions, I view sentiments as lacking conceptual content and being responsive to act-tokens; these sympathies derive from my view that the sentiments evolved by natural selection. First, I think that any account which ascribes conceptual content to the moral sentiments is problematic insofar as it would fail to appreciate the gradual emergence of these sentiments across evolutionary time and the role that such sentiments play in proto-moralities (e.g., those of higher primates) in which cognitive judgment is clearly lacking. In other words, such an account endorses a conception of moral sentiments which is clearly discontinuous with their evolution; this evolution is more plausibly linked to affective (i.e., non-cognitive) states than to conscious judgments.

Second, I think that the sentiments are responsive to act-tokens rather than act-types because an individual directly interacts with act-types rather than abstract act-tokens. This latter category is “invisible” to selection insofar as agents’ responses to act-types would be screened-off by their responses to act-tokens (i.e., natural selection would not “gain” any information by being provided with an agent’s responses to act-types over and above the information being provided by his response to act-tokens), and this screening-off means that responses to act-types would be (proximately) irrelevant to selection pressures. Given these considerations, I propose that sentiments attach to act-tokens rather than act-types. There are, to be sure, relevant *features* of act-types that trigger affective responses, and these features could be instantiated in multiple act-types. All that I wish to claim, however, is that it is the tokens, not the abstraction to the types, activate the sentiments.

² For non-technical background in this regard, see Marc Hauser, *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong* (New York: Ecco 2006).

³ There is, of course, anthropological data for these claims; see, for example, Donald E. Brown, *Human Universals* (Boston, MA: McGraw-Hill, 1991). See also Richard Joyce, *The Evolution of Morality* (Cambridge: MIT Press, 2006), pp. 134–7.

“evolutionary error theory.”⁴ For example, consider Michael Ruse, whose view is archetypal: Ruse holds that any evolutionary basis for the moral sentiments leads to moral nihilism. Why? Imagine that it did not. Imagine, for example, that evolution led us to judge that ϕ is morally permissible. And further imagine that ϕ is, in fact, morally permissible in virtue of according with some sort of moral fact (e.g., of the sort the moral realist would propose).⁵ Wouldn't this be an *incredible* coincidence? How could evolution *possibly* endow us with moral sentiments that track some external moral reality? How would evolution gain access to this external moral reality and give us the moral sentiments to track it? This does not seem plausible. And, if we think of *all* the moral sentiments we have (or even some restricted group), it is *very* unlikely that evolution and morality would just happen to line up in every case.

This argument, of course, is not deductive. It is (at least logically) possible that evolution has given us moral sentiments that coincide with some independent moral realm. Or, in other words, moral nihilism does not deductively follow from a naturalistic account of the moral sentiments. Rather, this is an inference based on simplicity. To wit, Ruse thinks that there is *no reason* to postulate moral truths given that we can explain all the salient features of our moral lives without appealing to them since they are metaphysically unnecessary and otherwise unmotivated. Therefore, he concludes, we should adopt an error theory approach to morality, by which we make moral claims that are all false because they assign moral predicates that are never instantiated. I think that this is wrong and that the skeptical conclusion is unwarranted.

To see why, let us look at his reasoning. Ruse seems to think that there are only two options: full-blown metaphysical moral realism and moral nihilism. When Ruse says that it is unlikely that our moral sentiments could track an external moral reality, he has in mind the sorts of moral metaphysics posited by either Plato or G. E. Moore. On either of these views, goodness exists *independently* of the moral agent, and moral goodness is something that is simply woven into the very nature of the universe. So, for ϕ to be permissible, the universe must exist in a certain way, and Ruse thinks that it is highly improbable that evolution could access/track the moral constitution of the universe. And I agree with Ruse

⁴ Michael Ruse has written extensively on this subject. He first argued for his view in Ruse (1984) and “Evolution and Morality,” *Philosophic Exchange* 16 (1985): 5–26, but his most substantial theoretical formulations can be found in a paper and book from the next year; see Ruse (1986a) and *Taking Darwin Seriously: A Naturalistic Approach to Philosophy* (Oxford: Blackwell, 1986b). He has continued to advance his program in numerous writings: *The Darwinian Paradigm: Essays on Its History, Philosophy and Religious Implications* (London: Routledge, 1989); “The New Evolutionary Ethics” in *Evolutionary Ethics*, ed. Matthew Nitecki and Doris Nitecki (Albany, NY: State University of New York Press, 1993), pp. 133–62; Ruse (1999a); and *Monad to Man: The Concept of Progress in Evolutionary Biology* (Cambridge, MA: Harvard University Press, 1999b). Ruse has also co-authored two papers with E. O. Wilson, see “The Evolution of Ethics,” *New Scientist* 108.1478 (1985): 50–2 and “Moral Philosophy as Applied Science,” *Philosophy* 61 (1986): 173–92. As an interesting note, Ruse initially argued against the relevance of evolution for ethics and then changed his mind—for the original view, see *The Darwinian Revolution: Science Red in Tooth and Claw*, 2nd ed. (Chicago: University of Chicago Press, [1979] 1999c).

Richard Joyce (2000) has advanced an evolutionary error theory; see “Darwinian Ethics and Error,” *Biology and Philosophy*, 15, 713–32, *The Myth of Morality* (Cambridge: Cambridge University Press, 2001), and *The Evolution of Morality* (Cambridge: MIT Press, 2006). Joyce advances a motivational skepticism wherein he argues that morality cannot maintain various of its “non-negotiable” features, such as the provision of reasons for action to the amoralist; Joyce then presents a moral fictionalism as an alternative to morality.

⁵ For present purposes, the precise details of such an account are not important. For example, though, see Russ Shafer-Landau, *Moral Realism: A Defense* (Oxford: Oxford University Press, 2003).

that naturalistic accounts of the moral sentiments are not very congenial to Platonic or Moorean conceptions of the good. But, there are conceptions of the good other than the Platonic or Moorean ones, and I argue that at least some of these conceptions are unchallenged by naturalistic accounts of the origin of the moral sentiments. In particular, I will develop a contract-based account that can accommodate these naturalistic sympathies while nevertheless being insulated from the arguments of the error theorist.

2 Subjectivism and Constructivism

Before developing that contract-based account, let me acknowledge a wide class of positions vis-à-vis moral metaphysics that escape the error theorists' attacks. I think that most of these positions fail for independent reasons, but it is worth noting this class of positions, from which I will subsequently develop my preferred one. The challenge, then, is to accept the evolution of the moral sentiments yet prevent the slide to moral nihilism. I agree with the evolutionary error theorist that this eliminates moral realism as a candidate position. But, if moral facts could be understood as more "metaphysically slender" (Blackburn 1998, 312). (i.e., without the metaphysical commitments of moral realism), then these criticisms can be avoided.⁶

There are various ways to have moral facts without being a moral realist, and these positions are sometimes grouped under the headings of "subjectivism" or "constructivism". I take it that the latter term is more en vogue and represents more sophisticated views but, regardless, all of these positions share two key features. First, they allow that moral claims are truth-evaluable. And, second, they allow that some moral claims are, in fact, true. The truth conditions are, of course, different than they would have been for moral realism, but some moral claims come out being true nonetheless. This second feature distinguishes them

⁶ It is worth acknowledging that all of the meta-ethical positions I will consider are cognitivist (i.e., they take moral claims to be truth-evaluable), and that non-cognitivism seems like a profitable avenue for someone who thinks that natural selection shaped the moral sentiments. If, as the non-cognitivist holds, moral discourse is constituted by the expression of attitudes rather than reports of beliefs, then this could align well with a picture that holds we have evolved to feel (dis)approbation toward various acts or states of affairs since these expressions would be directly tied to evolved sentiment; a cognitivist picture would have to postulate some further step by which the sentiments were converted into judgments. In fact, evolutionary considerations have been invoked by Allan Gibbard to defend his norm-expressivism, and these considerations could easily be co-opted by other non-cognitivist projects, such as A.J. Ayers's emotivism or Simon Blackburn's quasi-realism. See Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, MA: Harvard University Press, 1990); A.J. Ayer, *Language, Truth, and Logic*, 2nd ed. (New York: Dover, [1936] 1946); and Simon Blackburn, *Spreading the Word* (Oxford: Oxford University Press 1984), *Essays in Quasi-Realism* (Oxford: Oxford University Press, 1993), and (1998).

While the cognitivism/non-cognitivism debate falls outside the scope of this paper, my sympathies lie with cognitivism: it seems to me that, when we make moral claims, they are either true or false. To be sure, some of these sophisticated views, like Gibbard's and Blackburn's, nevertheless want to accommodate this apparent feature of moral discourse, though I think that the consensus on these attempts is equivocal at best. Terry Horgan and Mark Timmons have defended another sophisticated position that aspires to be both cognitivist and expressivist; this would be a non-starter on traditional taxonomies. See Terry Horgan and Mark Timmons, "Cognitivist Expressivism", in Terry Horgan and Mark Timmons (eds.), *Metaethics after Moore* (Oxford: Oxford University Press, 2006), pp. 255–298.

At any rate, I shall not consider non-cognitivism in this paper: my goal is to show that contract-based approaches offer a strong (cognitivist) response to the evolutionary error theorist. And, furthermore, this response is offered on terms that the evolutionary error theorist would be willing to accept vis-à-vis the truth evaluability of moral claims.

from error theory; error theory allows that moral claims are truth-evaluable, just denies that any of them is true.⁷

Consider the following sorts of views, which I take to be examples in this regard:

1. Simple Subjectivism: “ ϕ is wrong” =_{def} “I disapprove of ϕ ”
2. Cultural Relativism: “ ϕ is wrong” =_{def} “We disapprove of ϕ ”
3. Divine Command Ethics: “ ϕ is wrong” =_{def} “God disapproves of ϕ ”⁸

None of these analyses of moral claims carries the metaphysical commitments associated with moral realism, since their only commitments are to metaphysically innocuous concepts like approval. Alternatively, consider:

4. Moral Realism: “ ϕ is wrong” =_{def} “There exists some mind-independent moral fact such that ϕ is wrong”

These former positions are immune to the sorts of arguments that the evolutionary error theorist offers since those arguments are only effective against some sort of robust metaphysics (e.g., of the sort that moral realism would offer). Therefore, we have already found a way to get moral facts without risking the error theorists’ critique. Nevertheless, there are various well-rehearsed problems with holding that moral claims are to be analyzed in the ways suggested by the aforementioned meta-ethical views: that they misunderstand the “aboutness” of moral claims;⁹ that they preclude the possibility of moral disagreement;¹⁰ that they allow for moral infallibility,¹¹ and so on.¹²

⁷ For more discussion of these various categories, see Geoffrey Sayre-McCord, “The Many Moral Realisms”, *Southern Journal of Philosophy* 24.supplement (1986): 1–22. Reprinted in Geoffrey-Sayre McCord (ed.), *Essays in Moral Realism* (Ithaca, NY: Cornell University Press, 1988), pp. 1–26. See also R.M. Hare, *Sorting out Ethics*, New. Ed. (Oxford: Oxford University Press, 2000), esp. ch. 3.

⁸ This analysis would only be appropriate for one horn of Euthyphro’s dilemma (i.e., “ ϕ is good because it is loved by the gods); the other (i.e., “the gods love ϕ because it is good) would yield moral realism. See Plato’s “Euthyphro” in *Five Dialogues*, trans. G.M.A. Grube (Indianapolis, IN: Hackett Publishing Co., 1981), 10a1–11b8.

⁹ If the claim “ ϕ is wrong” were to be properly analyzed as “I/we/God disapproves of ϕ ”, the subject of the analysandum would have to be the same as the subject of the analysans. But, in these cases, the subjects are not the same: the subject of the analysandum is ϕ and the subject of the analysans is I, we, or God. Therefore, the analysans does not provide the correct analysis of the analysandum.

¹⁰ This argument goes back to Moore; see his *Ethics* (London: William and Norgate, 1912), ch. 3. Consider, for example, simple subjectivism. Imagine that two individuals, A and B, are debating the moral permissibility of ϕ ’ing. A says “ ϕ ’ing is wrong” which would mean “I, A, disapprove of ϕ ’ing.” And B says “ ϕ ’ing is permissible” which would mean “I, B, approve of ϕ ’ing.” But now notice what has happened: the disagreement has disappeared. Note that, with the original claims, A and B *disagreed* with each other; A did *not* agree that ϕ ’ing is permissible, nor did B agree that ϕ ’ing is impermissible. But the analysis dissolved the moral disagreement since A would *agree* that B approves of ϕ ’ing, and B would *agree* that A disapproves of ϕ ’ing. So, since simple subjectivism interprets moral claims as reports of an individual’s moral attitudes, moral disagreement would be impossible.

¹¹ For example, if “The Holocaust was morally permissible” means “I approve of the Holocaust”, it would be *true* (so long as I am speaking sincerely). In fact, I could *never* utter a false moral claim so long as I was speaking sincerely. These same points could be made regarding any subjectivistic theory, including cultural relativism or divine command ethics (though the problem of infallibility might be less of a concern for the latter). While infallibility is likely a problem on its own, it also precludes the possibility of moral progress insofar as, were nobody to ever be wrong about their moral claims, those claims could never come closer to moral truth since, *ex hypothesi*, they were already veridical.

¹² Many of these are so simple that we cover them in introductory ethics classes. See, for example, James Rachels, *The Elements of Moral Philosophy*, 3rd ed. (Boston: McGraw-Hill, 1999), esp. ch. 3.

My point in mentioning these above views is not to defend them, but rather to show that, structurally, there are straightforward ways to block the error theorists' argument: all of these views give us moral facts without complicated metaphysics. While I have only pointed to the most naïve of these accounts, this point should be uncontroversial. Certainly there are independent reasons to reject these naïve versions, and, I think, probably even their more sophisticated variants. But, for present purposes, I merely mean to gesture to the entire class of constructivist positions before turning to develop my preferred account.

3 Social Contracts: Contractarianism and Contractualism

Contract-based approaches to morality are, as with the views mentioned above, constructivist. They therefore lack the metaphysical commitments of moral realism that the evolutionary error theorist meant to leverage toward nihilism. And, furthermore, contract-based approaches are immune to some of the criticisms invoked against other subjectivisms above.

As a first proposal, let us consider:

5. Social Contract: " ϕ is wrong" =_{def} " ϕ was (or would have been) prohibited by rules agreed upon by informed, rational, and autonomous agents."¹³

A social contract approach avoids the metaphysical commitments of moral realism because, on this approach, there are no moral facts which exist *independently* of us; morality is something that we *construct* as informed, rational, autonomous agents. For example, consider John Rawls, who writes: "What justifies a conception of justice is *not* its being true to an order *antecedent* to and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us" (Rawls 1980, 519; emphasis in original). Rawls, like other social contract theorists, clearly denies the metaphysics that would be associated with moral realism.

In addition to an attractive metaphysics, social contracts also avoid the problems associated with many other subjectivisms. Of course, social contracts are subjectivistic in some sense: the moral edicts that they espouse are predicated upon the subjective wills of the contracting parties. Nevertheless, this need not be a problem since the *analysis* of moral claims that contract approaches would suggest is impervious to the criticisms mentioned in §2. For example, those more naïve subjectivisms analyzed moral claims to be reports regarding the (dis)approbative states of the speaker, and this was a problem since a proper analysis of moral claims should preserve the subject of those claims (i.e., the subject should be some act or practice, not the speaker). On the contract analysis that I proposed above, the aboutness of moral claims is preserved since both the claim and its analysans take the same subject (*viz.*, ϕ). Furthermore, moral debate is still possible since interlocutors could

¹³ Relatedly, we might say " ϕ is permissible" =_{def} " ϕ was (or would have been) allowed by rules agreed upon by informed, rational, and autonomous agents" and " ϕ is obligatory" =_{def} " ϕ was (or would have been) required by rules agreed upon by informed, rational, and autonomous agents."

disagree as to what was (or would have been) contracted by informed, rational, and autonomous agents; since the contract analysis does not reinterpret moral claims to be about the speaker, debate can be preserved.

Finally, social contracts *can* preserve the possibility of moral progress (or, relatedly, allow for moral fallibility) because of the “informed” provision that I have built into the analysans. Because of this provision, epistemic progress could lead to moral progress insofar as *informed*, rational, and autonomous agents would reach different moral conclusions given the accumulation of non-moral knowledge. For example, once we learn that the members of different races all have equivalent moral standing, we would be remiss to endorse slavery (or other repressive practices) on the basis of race; we could now say that moral progress was enabled by the acquisition of knowledge. Alternatively, we might *deny* that *informed*, rational, and autonomous agents could ever have constructed repressive practices in the first place. This line would hold that the parties that contracted those processes were not properly informed and that the measure of knowledge should not be against facts known at the time, but against a list of objective facts that always existed. If we adopted this second line, then moral progress would still be possible insofar as contracts were brought into line with what such idealized agents would have always chosen. So, either way, the possibility of moral progress is still preserved through a contract approach to morality.

There are, of course, numerous different contract approaches to morality, and I cannot address them all. Rather, I want to establish how the contract approach in general is congenial to evolutionary accounts of the moral sentiments. I will say more about this in §4, but I mean roughly that the evolutionary story can provide an empirical basis for the motivations of the contracting agents. For the remainder of this section, though, let me draw a distinction between two important contract approaches; this distinction will form much of the basis for the remainder of this paper.

All social contract theories share some set of central features, of which the most notable might be: “The idea that morality is deeply implicated in the very notion of agreement, and vice versa, so that whether an action is right or wrong must depend on whether the act accords with or violates principles that are, or would be, the object of suitable agreement between equals” (Darwall 2003, 1). As Stephen Darwall points out, it is important to realize that social contracts do not require that the contracting parties *actually* sat down and hammered out the details of a contract. Rather, contracts can be understood counterfactually: the legitimate contract is one that the parties *would have* chosen had they ever met at the bargaining table. While I will return to this idea in the next section, I want to focus another concept invoked by Darwall, that of “equals”. What does it mean for a contract to be the result of mutual agreement among *equals*? In what way are the parties supposed to be equal?

The answer to this question establishes a bifurcation into the two main branches of social contract theory: contractarianism and of contractualism. While the words are otherwise unimportant, each has been invoked by a representative of each camp and, by transitivity, has been attached to the entire tradition. So the contemporary flag bearer of contractarianism, David Gauthier, coined the nomenclature for his approach and, analogously, Thomas Scanlon is credited with the coinage of ‘contractualism’ (Gauthier 1991a, Scanlon 1982). What, then, is the substantive difference between the traditions and why does it matter?

For the contractarian, the equality of the contracting parties is “merely *de facto* and their choice of principles rationally self-interested”; this can be contrasted with contractualism which proceeds from “an ideal of *reasonable* reciprocity or fairness between *moral equals*”

(Darwall 2003).¹⁴ The roots of contractarianism can be uncontroversially traced to Thomas Hobbes, who wrote:

Nature hath made men so equal in the faculties of body and mind as that, though there be found one man sometimes manifestly stronger in body or of quicker mind than another, yet when all is reckoned together the difference between man and man is not so considerable as that one man can thereupon claim to himself any benefit to which another may not pretend as well as he. For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination or by confederacy with others that are in the same danger with himself. (Hobbes 1994, 74)

The equality among men is, for Hobbes, a *natural* equality as opposed to a moral equality because his contention is merely that individuals have nearly equal natural endowments (*viz.*, physical and mental). A second notable idea in Hobbes's contractarian view, shared by all contractualists, is the idea that morality does not even *exist* until after the covenant has been forged:

To this war of every man against every man, this also is consequent; that nothing can be unjust. The notions of right and wrong, justice and injustice, have there no place. Where there is no common power, there is no law; where no law, no injustice. Force and fraud are in war the two cardinal virtues. Justice and injustice are none of the faculties neither of the body nor mind. If they were, they might be in a man that were alone in the world, as well as his senses and passions. They are qualities that relate to men in society, not in solitude. It is consequent also to the same condition that there be no propriety, no dominion, no mine and thine distinct; but only that to be every man's that he can get, and for so long as he can keep it. (Hobbes 1994, 78)

This notion is the one that makes the contract approach to morality so metaphysically attractive, since it denies any sort of robust moral realism.

The final important feature of Hobbes's contractarian view has to do with the explanation for the formation of the contract in the first place. As Darwall noted, the contractarian tradition holds that morality is founded out of the self-interest of the contracting parties, and this is certainly an idea that has been supported by Hobbes (and, later, by Gauthier). For Hobbes, pre-societal life is unbearable; consider his famous proclamation that life in the State of Nature is "solitary, poor, nasty, brutish, and short" (Hobbes 1994, 76). Because this situation is so undesirable, the contracting parties form their covenant and transfer some of their rights to the sovereign. While they have fewer rights after the formation of the contract, they are nonetheless better off (because their adversaries are similarly restricted) than they were in the State of Nature. The contract therefore subserves self-interest and, in fact, would be motivated by it. While a lot more could be said about this idea, I only mention it in passing so that I may draw a contrast with the contractualist approach.

For the contractualist, there exist at least some norms prior to the contract, though those norms might vary. For example, consider Rawls' veil of ignorance, which restricts the extent to which we are able to pursue our self-interest since we are deprived of the

¹⁴ This notion of "moral equality" is a tricky one, though I will return to it briefly in §4. While Darwall does not provide an account in his (2003), he has gone on to develop an account that holds that moral equality is "equality in a kind of dignity of authority (to make claims and demands of one another—in particular to command a kind of respect I have called 'recognition respect')" (personal communication). See his *Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University Press, 2006). See also Elizabeth Anderson, "What Is the Point of Equality?", *Ethics* 109.2 (1999): 287–337.

information that would allow us to differentially favor ourselves in the contracting process; this is an inherently normative construct. Alternatively, consider Scanlon's version which holds that "an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement" (Scanlon 1999, 272).¹⁵ While the nuances are certainly important, we should at least be safe in assuming that Scanlon thinks that *reasonableness* is necessary for legitimate moral regulations. And, like Rawls's veil of ignorance, I take Scanlon's requirement of reasonableness to be more restrictive than the provisions made under Hobbesian laws.¹⁶ Abstracting away more general features of Rawls's or Scanlon's contractualisms, contractualists think that the pursuit of self-interest should not be the only factor dictating the terms of the contract (though surely it can be one); some other features should be interjected which, depending on the theory, could be fairness, justice, empathy, mutual respect, common understanding, etc.

To make this point another way, the contractualist assumes (or argues for) *more* than the contractarian. The reason is that contractualists would agree with Hobbes that the contracting parties tend to be *de facto* equals and, in *addition* to this, they build in moral constraints prior to the contracting process; Hobbes explicitly denies these constraints. Accepting this point, why is it relevant? For the picture that I want to develop, it has several implications. First, think about the reason that I invoked the contract approach in the first place: a contract approach could be responsive to the concerns that the evolutionary error theorists had with moral realism while, at the same time, avoid the moral nihilism to which they were committed. Second, I want to develop a meta-ethic which is empirically plausible. I shall discuss these points in turn.

4 Developing an Evolutionary Meta-Ethic

So first I should state my position: I am more sympathetic to contractarianism than contractualism. Why? Let us start with the metaphysics. Again, the reason I turned to contract theory was because it was less metaphysically committed than moral realism. Within contract theory, contractarianism is less metaphysically committed than contractualism since the former does not posit the pre-contractual norms or moral equalities of the latter. Therefore, in trying to keep my metaphysics simple, it makes sense to pick contractarianism over contractualism. But now the critic could object that, if I really wanted a simple metaphysics, I should pick *nihilism* over contractarianism. After all, nihilism is less metaphysically committed than even contractarianism since, for the nihilist, there are *no* moral facts (of any sort). So while contractarianism improves upon the mind-independent facts of moral realism, nihilism seems to improve upon contractarianism since the latter allows for the creation of moral facts through the contracting process.

This is challenge that should be taken seriously. But one obvious point to make is that metaphysical considerations are not the *only* ones that matter. The contractarian holds that

¹⁵ A somewhat simplified (as well as less rigorous) version of this statement might be: "If an act would be disallowed by any set of principles that no one could reasonably reject, then it is wrong."

¹⁶ For example, a principle which permitted the killing of non-philosophers would not be permissible under Scanlon's framework since this rule would *not* be part of a system that *no one* could reasonably reject. Rather *someone* (*viz.*, any non-philosopher), *could* have reason to reject this rule. Alternatively, I could endorse this rule under a Hobbesian system since it does not run contrary to my self-interest (nor any of the Laws of Nature).

agreement is somehow constitutive of morality, whereas the nihilist must deny this.¹⁷ Intuitively, the former position seems a lot more plausible to me than the latter. Imagine a case wherein two people each agree to ϕ , and then one breaks the agreement. Contrast this with a case where no agreement was ever made and one person ϕ 's while the other does not. In either case, the resultant outcomes are the same; the only difference is the agreement which existed in the former case. While the nihilist must deny that there are any moral differences in the cases, the contractarian can allow that the agreement in the first case constituted some moral feature that did not exist in the second case, and this is a strength of the contractarian position insofar as it has greater intuitive plausibility.

Another consideration that weighs against the nihilists' challenge would go as follows. Consider again the metaphysics postulated by the contractarian: the moral law comes into existence only after legislation and does not exist prior to the contract. On this view, the moral law is *continuous* with (or *emergent* from) features that the nihilist would find unproblematic: consent, agreement, autonomous will, etc. What it *means* for us to say that ϕ is wrong is no more nor less than " ϕ was legislated against by the contracting parties." This "wrongness" is, while morally substantial, quite metaphysically weak. Rawls, for example, does not even think that the contract theorist need to hold on to the *truth* of moral principles: "...[R]ather than think of the principles of justice as true, it is better to say that they are the principles most reasonable for us, given our conception of persons as free and equal and fully cooperating members of a democratic society" (Rawls 1980, 554). We might contrast this position with moral realism, wherein moral facts are *wholly discontinuous* with any features the nihilist would endorse. So, ultimately, the point is that the metaphysics of the contractarian really are not that much more committed than those of the nihilist and, furthermore, are only committed in a way to which the nihilist need not object.

So, for considerations of metaphysical simplicity, I think that I can safely endorse contractarianism over contractualism without being too concerned with the nihilist's objections. But I should be quite clear that I do not take the comparative metaphysical simplicity of contractarianism to serve as a reason to think that contractualism is implausible. Rather, the above comments merely go to my view that contractarianism is more attractive than contractualism. In fact, I think that it is quite hard to argue for the contractarian approach over the contractualist one, and the literature reflects this insofar as most of the associative views are developed independently of each other while their respective defenders typically fail to even consider the alternative line. Contractualists certainly have arguments as to where the pre-contract moral equality comes from (e.g., rooted in practical reason, respect for autonomous nature, etc.), but contractarians tend to find the starting points unconvincing. Alternatively, contractualists find that contractarian approaches implicitly (or explicitly) endorse egoism, and they often think that self-interest cannot serve as any foundation for morality. Regardless, without arguing for contractarianism over contractualism, I merely assert that I find the former more compelling, at least in part because of its metaphysical simplicity. But, returning to the significance that evolution might have for meta-ethics, there is another reason to think that contractarianism is more attractive than contractualism.

¹⁷ The reason that the nihilist would have to deny this would go something as follows. For the nihilist, moral predicates are never instantiated. Yet surely she would have to agree that agreement occurs (e.g., such as when two people agree to ϕ). If agreement is constitutive of morality, then the nihilist would have to accept that moral predicates could be instantiated whenever agreement occurs. Given her original thesis (and an uncontroversial assumption) she must, *a fortiori*, deny the consequent of this conditional. Therefore, by *modus tollens*, she must deny the antecedent as well.

Part of the project of meta-ethics is, of course, to explain the genesis of morality (e.g., of moral sentiments, of moral codes, etc.). This project is logically independent of the justificatory project, which would seek to justify various features of morality (e.g., moral codes). The distinction between explanations and justifications usually parallels a distinction between causes and reasons: causes explain whereas reasons justify. To accept these distinctions is to accept a fact-value divide, which obviously has its detractors.¹⁸ Nevertheless, the idea is that explanations are independent of any value assessment. So we might say that ϕ was caused (i.e., lies at the end of some explanatory chain) without being justified. For example, imagine that I strike my friend because I was unhappy. We could certainly accept that my unhappiness caused me to strike him, yet my action was not justified. These points, while not above discussion, are ones that I take to be plausible enough to be able to assume henceforth.

So look, we might say, there are two separate projects that ethicists must concern themselves with: the explanation and the justification of various facets of morality. Of course, not everyone would agree with this claim. Rather, many people would think that explanations of various features of our moral lives have nothing to do with moral philosophy at all and rather should be relegated to sociology, anthropology, etc. Insofar as evolutionary theory could only contribute to explanatory (as opposed to justificatory) projects within moral philosophy, these people might think that it is similarly irrelevant. For example, consider Wittgenstein: “The Darwinian theory has nothing more to do with philosophy than any other theory of natural science” (Wittgenstein 1961, 49).¹⁹

But this has to be wrong. As I think that the evolutionary error theorists have successfully argued, naturalized accounts of the moral sentiments inveigh against moral realism. So this is at least one instance wherein biological considerations have meta-ethical implications. But there are others as well, including those at the normative (as opposed to meta-ethical) level. For example, consider the “ought implies can” principle: it is common to think that obligation cannot persist given an inability to satisfy the obligation. If this principle is accepted, then empirical constraints have the ability to negate moral obligation. Insofar as moral philosophers care what an agent’s obligations actually are, it would be useful to think about what agents are able to do. Empirical results, (especially?) including those from biology, could inform this enterprise.

After this digression, we now return to another reason contractarianism is more attractive than contractualism: contractarians have a more empirically plausible account of the motivations that would drive the formation of a social contract. While I will say more about this below, it is worth reiterating that contractarians hold that social contracts are formed out of self-interest whereas contractualists hold that the contracts are formed, at least in part, because of mutual respect, understanding, etc.²⁰ If natural selection played an important

¹⁸ See, for example, W.D. Hudson, *The Is/Ought Question: A Collection of Papers on the Central Problems in Moral Philosophy* (London: Macmillan, 1969). For contemporary opposition to the fact-value divide, see Hilary Putman, *The Collapse of the Fact/Value Dichotomy and Other Essays* (Cambridge, MA: Belknap Press, 2004).

¹⁹ Wittgenstein allegedly did not even believe Darwin’s *biological* thesis, much less care for any of its purported philosophical implications.

²⁰ Some contractualists (such as Rawls) allow that self-interest could *also* play a role in the development of the contract. Such a view would not affect my conclusion that contractarianism offers a better motivational account since I view self-interest alone as more plausible than self-interest *and* anything else. I should clarify that I am speaking of *ultimate* (as opposed to *proximate*) motivations: I am not claiming that the contracting parties would necessarily think or feel that they were wholly being motivated by self-interest, but rather that any other putative motivations could be assimilated into a self-interested account (through, let us say, kin selection, reciprocal altruism, etc.).

role in the development of our moral sentiments, then those sentiments must contribute to their possessor's reproductive fitness. Even if we adopted a pluralistic approach to selection, positing both individual and group selection pressures, it is very unlikely that the moral sentiments would not confer *any* evolutionary advantages on their possessors individually (i.e., independent of group membership). The reason is that the group would be "invaded" by purely selfish individuals and the frequency of the group-altruistic phenotypes would diminish in future generations.²¹ So, for the moral sentiments to maintain their representation in the population (as they undoubtedly do), there would have to be at least some advantages that they confer upon their possessors. One such advantage is that the moral sentiments motivate the transition from the State of Nature into civil society; the reason is that we are predisposed to be sympathetic to the notion that certain legislations (e.g., against murder) are appropriate. Since individuals are better off in civil society than they are in the State of Nature, this transition is in their evolutionary interests. Contractarianism, with its emphasis on self-interest as the motivating factor for the creation of the social contract, is respectful of the motivational impetuses that evolution would have been likely to provide.²²

Alternatively, contractualism suffers serious deficiencies as an explanation for the emergence of the social contract. For the contractualist, the contracting parties are motivated, at least in part, by mutual respect and understanding. But, in virtue of natural selection's shaping of our moral sentiments, these are not likely to be strong motivators. If group selection played a substantial role in our evolution, then these motivations might be expected to exist to some degree, however the favored view seems to be that group selection did not play such a role. Rather, it is more plausible that the dominant mechanism in our evolution was individual selection, in which case we would have evolved to be self-interested rather than to be interested in mutual respect and understanding. Even if someone did think that mutual respect and understanding were strong motivators, I would offer two

²¹ For anyone interested in following this literature, the group selection was first invoked to support the evolution of our "moral sense" by Charles Darwin. See his *Descent of Man* (Amherst, NY: Prometheus Press, [1871] 1998), chs. 4–5. Group selection was powerfully challenged in George C. Williams, *Adaptation and Natural Selection* (Princeton: Princeton University Press, 1966). Richard Dawkins has continued the assault against group selection, and even against the traditional understanding of individual selection, by arguing that selection operates on the level of the gene rather than the group (or even the individual). See his *The Selfish Gene* (Oxford: Oxford University Press, 1982a) and *The Extended Phenotype* (San Francisco, CA: Freeman, Cooper, and Co., 1982b). Elliott Sober and David Sloan Wilson have attempted to resuscitate group selection by presenting a more sophisticated model than the naïve version which was exorcised in Williams's work. However, some critics regard their revised version as a theoretical possibility which was unlikely to have played a substantial role in the evolution of most species. See Elliott Sober, *The Nature of Selection* (Chicago: University of Chicago Press, 1984); "Screening-Off and the Units of Selection," *Philosophy of Science* 59 (1992): 42–52; and *From a Biological Point of View: Essays in Evolutionary Biology* (Cambridge: Cambridge University Press, 1994). See also David Sloan Wilson, "Levels of Selection: An Alternative to Individualism in Biology and the Human Sciences," *Social Networks* 11 (1989): 257–72; David Sloan Wilson and Elliott Sober, "Re-introducing Group Selection to the Human Behavioral Sciences," *Behavioral and Brain Sciences* 17 (1994): 585–654; and Elliott Sober and David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge, MA: Harvard University Press, 1998). For a critical response, to Sober and Wilson (1998) see Leonard D. Katz (ed.), *Evolutionary Origins of Morality* (Bowling Green, OH: Imprint Academic, 2000), pp. 207–56 and, for a general critique, see Robert Brandon, *Concepts and Methods in Evolutionary Biology* (Cambridge: Cambridge University Press, 1996).

²² I will offer a caveat to this claim in §5, particularly as pertains to kin selection and reciprocal altruism.

responses. First, it seems to me that, as a matter of phenomenological fact, they are not. Second, even if they were, the contractualist has not carried his burden. The reason is that mutual respect and understanding might be some proximate psychological mechanism that natural selection co-opted to serve its ultimate promotion of evolutionary interests. If my empirical thesis regarding the origin of the moral sentiments is correct, then contractualism *has* to be wrong regarding the (ultimate) motivations of the contracting parties since such motivations are incompatible with natural selection.^{23,24}

5 Further Considerations

In this last section, I would like to briefly discuss three further issues that surely warrant more discussion than I will provide herein. First, contractarians talk about the social contract being motivated by self-interest, and I have claimed that that evolutionary considerations will provide this motivation. But it could be objected that there are important incongruities between self-interest and evolutionary interests since self-interest is solely concerned with the interests of the agent *himself* whereas evolutionary interests track inclusive fitness (i.e., self-interest *plus* the interests of relatives, discounted by degree of relation). Presumably we can get these two interests to diverge: imagine that someone had to choose between saving himself and his three children. In this case, his self-interest mandates that he save himself, whereas his evolutionary interests mandate that he save his three children (since his aggregate genetic representation in them would *exceed* his genetic representation in himself). So, given the possibility of such divergences of self- and evolutionary interests, is contractarianism still as attractive as I have claimed?

I think so, though more would need to be said. In particular, we would need to re-equilibrate the ends of the social contract and of evolution, and this could proceed in either of two ways: we could argue that evolution perpetuates *individual* self-interest or we could argue that social contracts subserve inclusive fitness. The first line obviously fails since evolution does not care about individual organisms and only cares about propagating genetic material to future generations. But what about the second? Certainly it would not take a dramatic reconception of contractarianism to propose that social contracts would be made to subserve inclusive fitness as opposed to individual fitness. Look, we might say, the State of Nature is a really bad place, and it would be nice to get out of it and into civil society. But it would not be bad just for us, it would also be bad for all of our relations, and surely we would not want to leave them unprotected. In fact, our concern for them would undoubtedly play a significant role in the societal structures that we would go on to legislate since we would not be impartial toward their welfare (cf., Hume's expanding circle of concern (Hume 1978, pp. 351–7, 483–4, 602–6; Hume 1975, 295–302)).

It might appear that our desire to extract our relations from the State of Nature is altruistic, though a better interpretation is that this desire is (evolutionarily) “self”-interested

²³ For alternative empirical presuppositions, particularly regarding the possibility of genuine altruism, see Sober and Wilson (1998).

²⁴ I suppose the contractualist could concede that some of his proposed motivations reduce to self-interest without being willing to abandon his theory. Rather, he might argue that contractarianism, while ultimately correct, nevertheless needlessly collapses descriptively varied motivations into a unitary motivational scheme. However, I take such a concession to be quite extreme for the contractualist, since it throws away the primacy of the motivations that distinguished him from the contractarian.

insofar as our relations share genetic material with us. In other words, the rational approach, from an evolutionary perspective, would be to help those to whom we are related. And, even if our compassion seems altruistically motivated, we could always ask where altruism would have come from in the first place. A likely explanation would be that evolution has endowed us with *apparently* altruistic tendencies that ultimately promote inclusive fitness.²⁵ Regardless, I think that it is quite plausible to suggest that contractarian motivations would not, as a matter of empirical fact, be merely confined to rational *self*-interest, but would extend to include the welfare of relatives. If this is true, we might need some refinements of the contractarian line, but we could probably get pretty far by replacing ‘self-interest’ with ‘evolutionary interest’ throughout.

A second issue worth mentioning concerns the dynamic by which the social contract would evolve. On traditional models, contractarian approaches could be explained by appeals to classical game theory: the actual contract would represent the optimal results that each agent could gain for himself given similar motivations by each other party to the contract.²⁶ For example, while I would be happy to secure the entirety of the social product, my co-contractors, each trying to maximize their own share of the social product, would not be willing to grant me a disproportionately large share. Imagine that empirical constraints dictated that, so long as the sum requisitions of each contracting agent exceeded the amount of the social product, it would be (wholly or partially) forfeited. This is certainly possible given that conflict over some resource could lead to the damage or loss of that resource. Presupposing this constraint, the rational course of action would be self-regulation such that each agent would put in for a smaller part of the social product than he would absent said constraint. All deliberations to the contract could presumably proceed along analogous lines and, in each case, the rational course of action for each agent would be that which sought to maximize his advantage through the associative provision.

But, again, we have a potential decoupling of contractarianism and evolution: contractarianism can be modeled along classical game-theoretic decision procedures and evolution cannot. Why would evolution not fit with classical game-theoretic models? It should be pretty clear that the dynamic that leads to evolutionary change is not rational choice (as in classical game theory), but rather reproductive success; only a vanishingly small percentage of the entities affected by evolution even *has* rationality, so it is obviously not the dynamic on which selection operates. Insofar as I want to argue that evolution gives rise to the social contract, it seems that I have a problem since the social contract operates on classic game-theoretic models whereas evolution operates on different dynamics.²⁷

Again, this is an important challenge and one that deserves more space than I will afford it. In order to restore the compatibility of contractarianism and evolution, I would presumably want to re-equilibrate the dynamics on which they operate. Since it would be wildly implausible to suggest that evolution operates on classical game-theoretic dynamics,

²⁵ Again, for a contrary proposal, see Sober and Wilson (1998).

²⁶ There is a huge literature regarding classical game theory and its applications, but some of the seminal works are John von Neumann and Oskar Morgenstern’s *Theory of Games and Economic Behavior*, 60th anniversary ed. (Princeton, NJ: Princeton University Press, [1944] 2004) and John Nash, “The Bargaining Problem,” *Econometrica* 18.2 (1950): 155–62 and “Two-Person Cooperative Games,” *Econometrica* 21.1 (1953): 128–40.

²⁷ For a nice introduction to the differences between classical and evolutionary game theory, see Herbert Gintis, “Classical versus Evolutionary Game Theory” in Katz (2000), pp. 300–4. A less technical explanation can be found in William Poundstone, *The Prisoner’s Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb* (New York: Doubleday Anchor, 1992), pp. 231–56.

I would need to construct a social contract that evolves by evolutionary dynamics. Fortunately, this task has been carried out by more able people, but there are some comments that I can make here.²⁸ First, there is no necessary reason that the contract must be held to classical models. Certainly Hobbes's contractarianism is congenial to classic game-theoretic renderings, and Gauthier explicitly bases his contractarianism on these classical models (Gauthier 1986, 157–189). Nevertheless, the central elements of contractarianism (e.g., *de facto* equality among contractors, self-interested motivation, etc.) are silent as to the procedure by which the contract develops and, while the dominant tradition has been to assume rational choice, evolutionary game theory might be able to provide a parallel explanation that could be more empirically plausible.

Note, however, that this alternative conception might not even be that novel since there are historical sources that are eerily consistent with evolutionary game theory. Consider David Hume:

When this common sense of interest is mutually expressed, and is known to both, it produces a suitable resolution and behavior. And this may properly enough be called a convention or agreement betwixt us, though without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are performed upon the supposition, that something is to be performed on the other part. Two men, who pull the oars of a boat, do it by an agreement or convention, though they have never given promises to each other. (Hume 1978, 490)

On Hume's notion of convention, no explicit guarantees were ever made among the participating parties. But the point goes even deeper than that since Hume does not even depend on the notion of either actual or counterfactual rational choice—herein signals his departure from classical game-theoretic models. Just as convention emerges among the oarsmen, we might imagine (moral) convention emerging across evolutionary time. This gradual emergence of norms is certainly what the evolutionary game theorist would want to posit. And, if we allow that the Humean account can be adopted to *explain* the emergence of norms, we could still hold that actual or counterfactual agreement would provide the relevant *justificatory* features.

The final remarks that I would like to make, and probably the most deserving of more room, concern the rationality of upholding the social contract. Even if we were to allow that, in general, the social contract subserves evolutionary interests, there could (and would) nevertheless be *specific* instances wherein the two are at odds. We all revel in the broad security conferred by the social contract since civil society is so much more amiable than the State of Nature. But imagine a specific instance, when nobody else is looking, such that I could easily dispose of an adversary and steal some of his resources. What should I do? Of course, I am overall better off given the existence of the social contract (since it restrains my enemies), but what if my transgressions would go rewarded and unpunished? If, for example, I had Gyges' fantastic ring, would there be any reason to honor the social contract?

²⁸ See Brian Skyrms, *The Evolution of the Social Contract* (Cambridge: Cambridge University Press, 1996), "Game Theory, Rationality, and Evolution of the Social Contract," *Journal of Consciousness Studies* 7.1–2 (2000): 269–84, and *The Stag Hunt and the Evolution of Social Structure* (Cambridge: Cambridge University Press, 2003). See also Ken Binmore, *Game Theory and the Social Contract, Vol. 1: Playing Fair* (Cambridge, MA: MIT Press, 1993) and *Game Theory and the Social Contract, Vol. 2: Just Playing* (Cambridge, MA: MIT Press, 1998).

Gauthier has been very interested in this problem, though his account has drawn much critical response.²⁹ As a first move, he asks that we imagine that moral agents are transparent or, more realistically, translucent such that their motives and future actions are at least somewhat apparent to those with whom they contract. For example, imagine that you and I are discussing whether to assist in mowing each other's lawns, and you are skeptical of my promise to help you *after* you help me because I could renege. If my motivations and future actions were wholly opaque (i.e., such that you would have *no* evidence either way how I would act after you completed your end of the bargain), it would be hard to ever get this agreement off the ground since I would have every reason to defect after you helped with my lawn.³⁰ But, as a matter of empirical fact, it is highly unlikely that my motivations and future actions would be wholly opaque. Rather, translucency seems much more plausible: we live in fairly small social networks, we gain fairly lasting reputations, most of us have fairly stable dispositions, etc.³¹ Given these features there would be, in most cases, ways of attaining evidence for whether contracting parties would fulfill their contracts. Therefore, it would be likely that the contracting parties would fulfill their contracts since, otherwise, the mutually advantageous contracts would never become established. To this end, Gauthier thinks that moral agents adopt certain dispositional stances toward the honoring of contracts (so-called "choice dispositions"), and these dispositions, coupled with translucency, get the contracting process started.

How does this story satiate the amoralist? My answer, different from Gauthier's, would go as follows. So long as the translucency fell short of transparency, the choice dispositions would have to be reasonably strong, otherwise co-contractors would expect some defections. And, given the advantages of the contracts, there would be every reason for the dispositions to develop the requisite strength to ensure the formation of the contracts.³² But dispositions are non-cognitive and, as such, are not (directly) within the reach of reason. If the disposition were strong enough then, even if the contracting parties recognized that defection was rational, they might nonetheless be *incapable* of violating their covenants. If this is right, then there would not actually be a reason to honor one's contracts, but the contracts would nonetheless be stable given the unlikelihood of defection. As a matter of empirical fact, it is clearly not the case that our choice dispositions are so

²⁹ See Geoffrey Sayre-McCord (1991), "Deception and Reasons to be Moral"; David Copp (1991), "Contractarianism and Moral Skepticism"; Holly Smith (1991), "Deriving Morality from Rationality"; Jody S. Kraus and Jules L. Coleman (1991), "Morality and the Theory of Rational Choice"; and Peter Danielson (1991), "Closing the Compliance Dilemma: How It's Rational to Be Moral in a Lamarckian World". All these articles, and Gauthier's (1991b) response, "Rational Constraint: Some Last Words," are in Vallentyne (1991).

³⁰ Assume that, unbeknownst to you, I am moving to another state the following day and that I would never suffer adverse consequences for my breach.

³¹ Note also that, in evolutionary environments, at least these first two conditions would likely have been met, thus making translucency a reasonable empirical hypothesis—at least historically if less so contemporarily.

³² One reason that I like the appeal to these dispositions as a cement for social contracts is that dispositions can have evolutionary explanations. Certainly some do not, such as my disposition to travel during the summers. But many do: dispositions to care for family members, dispositions to seek food and water, etc. Insofar as Gauthier's choice dispositions add stability to fitness-serving contracts, they quite likely would have an evolutionary basis. Gauthier would, I think, be congenial to this suggestion as he approvingly acknowledges Robert Triver's evolutionary account of reciprocal altruism (and its postulation of evolved guilt to secure reciprocity). See Gauthier (1986), pp. 187–9 and Robert Trivers, "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46.4 (1971): 35–57.

amazingly strong as to render contractual violation psychologically impossible. Nevertheless, I do think that it would be plausible to hold that psychological dispositions toward honoring a social contract are strong enough that we need not worry that the defections offered by the amoralist would completely subvert it.

References

- Anderson E (1999) What is the point of equality? *Ethics* 109:287–337 doi:10.1086/233897
- Ayer AJ ([1936] 1946) *Language, truth, and logic*, 2nd edn. Dover, New York
- Binmore K (1993) *Game theory and the social contract vol 1: playing fair*. MIT, Cambridge MA
- Binmore K (1998) *Game theory and the social contract vol 2: just playing*. MIT, Cambridge MA
- Blackburn S (1984) *Spreading the word*. Oxford University Press, Oxford
- Blackburn S (1993) *Essays in quasi-realism*. Oxford University Press, Oxford
- Blackburn S (1998) *Ruling passions*. Clarendon Press, Oxford
- Brandon R (1996) *Concepts and methods in evolutionary biology*. Cambridge University Press, Cambridge
- Brown D (1991) *Human universals*. McGraw-Hill, Boston MA
- Copp D (1991) Contractarianism and moral skepticism. In: Vallentyne P (ed) *Contractarianism and rational nature*. Cambridge University Press, Cambridge, pp 196–228
- Danielson P (1991) Closing the compliance dilemma: how it's rational to be moral in a lamarckian world. In: Vallentyne P (ed) *Contractarianism and rational nature*. Cambridge University Press, Cambridge, pp 291–322
- Darwall S (ed) (2003) *Contractarianism/contractualism*. Blackwell, Malden MA
- Darwall S (2006) *Second-person standpoint: morality, respect, and accountability*. Harvard University Press, Cambridge MA
- Darwin C ([1871] 1998) *The descent of man*. Prometheus, Amherst NY
- Dawkins R (1982a) *The selfish gene*. Oxford University Press, Oxford
- Dawkins R (1982b) *The extended phenotype*. Freeman, Cooper, and Co, San Francisco
- Gauthier D (1986) *Morals by agreement*. Oxford University Press, Oxford
- Gauthier D (1991a) Why contractarianism? In: Vallentyne P (ed) *Contractarianism and Rational Nature*. Cambridge University Press, Cambridge, pp 15–30
- Gauthier D (1991b) Rational constraint: some last words. In: Vallentyne P (ed) *Contractarianism and rational nature*. Cambridge University Press, Cambridge, pp 323–330
- Gibbard A (1990) *Wise choices, apt feelings: a theory of normative judgment*. Harvard University Press, Cambridge MA
- Gintis H (2000) Classical versus evolutionary game theory. In: Katz LD (ed) *Evolutionary origins of morality*. Imprint Academic, Bowling Green, OH, pp 300–304
- Gould SJ, Lewontin R (1978) The spandrels of san marco and the panglossian paradigm. *Proc R Soc Lond* 205:581–598
- Hare RM (2000) *Sorting out ethics*, new edn. Oxford University Press, Oxford
- Hauser M (2006) *Moral minds: how nature designed our universal sense of right and wrong*. Ecco, New York
- Hobbes T ([1669] 1994) *Leviathan*. Curley E (ed) Hackett Publishing Co, Indianapolis IN
- Horgan T, Timmons M (2006) Cognitivist expressivism. In: Horgan T, Timmons M (eds) *Metaethics after moore*. Oxford University Press, Oxford, pp 255–298
- Hudson WD (1969) *The is/ought question: a collection of papers on the central problems in moral philosophy*. Macmillan, London
- Hume D ([1739–1740] 1978) *A treatise of human nature*. Selby-Bigge LA, Nidditch PH (eds) Oxford University Press, Oxford
- Hume D ([1751] 1975) An enquiry concerning the principles of morals. In: Selby-Bigge LA (ed) *Enquiries concerning human understanding and concerning the principles of morals*, 3rd edn. Clarendon Press, Oxford, pp 167–323
- Joyce R (2000) Darwinian ethics and error. *Biol Philos* 15:713–732 doi:10.1023/A:1006742813396
- Joyce R (2001) *The myth of morality*. Cambridge University Press, Cambridge
- Joyce R (2006) *The evolution of morality*. MIT Press, Cambridge
- Katz LD (ed) (2000) *Evolutionary origins of morality*. Imprint Academic, Bowling Green OH
- Kraus JS, Coleman JL (1991) Morality and the theory of rational choice. In: Vallentyne P (ed) *Contractarianism and rational nature*. Cambridge University Press, Cambridge, pp 254–290
- Nash J (1950) The bargaining problem. *Econometrica* 18:155–162 doi:10.2307/1907266

- Nash J (1953) Two-person cooperative games. *Econometrica* 21:128–140 doi:[10.2307/1906951](https://doi.org/10.2307/1906951)
- Plato (1981) *Euthyphro*. In: Grube GMA (trans) Plato, five dialogues. Hackett Publishing Co, Indianapolis IN, pp 5–22
- Poundstone W (1992) *The prisoner's dilemma: john von neumann, game theory, and the puzzle of the bomb*. Doubleday Anchor, New York
- Putman H (2004) *The collapse of the fact/value dichotomy and other essays*. Belknap, Cambridge MA
- Rachels J (1999) *The elements of moral philosophy*, 3rd edn. McGraw-Hill, Boston, MA
- Rawls J (1980) Kantian constructivism in moral theory. *J Philos* 77:515–572 doi:[10.2307/2025790](https://doi.org/10.2307/2025790)
- Ruse M (1984) The morality of the gene. *Monist* 67:167–199
- Ruse M (1985) Evolution and morality. *Philos Exch* 16:5–26
- Ruse M (1986a) Evolutionary ethics: a phoenix arisen. *Zygon* 21:95–112 doi:[10.1111/j.1467-9744.1986.tb00736.x](https://doi.org/10.1111/j.1467-9744.1986.tb00736.x)
- Ruse M (1986b) *Taking darwin seriously: a naturalistic approach to philosophy*. Blackwell, Oxford
- Ruse M (1989) *The darwinian paradigm: essays on its history, philosophy and religious implications*. Routledge, London
- Ruse M (1993) The new evolutionary ethics. In: Nitecki MH, Nitecki DVN (eds) *Evolutionary ethics*. State University of New York Press, Albany NY, pp 133–162
- Ruse M (1999a) Evolutionary ethics: what can we learn from the past? *Zygon* 34:435–451 doi:[10.1111/0591-2385.00225](https://doi.org/10.1111/0591-2385.00225)
- Ruse M (1999b) *Monad to man: the concept of progress in evolutionary biology*. Harvard University Press, Cambridge MA
- Ruse M ([1979] 1999c) *The darwinian revolution: science red in tooth and claw*, 2nd edn. University of Chicago Press, Chicago
- Ruse M, Wilson EO (1985) The evolution of ethics. *New Sci* 108:50–52
- Ruse M, Wilson EO (1986) Moral philosophy as applied science. *Phil* 61:173–192
- Sayre-McCord G (1986) The many moral realisms. *South J Philos* 24:1–22
- Sayre-McCord G (1988) *Essays in moral realism*. Cornell University Press, Ithaca, NY
- Sayre-McCord G (1991) Deception and reasons to be moral. In: Vallentyne P (ed) *Contractarianism and rational nature*. Cambridge University Press, Cambridge, pp 181–195
- Scanlon TM (1982) Contractualism and utilitarianism. In: Sen A, Williams B (eds) *Utilitarianism and beyond*. Cambridge University Press, Cambridge, pp 103–128
- Scanlon TM (1999) *What we owe to each other*. Belknap, Cambridge MA
- Shafer-Landau R (2003) *Moral realism: a defense*. Oxford University Press, Oxford
- Skyrms B (1996) *The evolution of the social contract*. Cambridge University Press, Cambridge
- Skyrms B (2000) Game theory, rationality, and the evolution of the social contract. *J Conscious Stud* 7:269–284
- Skyrms B (2003) *The stag hunt and the evolution of social structure*. Cambridge University Press, Cambridge
- Smith H (1991) Deriving morality from rationality. In: Vallentyne P (ed) *Contractarianism and rational nature*. Cambridge University Press, Cambridge, pp 229–253
- Sober E (1984) *The nature of selection*. University of Chicago Press, Chicago
- Sober E (1992) Screening-off and the units of selection. *Philos Sci* 59:42–52 doi:[10.1086/289658](https://doi.org/10.1086/289658)
- Sober E (1994) *From a biological point of view: essays in evolutionary biology*. Cambridge University Press, Cambridge
- Sober E, Wilson DS (1998) *Unto others: the evolution and psychology of unselfish behavior*. Harvard University Press, Cambridge MA
- Trivers RL (1971) The evolution of reciprocal altruism. *Q Rev Biol* 46:35–57 doi:[10.1086/406755](https://doi.org/10.1086/406755)
- Vallentyne P (ed) (1991) *Contractarianism and rational nature*. Cambridge University Press, Cambridge
- Von Neumann J, Morgenstern O ([1944] 2004) *Theory of games and economic behavior*, 60th edn. Princeton University Press, Princeton, NJ
- Williams GC (1966) *Adaptation and natural selection*. Princeton University Press, Princeton
- Wilson DS (1989) Levels of selection: an alternative to individualism in evolutionary biology and the human sciences. *Soc netw* 11:257–272
- Wilson DS, Sober E (1994) Re-introducing group selection to the human behavioral sciences. *Behav Brain Sci* 17:585–654
- Wittgenstein L (1961) *Tractatus logico-philosophicus*. Pears DF, McGuinness B (trans) Routledge & Kegan Paul, London