William Harms
Draft, April 24, 1999
e-mail: bharms@interchange.ubc.ca

The Evolution of Normative Systems

Abstract: The key to the naturalistic analysis of norms is incorporating adaptive histories into standard functional analysis. The rules of adapted design specify truth conditions for signals in adapted signaling systems, and their failures constitute the truth conditions for signals in function enforcement mechanisms. If function enforcement mechanisms lie behind the normative impulses (intuitions) expressed in normative utterances, then they can have correspondence truth, and they refer essentially to objective, external rules. This analysis is easily formalized, and extends to incorporate cultural flexibility and the role of reason in normative deliberation. It accommodates the bulk of philosophical intuitions, and also promises a fruitful avenue by which the investigation of norms can proceed throughout the biological and social sciences.

Philosophers spend a lot of time worrying about rules. We worry about how one ought to live, about the rules of justification for beliefs and actions, about what it would be like if the rules of reason were rigorously followed, about what the rules are for scientific enquiry, about which rules govern the meaning of signs and the intentions of agents, and so on. Sometimes, we argue that there are no such rules as most of us want to believe there are, rules which apply to all of us collectively and to each of us individually, which are beyond our ability to change, and whose violation is in some simple sense *wrong*. To this we often respond that without such rules we are all made somehow less, that our normative deliberations are a sham, or even that the whole business of living becomes somehow pointless.

Naturalism, by which I mean the commitment to understanding human beings in terms of the natural sciences, creates a peculiar tension for its adherents. On the one hand, the naturalist is driven by the commitment to some very basic rules of admissible concepts and justification of belief. On the other hand, this very commitment has resulted in world-views in which there is nothing out there for us to be worrying about when we worry about rules. Hume's analysis of the status of "oughts" consisted largely of surveying the sorts of natural relations available to empiricism, and eliminating them one by one as candidates for the referents of "oughts". There is nothing in nature that sufficiently resembles the rules that concern us. Hence, normative utterances can be no more than the expression of *sentiments*, or in more contemporary terms, the expressions of emotions, desires, or feelings. Mackie's (1977) "argument from queerness" which casts such a long shadow over contemporary ethics reaches the same conclusion. While there is a good bit of controversy concerning what we can or ought to do about this disappointing state of affairs — how we might restore confidence in our justificatory practices — there is virtually none regarding whether or not science might somehow discover the natural basis for the rules that bind agents.

The purpose of the present paper is to argue that the contemporary scientific world-view includes certain elements which were not part of eighteenth century empiricism, and whose presence calls for a reassessment of what science has to offer our understanding of norms at the most basic level. In particular, current biology rests crucially on relationships that are essentially *historical* — the relations of common descent that exist between ancestors and descendants as well as between current cousins, and the relationships of *adaptation* that exist between traits and environments. They are uncontroversially *causal* (and thus explanatory) relationships, but of a kind to which empiricist attempts at rigor have traditionally been blind. What I want to suggest is that, for the *biological* naturalist, there is after all something out there for us to be worrying about when we worry about rules — namely the rules of adapted design — and that these rules have very much the character that we believe our rules to have. It is not the intent of this paper to claim that these are *always* the kinds of rules we are worrying about, but that it is plausible that much of the time we are in fact worrying about such rules, and when we are, there can be a distinctive sense in which it is objectively true that it is wrong for us to violate them.

The difficulty with such a proposal is the very depth of its implications. It will be helpful to establish some ground rules. Philip Kitcher (1994) , as part of his extended critique of E.O.Wilson's Sociobiological account of morality, distinguished four ways of applying biology and the social sciences to ethics. The first two involve mere identification of historical facts about human nature — helpful but not directly relevant to our normative concerns. The third consists in "explaining what ethics is all about". The fourth consists in using such an account to actually *guide* normative deliberations. This paper is devoted to the third sort of project — in particular, to taking steps toward providing resources which make possible a more satisfying naturalistic account of norms.

Kitcher posed a challenge for evolutionary theories which propose a complete account of norms. Such theories must either say what it is that makes normative utterances true, or else, if they are taken to be mere expressions of momentary impulses, the theory must say on what basis deviants are to be judged. Ethicists tend to have longer lists. Darwall, Gibbard and Railton, for instance, propose that,

> Understanding the commitments of ordinary moral or value discourse and practice would appear to involve accounts of at least the following: the semantics of morals and value; the apparent metaphysical status of moral properties or values; the putative epistemology of morality or value theory; and the relation of morality or values to practical reasoning. (1992, 127)

In addition to Kitcher's requirement that one supply truth conditions, we must at some point be able to address something like Mackie's "queerness" argument. We must also explain how it is that we can come to know about (e.g.) moral truth conditions. Finally, we will need say something about how reason which proceeds in terms of propositional language interacts with the epistemology of normative truth-makers. We can add to this Mackie's other main argument, the "argument from relativity". If the truth of ought-statements is a matter of correspondence to external facts, then how is it that moral codes vary from culture to culture as much as they do?

Inasmuch as dualism is out of fashion, one need not argue that there are in fact processes in the brain which underlie normative impulses and deliberation. Nor is it necessary to argue that many such processes are products of evolutionary design. Consequently, we can restrict ourselves to showing how incorporating the process of evolutionary design changes the naturalistic picture of norms. My proposal, in a nutshell, is this: The evolutionary design process involves rules which pertain to adapted mechanisms. In the case of adapted signaling systems, these rules specify correspondence maps which give truth (or satisfaction) conditions for the signals involved. Such maps are specified even for signals which express no propositional content. So if normative impulses are issued by adapted signaling systems, they can be true or false by correspondence to the world, and yet irreducible to propositional language, without the introduction of "queer" properties or facts. Normative systems (function enforcement mechanisms) are distinguished by the fact that their corrective signals represent the violation of other rules of design. Finally, behavioral control systems can be remarkably sophisticated in the kinds of flexibility that are implemented and stabilized. Cultural variability is allowed by this flexibility, and the peculiar relation of reason to norms can be understood in terms of the interdependencies that have developed between various control systems.[1]

The proposal, then, is that much of the time, when people worry about *rules*, what they are in fact worrying about are rules of adapted design. The strategy will be to provide a theory of meaning in order to give an account of how these worries might come to be *about* such rules. The plausibility of such an account rests, for the philosopher, on the extent to which it can accommodate and illuminate basic features of normative deliberation. As an empirical hypothesis, its ultimate tenability rests on whether or not the facts cooperate. I will address a sort of ethicist's "wish list" for accounts of norms at the end. But this does not signal an intent to build normative theory, as that is usually understood. There is ample reason to expect that the present account falls short of yielding everything normative theorists have wanted from an account of morality, knowledge, or reason. In particular, it does not seem as though the account can tell us *what we ought to do*. But it can do a number of other things that are worth doing. First, it can explain with some precision naturalism's limitations *vis a vis* normative deliberation. The is/ought distinction need not remain the mystery it seems to be. Second, by recognizing external standards and truth conditions for normative intuitions, it can counteract the generally corrosive effect of scientific materialism on realism about morality, knowledge, and reason. Third, it can provide basic theoretical tools which allow more satisfying accounts of meaning and normativity throughout the biological and social sciences. I suppose that, if one had it in mind to build realist normative theories, the present account might provide resources for such an endeavor. Such considerations lie beyond the scope of the present paper, however.

Empiricist Biases and Biological Resources

If history is any indication, there are two sorts of rules that pass empiricist muster. First, there are the unalterable regularities we call "laws of nature." Second, empiricists recognize rules

[1] The proposed hypothesis owes an enormous debt to the work of Ruth Millikan (1984, 1993). The central insight regarding biological semantics is hers. For present purposes, however, her theory depends overly much on stipulative definitions of terms, and is directed primarily at the analysis of propositional language, rather than at the non-propositional semantics of adapted signaling systems crucial to the understanding of norms.

issued by cultures, either explicitly, or in enforced patterns of behavior. It would seem that, taking physics as the paradigm science, empiricism has sought to impose rigor on our investigation of the world by focusing on physical structures, physical similarity, and actual causal interactions. Bringing such a set of tools to the analysis of norms, there is not much that one can say that is very satisfying. I can assess the proximal causes of my normative utterances — to say that something is "wrong" is to express a particular sort of feeling of revulsion. Moving back a bit, I can say that my normative utterances express the rules of my culture, which have played a causal role in my revulsion having the character it does.

The difficulty that such a meager set of resources leaves us with is, of course, quite familiar. If there are normative truths, what grounds them? And if not, then what does one say about deviants? Put another way, our normative rules such as we understand them can not be laws of nature, since we violate them all the time. Indeed the whole point of such rules seems to be that we *can* violate them. If we couldn't, the norms wouldn't be necessary. On the other hand, our ordinary understanding of norms is such that cultures can be wrong — everyone can be wrong. It seems that the rules that interest us are not the rules that we actually follow, but the rules that we are *supposed to* follow, in some sense that needs to be made clear.

In the face of such difficulties, the usual thing for the hard-nosed naturalist to do is admit that there are no objective norms, though it may be that the illusion that there are such norms is useful.[2] There are, of course, any number of roughly internalist strategies for attempting to justify our normative appeals.[3] The point of the present paper, however, is to suggest that naturalists have been precipitous in their bullet-biting, and to show how one can construct a externalist/realist semantics for norms using only uncontroversial resources of the current scientific world-view. The hope is that the extension of biofunctional semantics can provide the resources for a richer descriptive account of norms, which may, if the facts cooperate, go some ways toward corroborating our ordinary confidence in our normative practices and intuitions — an evolutionary epistemology of pre-linguistic intuitions, if you like.

---

[2] Ruse and Wilson 1986, Dawkins 1976.

[3] Darwall, Gibbard, and Railton (1992) give a nice overview of the current state of the debate.

How does this work? It is uncontroversial that many of the traits carried by extant organisms are adaptations.[4] A trait is an adaptation if it has been selected for — if it has through its functioning contributed to the reproductive success of the lineage that carries it. There are *rules* implicit in the process of adaptation. The rules are described by what the trait has in fact done in the past that caused it to be selected *for*, and in particular, *how* the trait went about making that contribution. There are rules for the formation and operation of the lens in the human eye, for the unfolding of a bat's wing, for the arrangement of chemical sensors in bacteria, for the distribution of photo synthetic chlorophyll in a sunflower, and so on. Such rules are as ubiquitous and uncontroversial as adaptations.

The reader will be relieved to know that I am not going to suggest that the rules of evolutionary designs are in themselves normative. There is no plausible norm that says that one ought to use traits according to design[5], nor do I think that there is any reason why promoting the good of the species or the process of evolution is directly normative.[6]

Nonetheless, the rules of evolutionary design are the key to the naturalistic analysis of norms. The first thing to realize is that, despite the obvious epiphenomenality of such rules, their existence is objective, and thus proper subject for naturalism. The facts that determine what rules apply to a trait are quite as difficult to assess as any other fact about evolutionary history. But that there are such facts is not in doubt, nor would the resulting rules be any more mysterious than any other fact if the actual selective history were known.

More important is the character of the rules. What we have been looking for, and have had such difficulty finding within traditional empiricism, are rules which are objective, at least to the extent of being extra-cultural, and yet can be violated, unlike the laws of nature. The rules of evolutionary designs have both of these properties. Being entrenched in the history of adapted traits, they are standards which are beyond the ability of cultures to dictate, and moreover we share them in common insofar as we share common descent. On the other hand we are free to

---

[4] Sober 1984 and especially 1993 chapter 2 convinced me that one can let the biologists worry about defining selection, adaptation, and function. Cf. Millikan 1984, chapters 1&2.

[5] There are those who will argue otherwise. See for instance Levin 1984.

[6] See Bradie 1994 for the history of the latter sort of thinking.

violate the rules of evolutionary designs. We do so with impunity.

In sum, biological naturalism recognizes three kinds of rules, whereas traditional empiricist naturalism has traditionally recognized only two. That this might make a profound difference in the available accounts of norms is suggested by the fact that the third kind of rule shares with norms properties whose absence has created much of the difficulty faced by traditional accounts.

Simple Functional Semantics

Getting norms out of the rules of design takes two steps. The first is showing how signals can get correspondence truth (or satisfaction) conditions. The second is showing in what kinds of control structures signals refer to the satisfaction or failure of a rule. We will address these in order.

Common wisdom has it that evolution gives us a world of causal processes devoid of intentionality, that elusive relationship that exists between sign and signified. We seem to think that, somehow, you need a *mind* to generate correspondence between signs and their satisfaction conditions. Nothing could be further from the truth. Indeed, the big question is not how there can be correspondence relationships in the natural world, but rather, *which* correspondence relationships we are concerned with when we think about language, meaning, and normativity. An example will be helpful at this point. Cheney and Seyfarth (1990) discuss a signaling system in vervet monkeys in Kenya. It seems that vervets must cope with three kinds of predators, pythons, eagles, and leopards. Vervet sentries issue three kinds of warning cries and hearers engage in three kinds of evasive strategies. The cry for eagles causes vervets to look up into the sky. The cry for pythons causes them to stand up and look around on the ground. The cry for leopards causes them to run up the nearest tree. The parts of this signaling system are designed to work together. The vervets' perceptual system must result in the issuance of the right cry for the circumstances. What makes it the right cry depends on what the vervets are supposed to do in response to the cries. The design of the signaling system is such that there are rules for the issuance of signals, and rules for responding to those signals. The two sets of rules must fit together as part of a unified design, otherwise the system as a whole does not function.[7]

Neither the rules for signal production nor the rules for signal interpretation (which behavior is appropriate) give the correspondence maps, however. The mapping we are after derives from the kind of coordination with the world that the system has maintained which explains its historical benefits, that is, explains why the signaling system has been selected for.[8] For instance, the state of affairs that constitutes the truth conditions of the vervets' leopard cry is just that state of affairs in which it has been advantageous to run up a tree in response. Given the vervets' history, this is just when there is a leopard close enough to pose a danger. The reason that the cry means "leopard", is that the signaling system is designed to get the cry to covary with the presence of leopards. Notice that in simple signaling systems it is not at all clear whether the signals are indicatives or imperatives. They seem to be a bit of both.[9]

These historically determined (and thus objective) correspondence maps exist for any adapted control system which mediates control via signals. The general architecture is as follows. The rule that applies to an adapted trait **AT** is just a map from conditions to processes — the set of all ordered pairs with a condition in the first place and a process in the second such that **AT** was selected for performing the process in the conditions, or:

$$\mathbf{R}_{\mathbf{AT}} = \{\langle \mathbf{condition}, \mathbf{process} \rangle \mid \mathbf{AT} \text{ was selected for performing } \mathbf{process} \text{ in } \mathbf{condition}\}$$

(The presumption is that the naturalist can simply assume basic notions from evolutionary biology like "selected for", rather than needing to define them. We can define our formalisms over operationally justified primitives of our current best science, obviating the need for the definitions that more foundationalist approaches require.) The rule for a complex mechanism will contain various sorts of individual mappings. In each case, the processes are something that the

[7] The example is borrowed from Skyrms (1996, Ch.5), whose game-theoretic analysis of the evolution of meaning rests on functional architectures similar to Millikan's (1984).

[8] Those familiar with the teleosemantic literature will recognize that I am doing things Millikan's way. Dretske (1986), for instance, attempts to derive satisfaction conditions via the causal pathway responsible for the production of the signal. See Millikan 1990 for a discussion.

[9] Cf. Millikan, "Pushmi-Pullyu representations" (1996) for a discussion of this point.

mechanism can *do*. The conditions are more varied. They may be proximal causes of the specified process, in which the rule describes a causal chain of events. They may be states of the world which obtained when the processes were adaptively performed. Biofunctional correspondence maps involve the latter sort of condition.

The minimal architecture for biofunctional semantics is as follows: There is a signal producing mechanism **P** which issues a set of signals $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. The signals in turn elicit a set of responses $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ from a response mechanism **RM**. From the general rule governing an adapted signaling system we can extract three kinds of sub-rules relevant to the present analysis. The rule governing the response mechanism includes specification of the *interpretation of signals*.

$$\mathbf{R}(\text{interpretation})_{\mathbf{RM}} = \{\langle \mathbf{s},\mathbf{b}\rangle \mid \mathbf{RM} \text{ has been selected for } \mathbf{b}\text{-ing when } \mathbf{s} \text{ is received}\}$$

The rule governing the signal producing mechanism **P** includes a correspondence map from states of the world ($\mathbf{W} = \{\mathbf{w}_1 \ldots \mathbf{w}_m\}$) to signals.

$$\mathbf{R}(\text{correspondence})_{\mathbf{P}} = \{\langle \mathbf{w},\mathbf{s}\rangle \mid \mathbf{P} \text{ was selected for sending } \mathbf{s} \text{ in } \mathbf{w}\}$$

Notice that the correspondence rule is different from the production rule for **P**.

$$\mathbf{R}(\text{production})_{\mathbf{P}} = \{\langle \mathbf{stimulus},\mathbf{S}\rangle \mid \mathbf{P} \text{ was selected for sending } \mathbf{s} \text{ in response to } \mathbf{stimulus}\}$$

Note here that the correspondence, interpretation and production rules serve to specify something like the *extension*, *intension*, and conditions of *justification* of the signals involved. On the current account, these primitive versions of extension, intension, and justification are not normative as they are in human language, but have the same form.

What exactly are the world states **w**? They are not, in the first analysis, specified by the attribution of properties to objects. Rather, if this arrangement has been selected for, then the history of the signaling system induces a set of partitions over states of the world. Consider the simplest case in which each signal is designed to elicit one and only one response. The adaptive history of *each* signal/response pair carves the space of states of the world into two parts — those states in which the response has been adaptive, and those in which it hasn't, the former constituting the truth conditions for the signal. Notice that *each* signal/response pair independently induces a full partitioning of states of the world, whose two cells determine the truth and falsity of the signal involved. Unlike contemporary propositional semantics in which propositions are commonly interpreted as sets of *possible worlds*, the biofunctional extension of signals in adapted signaling systems are sets of states of *this* world which have played a role in the system's adaptive history.

There is no reason, in general, why signals in this kind of system cannot have overlapping truth conditions. It is possible that vervets respond to rustling in the grass by issuing the leopard cry and running up a nearby tree. If sometimes it is a python rather than a leopard that causes the rustling, and tree-climbing has been an effective (though perhaps inefficient) strategy for avoiding pythons in such cases, then the leopard cry may be true also when a python is present so that both python and leopard cries can be true in the same circumstances. One would expect, however, that optimizing selection on perceptual discrimination would tend to reduce this overlap. The result might be an "optimized" signaling system where no two signals are true at the same time.

Summing up again, the evolutionary design process results in rules which apply to adapted mechanisms. Some of these rules specify which states of the world signal types are supposed to correspond to. Such correspondence rules exist for any adapted control system which mediates control via signals.

Representing Rules

The hypothesis we are pursuing is that some of the time, when people worry about rules, they are actually worrying about rules of adapted design. This requires that we are somehow able to represent those rules. Simple signaling systems like the vervets' system of warning cries have correspondence maps which specify states of the world external to the organism, though in the case of the vervets, the signals' truth depends not simply on some "neutral" state of the world (like the presence of a leopard at some location) but crucially, upon the leopard being close enough to constitute a threat. In general, one might expect that the truth conditions for simple signaling systems tend to involve some relationship involving the organism(s). Moreover, there is no reason why external states need to be involved at all. The signal I experience as hunger has a correspondence rule of the sort we have been discussing, but what makes it true is just that fact that my stomach is empty. The variable states of the world outside my skin don't seem to get involved. On the other hand, if adrenalin coursing through my veins means "danger", this is true if I currently stand in some relation (being in danger) to things outside of me. According to the non-propositional semantics of adapted control systems, all it takes for a signal to represent something is the appropriate adaptive history. Adaptation forges semantic links. The question here is, what kinds of histories might result in the representation of rules of adapted design?

One answer is the accumulation of function stabilizing mechanisms. Due to the way that nature goes about providing solutions to adaptive problems, selection and (so-called) random variation, the preliminary model tends to be rather inefficient. But once something that is at least better than nothing is in place, optimization can commence. Small modifications of the system arise via the usual inaccuracies of biological reproduction, and barring accident and given time, those that are superior with respect to the particular function will be selected for. For our purposes, the kinds of modifications that arise fall into two categories. The first is the most familiar. The *existing* structure might be modified, for better or worse. On the other hand, instead of modifying existing structures, *new* mechanisms might arise which improve the performance of existing structures by interacting with them. The common bacterium *e. coli* has, along with an ingenuous system of motorized flagella dedicated to foraging and toxicity avoidance, a collection of chemical sensors. Presumably, some of these sensors have been *added* to the existing system to improve the functioning of the older mobility system. Genes are commonly divided into two categories, *structural* genes which code for proteins and enzymes, and *regulatory* genes which turn them on and off. Again, there is an asymmetrical functional dependency that arises from the addition of regulatory mechanisms to a pre-existing system. The particular relationship we are after here is one where some new regulatory mechanism is selected which *enforces existing function*.

In his classic article "The Evolution of Reciprocal Altruism", Robert Trivers presented a sketch of the system underlying human altruism. The proposed reconstruction of the evolutionary history, based mostly on anthropological studies of tribal peoples and laboratory studies of human moral and cooperative behavior, is as follows. The economics of cooperative behavior are such that cooperation is unstable, as exhibited in the familiar "prisoner's dilemma" of game theory.[10] Nature's initial solution has been to provide "strong positive emotions" favoring cooperation. This may be an adequate solution where one usually plays against close kin, as in Hamilton's (1964) kin selection model. However,

> Once such positive emotions have evolved to motivate altruistic behavior, the altruist is in a vulnerable position because cheaters will be selected to take advantage of the altruists' positive emotions. This in turn sets up a selection pressure for a protective mechanism. Moralistic aggression and indignation in humans was selected for in order
>
> (a) to counteract the tendency of the altruist, in the absence of reciprocity, to continue to perform altruistic acts for his own emotional rewards;
> (b) to educate the unreciprocating individual by frightening him with immediate harm or with the future harm of no more aid; and
> (c) in extreme cases, perhaps, to select directly against the unreciprocating individual by injuring, killing, or exiling him. (Trivers 1971,49)

Trivers goes on to suggest that a sort of "arms race" can ensue between cheating and the detection of cheating.

> Sham moralistic aggression when no real cheating has occurred may nevertheless induce reparative altruism. Sham guilt may convince a wronged friend that one has reformed one's ways even when the cheating is about to resume. (Trivers 1971,50)

Such innovative deceptions involving enforcement mechanisms create selective pressures for new detection and enforcement mechanisms, whose discriminating responses can then be exploited, and so on. The result is a hierarchical system of controls in which patterns of enforcement are themselves enforced, resulting in many levels of regulatory hierarchy. Notice that, in principle, all this regulatory complexity can arise through the evolution of "instinctive" behaviors.

Whether or not Trivers has the details of the adaptive history of human altruism exactly

[10] See especially Axelrod 1984.

right is tangential to the point being made here, since it is evident that human social norms have the kind of hierarchical regulatory structure described, and that structure is universal enough to make an adaptive history of accumulating regulatory mechanisms plausible. Our question concerns the correspondence rules of the enforcement mechanisms. What makes cheater identifications true, and what if anything is the difference between the associated correspondence rules and those of more basic signaling systems?

Say that a mechanism of “moralistic aggression” arises and is selected for in order to compensate for cheating (non-reciprocating behavior). Such a mechanism is another example of an adapted signaling system, and so there will be production rules governing the issuance of the cheater recognition signal, and interpretive rules specifying the appropriate response according to the design process. But as before, the correspondence rules for the signals are separate from (though complementary to) the rules governing production and interpretation. The correspondence rule specifies the state in which cheater detection has been advantageous. If the enforcement mechanism has been selected specifically in order to eliminate the compromises to the design of the cooperative system posed by cheaters, then the cheater identification signal is true just in case the rule governing the operation of the system of cooperation has been violated. What the signal must correspond *to* in order to be true, is the *failure of a rule* of adapted design.

Our simple formalization of biofunctional semantics can easily be extended to the representation of rules of adapted design. Let us say that a rule of design *fails* when one of the conditions specified by the rule is not accompanied by the indicated process.

$$\text{Failure of } \mathbf{R}_{\mathbf{M}} : (\mathbf{condition} \;\&\; \neg\mathbf{process}) \text{ where } \langle \mathbf{condition,process} \rangle \in \mathbf{R}_{\mathbf{M}}\,.$$

For some stabilizing mechanism (**SM**) the correspondence rule for its corrective signals $\mathbf{CS} = \{\mathbf{cs}_1, \ldots, \mathbf{cs}_n\}$ are given by

$$\mathbf{R}(\text{correspondence})_{\mathbf{SM}} = \{\langle \mathbf{w}\&\mathbf{m},\ \mathbf{cs} \rangle \mid \mathbf{SM} \text{ was selected for sending } \mathbf{cs} \text{ when } \mathbf{w}\&\mathbf{m}\}$$

where the **m** are individual states of the stabilized mechanism. (This includes the processes and some of the conditions in the general specification of the rule for **SM**.) Since **SM** was selected for stabilizing **M** then the states of the joint system **W+M** in which **SM** was selected for sending signals are just conditions in which some component of **M**’s rule $\mathbf{R_M}$ was violated. Which is to say,

$$\mathbf{R}(\text{correspondence})_{\mathbf{SM}} = \{\langle \textbf{condition}\,\&\; \neg\textbf{process} \text{ where } \langle\textbf{condition,process}\rangle \in \mathbf{R_M}, \textbf{cs}\rangle \mid \textbf{SM} \text{ was selected for sending } \textbf{cs} \text{ when } (\textbf{condition} \;\&\; \neg\textbf{process})\}.$$

Again, the corrective signal is true when elements of the rule implicit in **M**’s history have failed.

What has puzzled philosophers about norms is what kinds of properties they must attribute in order to be true, and why it is that the attribution of those properties has normative force. We are now in a position to say something about this puzzle. In the first place it is not the attribution of properties that makes the corrective signal true, but the failure of a particular historically established relation between conditions and processes. The corrective signal is true when the relation that it has evolved to enforce has failed. But this is not the deep part of the puzzle.

Suppose that some authoritative normative impulse is in fact a corrective signal in a function enforcing mechanism as above, and that it is truly issued. I claim to have provided a schema that allows us to state the satisfaction conditions for the corrective signal in descriptive terms. Which is to say, we can create descriptions which are true if-and-only-if the corrective signal is true. Why isn’t the description normative in the same way as the corrective signal? The answer is that there is nothing in the represented relation that is in itself normative. The relation is simply a rule of adapted design. The normativity of the rule-component derives from its role in the truth conditions of the enforcement system. The reason that the corrective signal confers normativity is not that it truly represents an independently normative relation (or its failure), but that the adapted role of the corrective signal, from which it derives its truth conditions, is to constrain adherence to the rule. To put this in more conventional terms, the descriptive theory, if successful, can *duplicate the extension* of the normative signal. What it fails to do is duplicate the *intension* (the normative interpretation) of that signal, though that too can be specified descriptively. The phenomenon is not unlike that where substitution fails in intensional contexts (e.g. substituting "evening star" for "morning star" in the mind of an individual who doesn't realize that the terms are coextensive). It is in the failure to duplicate intensions that descriptive accounts of normativity lack normative authority, rather than any unavoidable incompleteness in the descriptions themselves.

The discussion of realism with respect to the referents of normative utterances often centers around the claim that moral "properties" supervene on physical properties. Horgan and Timmons (1992) argue that while it is certainly possible that such supervenience relationships exist, the realist must do more than simply defend the possibility of such relations. One must *explain* what determines which supervenience relation holds between truth-makers for normative and physicalistic language. The approach taken in this paper does not rely on the notion of supervenience to relate moral and physicalistic truth-makers, principally because the reification of moral truth-makers as moral *properties* seems unhelpful. But it may help clarify the relation of the present proposal to current alternatives to rephrase it in terms of the (to some) more familiar supervenience relations.

The most important characteristic of the supervenience relation is that the higher level property supervenes on a *disjunction* of physical base properties. That is, the collection of different physical states that are sufficient for the instantiation of the higher level property may form a rather motley collection of various sorts of physical states with *ad hoc* restrictions and the collection may be governed by a non-systematic array of physical laws. The challenge is to offer a systematic way of saying just what it is that collects the disjuncts together in the set that forms the subvenient base of the higher level property. In the present case, it is quite clear that the portion of the rule $\mathbf{R_M}$ whose failure (the various [**condition** & ¬**process**] configurations) forms the correspondence conditions for some corrective signal **cs** is in fact just such a disjunctive set as supervenience relations are invoked to accommodate. The corrective signal has as many distinct truth conditions as it has been selected for co-occurring with. More generally, the truth conditions for signals in adapted signaling can be expected to be disjunctive in physical terms, since the set of correspondence conditions (extension) for a given signal {**w** | **P** was selected for sending **s** in **w**} is determined by the historical efficacy of the adapted response to the signal, *not* by whether or not members of the set form a proper natural kind. Consequently, the biosemantic approach may be unique in its ability to create objective disjunctive truth conditions for normative utterances. And if one can't resist the temptation to reify them into special sorts of disjunctive supervening properties, then the adaptive history of signaling systems provides a systematic factual basis for specifying the supervenience relations.

A normative system on this account is a function enforcement mechanism — an adapted signaling system in a regulatory hierarchy whose correspondence rules specify the failure of some design rule of the system it regulates. On the present account, the normativity of the system does not derive from some further reason why we ought to care about function enforcement mechanisms. Looking to the semantics of the very "oughts" in question, we provide an analysis of normativity itself, rather than deriving it from some further evaluation. Caring about rules simply *is* the operation of function enforcement mechanisms. Thus, as emotivism had it, normative utterances are the expression of more inarticulate feelings or intuitions. Where the emotivist went wrong was in concluding that this deprived them of truth conditions and reference to externally given rules.

Normative Interpretation of Signals

The main focus of the semantic analysis in this paper is on reference relations (extension) in adapted signaling systems. The reason for this focus is that the reference of oughts has been the primary mystery for the analysis of norms. However, the resources deployed so far allow a complementary analysis of the other part of semantics — the normative interpretation of signals (intension). Note again that these are the "oughts" that are missing from descriptive accounts of normativity.

Recall that among the rules of adapted design for a signaling system are rules of interpretation. For signals **S**, behaviors **B**, and a response mechanism **RM**, the interpretive rule was given by,

$$\mathbf{R}(\text{interpretation})_{\mathbf{RM}} = \{\langle \mathbf{s},\mathbf{b} \rangle \mid \mathbf{RM} \text{ has been selected for } \mathbf{b}\text{-ing when } \mathbf{s} \text{ is received}\}.$$

If **RM** acquires a function stabilizing mechanism, then on the present account, elements of

R(interpretation)$_{\mathbf{RM}}$ whose failures constitute the truth conditions for the stabilizing mechanism's corrective signals become normative. This gives us what one might call *normative practical implication*, which is central to normative motivation.

In the case of simple signaling systems like the vervet's warning system, the normative interpretation of the signal is a standard evasive behavior. In more complex signaling systems (like reason) which are selected for performing operations on signals, the normative interpretation may be another signal. Which is to say, for signal types **P** (for "premise") and **C** (for "conclusion"), the rule governing an inference mechanism **IM** would be,

$$\mathbf{R}(\text{interpretation})_{\mathbf{IM}} = \{\langle \mathbf{p},\mathbf{c}\rangle \mid \mathbf{IM} \text{ was selected for sending } \mathbf{c} \text{ on receipt of } \mathbf{p}\}.$$

If **IM** acquires a function stabilizing mechanism, then as before, elements of **R**(interpretation)$_{\mathbf{IM}}$ whose failures form the truth conditions of the stabilizing mechanism's corrective signals become normative. If mechanisms exist to enforce elements of **R**(interpretation)$_{\mathbf{IM}}$ which have the form of familiar normative inferences (non-contradiction, modus tollens, etc.) then the so-called laws of reason are grounded in the adaptive history of reason's endogenous normative systems. This gives us *normative inference rules*, naturalistic grounds for the "laws" of logic.

Normative Systems and Culture

One of the most remarkable things about human behavior is how variable it is under cultural influence. How does one go about integrating this remarkable fact into evolutionary stories which seem to proceed as though all behavior were instinctive? The answer is actually quite simple. The systems that regulate human behavior are designed to be flexible. They are designed to be able to accommodate environmental novelties. They are designed to allow the transmission of information between conspecifics via language. They are designed to allow the formation and adoption of rules of social behavior. And they are designed to provide for the enforcement of rules so formulated and adopted. Doesn't this flexibility threaten to break the systems involved loose from their adaptive histories, and thus from the correspondence rules those histories provide? Not necessarily.

Consider that most arbitrary of social rules, the traffic convention. In the U.S., we drive on the right. In England, they drive on the left. And, it seems, there is something wrong with not following those conventions when you are in those places. Two questions arise: what exactly is it that is wrong with driving on the wrong side of the street? And, aren't these standards just the arbitrary dictates of culture?

There seem to be at least three different reasons why driving on the wrong side of the street is wrong. First, it's stupid. Second, you pose a danger to others. Third, there is a convention (ensconced in law) which *says* that it's wrong to drive on that side. In all three cases it is possible, given the appropriate adaptive history, for the judgement to be semantically grounded in that history. In the first case, if humans are equipped with a normative system which rides herd on flexible instrumental behavior and corrects it in cases where the function of instrumental behavior is compromised, and if "is stupid" is just a linguistic proxy for the correcting signal in that normative system, then it *is* stupid to drive on the wrong side of the street just in case the appropriate part of the function of instrumental behavior is violated. In the second case, if humans are in fact equipped with a normative system whose function is to minimize the danger we pose to others, and if "is wrong" is a linguistic proxy for the correcting signal in that system, then it is wrong to drive on the wrong side of the street just in case we are posing a danger to others. Finally, if humans are equipped with the abilities to formulate and follow conventions, and there is a normative system in place whose function is to enforce conventions so adopted, and "is wrong" is a linguistic proxy for the correcting signal in that system, then it is wrong to violate the convention just in case the system of convention following that the normative signal is designed to regulate is in fact not functioning according to design. If all three of these hypotheses seem plausible, then perhaps it is objectively wrong to drive on the wrong side of the street in three different *senses* (according to 3 different semantic mappings). The nice thing about traffic conventions (and one of the reasons they are so stable) is that the three sets of norms seem to agree. This is not always the case, however.

As for the second question, which side we drive on is *of course* an arbitrary dictate of culture, but it is not *just* that. It is an arbitrary dictate of culture that plays a small though decisive role in governing the behavior of an immensely complicated system of behavioral controls. It is an arbitrary dictate of culture that may do a good or bad job of regulating that system from the point of view of the system's design. And *if* there is a normative system in place whose function is to evaluate arbitrary dictates of culture *vis a vis* their efficacy in contributing to the function of the systems they regulate, then accordingly, some rules are objectively, truly, bad, and others good, at least according to the semantics of that normative enforcement system. But once again, rules of adapted designs are not normative in general. The proposal is that true attributions of wrongness are the expression of normative systems whose function it is to enforce those designs.

The Normative Role of Reason

The models that have been developed so far suggest a picture in which human minds consist of a (possibly large) number of independent control structures. Each proximal behavioral control has one or many distinctive functions described by its rules of adapted design. Each signaling/control system has a distinctive and *prima facie* untranslatable correspondence semantics. Each level of normative stabilizing mechanisms issues a *sui generis* species of imperatives, which while they may possess objective correspondence maps, do not intrude and compete in consciousness in the way that human norms seem to. So even if I have demonstrated that there is something out there for us to be worrying about when we worry about rules, and that it is perfectly natural for us to be worrying *about* them, and that it is not so hard to understand what it would take for it to be simply *wrong* for us to violate them, still, most philosophers will feel that something important has been left out of the picture. Namely, the way in which instrumental reason plays a central and ineliminable role in normative deliberation. For instance, it seems that much of moral reasoning involves not the stabilization of ancestral responses or even the optimization of those responses to local environments, but crucially, the creative use of reason in generating new strategies to pursue moral ends. Moreover, it seems that reason has distinctively *moral* authority in this role. How is this possible?

A change of strategy is called for. So far I have been talking about general purpose structures and arguing that at least some of the time when we worry about rules we may in fact be expressing those structures and their attendant semantics. Here, the problem concerns a particular system and one is forced to speculate a bit more. So here is an evolutionary fable:

****

Suppose that humanity's distant ancestors were in fact characterized by a collection of isolated behavioral control systems. One regulated social behavior. One regulated foraging. One regulated reproduction. One regulated the seeking of shelter and other thermally important tasks. And so forth. All of these systems were implemented in the same brains, but had somehow managed to remain functionally isolated from one another. At some point, the community was forced to move into a new environment in which there was much greater variability in available kinds of food. Consequently, the foraging system began to evolve for greater flexibility. Neural pathways into the other systems were forged which allowed the foraging system to make use of perceptual discrimination developed for other tasks. The new wealth of sensory input made possible associative learning and deliberate behavior. Strategies for opening nuts worked well also for shellfish, and so forth. One unexpected consequence of the foraging system's raid on information resources is that it started to accept imperatives of the raided systems as motivating. Just as behavioral creativity and associative learning had been directed toward bringing about the cessation of the hunger signal, it began to be directed to bringing about the cessation of moral outrage, cold, reproductive longing, and so forth. Individuals with these aberrant connections ended up doing much better than their conspecifics and took over. In fact, the new trial and error learning system was so successful in its new more general role that eventually it no longer had any distinctive connection to foraging. It's newly stabilized purpose was to satisfy the demands of the various dedicated systems which it recognized as "desires". In time, Reason (as it had come to refer to itself) developed an elaborate general purpose representational system in which "properties" were attributed to "objects". This distinctive "propositional" system of representations was so powerful that Reason began using it to guess at the ends of the various subsystems it served to satisfy. Success at these tasks led to the accumulation of stabilizing mechanisms distinctive to Reason, which Reason conceived of as "eternal laws" that governed Reason wherever it might occur. And indeed, Reason came to think that rather than being a general purpose problem solver with no ends of its own, it was the real reason for the existence of the system. Desires became mere non-rational "attitudes" or preferences toward propositional states, mere animal impulses with no more normative authority than, well, animal impulses.

Reason came to find this rather puzzling...

****

The evolutionary fable provides a model with enough structure to capture a number of interesting features of reason and its role with respect to more dedicated normative systems. (1) In the first place, the propositional form of representation in which properties are attributed to objects is only one among many semi-independent representational systems occurring in the mind. There are, if you like, many "languages of thought," and many kinds of meaning. Truth is more general than propositional content. The flexibility and specificity of propositional representation was selected for its ability to support associative learning and the forward-looking matching of means to ends necessary for a general purpose problem-solving system. (2) Desires are understood as the demands of dedicated systems whose ends reason serves. The very separation of belief from motivating desire is not a basic feature of biological representation, but is a rather peculiar feature of this general purpose system's relationship to the various needs of the organism. (3) The *endogenous* norms of reason, which rational choice theory codifies into laws of logic and rules of consistency in the ordering of preferences, are aspects of normative systems which enforce the rules of design implicit in the adaptive history of reason. They are emphatically *not* axioms for ideal systems, but rather the relations whose failure constitutes the truth conditions for corrective signals which oversee the operation of systems of belief and desire-satisfying action. It is here that the other part of semantics — normative implication — finds its place. Modus ponens is a rule of adapted design. Its normativity derives from the fact that there are enforcement mechanisms which correct deviations. (4) Reason was selected for the satisfaction of some desires under some circumstances. If there are normative systems which enforce these arrangements, then it may be the case that there are some desires reason objectively *ought* to satisfy, or conditions under which it ought not. (5) Reason may do more than attempt to bring about the cessation of occurrent desires like hunger. It may anticipate the occurrence of desires, and it may attempt to identify the ends toward which desires are directed. Given reason's preferred representational framework, this will involved specifying those ends in propositional/indicative terms. The corrective signals of the dedicated normative systems appear as *intuitions* regarding the proper end of desires. The fact that the authority of corrective signals is tied up with non-propositional correspondence rules accounts for difficulty of incorporating oughts into the descriptive framework. Being parts of different systems they have different jobs and thus different normative implications.

This is all very speculative and sketchy. Moreover, much of it is probably false. But it may be that there is a time for stories like this. It was once said that there was no way that mere natural processes could explain the design of the human eye. What we have found is that, if anything, the opposite is the case. Such things are far too easy to explain via selection and variation, and we need to hold ourselves to strict standards of evidential support when actually claiming that some particular evolutionary hypothesis is true. The same is true of the adaptive history of normative systems. Given the wide availability of rules for adapted systems, it is far too easy to invent merely plausible stories that explain why complex human agency has the character it does, and would if true justify much of our ordinary confidence in the existence and objectivity of the rules we try to follow. In time, the study of normative systems will turn from what could be true to what is in fact true. But when common opinion is that there can be no such story, evolutionary fables like the one above have their place.

The Ethicist's Wish-List

I have argued that there is after all something out there for us to be worrying about when we worry about rules. The "furniture of the world" unproblematically includes the rules of adapted design. If the normative systems I have described underlie human normative intuitions, then the failures of these rules form correspondence conditions for expressions of those intuitions, yielding support for realism about normative standards. Despite the fact that it is a descriptive rather than normative theory that is being proposed, it is instructive to see how far the current proposal can go toward satisfying the concerns voiced by moral theorists.

Semantics: Not to belabor the point, but correspondence truth conditions are easy to come by for adapted signaling systems. This covers the case of morality and value judgements as well if the "impulses" expressed by normative utterances are signals in such systems. Emotivists were right in saying that normative utterances are the expressions of sentiments. Where they were wrong was in inferring that such sentiments have no objective meaning, no reference or truth.

Queerness: The primary impediment to the naturalization of normativity is the widespread belief that if normative utterances are true by correspondence, they must correspond to some exceedingly queer properties. The rules of adapted design, on the other hand, are as ordinary as one could wish. They are first and foremost natural historical relations between states and processes, and between signals and states. They are identified and discussed by scientists as an unremarkable matter of course. There is nothing any more queer about these rules than there is about kinship. Fatherhood is not a property, but a relation, and a relation just as contingent on past history as the rules of design. Normative authority comes not from the relationship represented, but from the role of the representation itself in the economy of function/rule enforcement mechanisms.

Epistemology: There is no mystery as to how it is that we “know” about such relations, even when we are unable to identify them rationally. Each normative system includes endogenous perceptual systems with their production rules (justification) and semantics (truth and implication). Just as there are many languages of thought, there are many kinds of knowledge. Of course, this does not yield certainty, but reliably true and properly formed normative signals.

Reason and Desire: Reason, if it is understood along the lines of the above evolutionary fable, has a different system of representations than the special purpose systems whose ends it serves. This explains the familiar *discontinuity* between factual meaning and normative meaning. It also explains Hume’s dictum that reason is the slave of the passions, as well as the difficulty in deciding whether or not desires are propositional attitudes. On the present account, the primary semantics of desires is non-propositional, but reason represents the ends of desires propositionally as part of a strategy for inventing novel strategies for attaining those ends. The endogenous norms of reason specify strategies for matching means to its proper ends — those of the dedicated systems reason has been optimized to serve.

Cultural Conventions: Again, cultural variability can be understood as designed flexibility. Think of culturally variable “codes” as environmentally variable “representations” tailored to the purpose of the interpreting systems. Whatever can vary, can serve a purpose in that variation. If the appropriate stabilizing mechanisms are present, there can be norms that determine the appropriateness of culturally variable codes *in situ*.

Action-Guidingness: It is often said that norms are intrinsically action-guiding, in the way that factual beliefs are not. If we understand factual belief as part of a representational system which serves many ends, and dedicated normative systems as representational systems which serve particular ends, then it is not at all surprising that signals in the latter should have a more direct connection to action than the former. This is to say, the signals are essentially *motivating* in a way that factual beliefs are not.

A Special Sense: Intuitionists are famous for claiming that we perceive moral truth via a "special sense" separate from that via which we perceive ordinary facts. They may be right. Each normative system includes a special sense (the stabilizing mechanism) which produces the normative corrective signals. The meaning of norms becomes muddled once reason attempts to mirror the semantics of the normative systems it serves and characterizes their ends propositionally. But the intuitions which tell us whether reason has gotten it right may be "special" in much the way the intutionist supposes.

Disjunction: The correspondence conditions of signals in adapted systems are typically disjunctive in terms of physical properties. Thus, the challenge to specify the disjunctive relationship between normative and factual truth-makers is met at the most basic level. If anything, the challenge becomes explaining how a signaling system could acquire non-disjunctive referents. (We apparently believe that scientific language has this property.) Presumably the story has something to do with the optimization of a general purpose forward looking problem solving faculty.

In sum, rather than the various features of normative deliberation being mysterious, they are more or less what one would expect from a complex system like the one sketched above. But again, the truth of the proposed hypothesis about human adaptive history and cognitive architecture is an empirical matter, whose truth must be supported in appropriate ways.

Conclusion

I began by suggesting that traditional empiricism's focus on proximal causal relations has been primarily responsible for the difficulty that philosophers and scientists have had understanding norms. The above sketch of a theory of normative systems should make clear the immense resources that evolutionary biology adds to the naturalist's toolkit. Adaptive histories are not limited to simply *explaining* how we got to be the way we are. They may go further, and corroborate many of the intuitions about normativity which have seemed so threatened by traditional naturalism. Adaptive histories provide subject matter for our worries about rules and forge semantic links between normative sentiments and the failure of those rules. In short, *evolution matters,* at least to the descriptive understanding of human normativity. The biosemantic approach to norms results in descriptive stories which have a *lot* of features that philosophers have thought could not be part of the natural world.

These are philosopher's criteria, however. An empirical hypothesis must do more than be defensible against objections. A hypothesis which proposes a research framework like the present one must promise a fruitful approach to the study of its subject matter as well. Several features deserve note: the central concepts of adaptation and regulatory architecture are uncontroversial core concepts in current biology, though they are not often studied in conjunction.[11] The architecture of normative systems is simple enough to support a consistent interpretation across the scientific community. Moreover, no more is needed to pursue the investigation of normative systems than the integration of existing research techniques. Currently, scientists do not identify correspondence rules in signaling systems, though such systems are studied widely in all areas of the biological and social sciences, presumably because there is no paradigm which demonstrates why those relationships are of interest. Nonetheless, current research *does* provide a variety of evidence concerning the design of normative systems, and it is reasonable to suppose that research focused on that question would provide more compelling evidence. At any rate, there is no theoretical impediment to the descriptive study of normative systems throughout the biological and social sciences.

[11] Developmental biologist Rudolph Raff (1996) has urged that the integration of developmental and evolutionary biology is overdue, and promises great rewards.

References

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

Bradie, M. (1994): *The Secret Chain: Evolution and Ethics*, Albany: SUNY.

Cheney, D. L. and R. M. Seyfarth (1990): *How Monkeys See the World*, Chicago: University of Chicago Press.

Darwall, S., A. Gibbard and P. Railton (1992): "Toward Fin de Siecle Ethics: Some Trends," *The Philosophical Review*, 101(1), Januaru, 115-189.

Dawkins, R. (1976): *The Selfish Gene*, New York: Oxford University Press.

Dretske, F. (1986): "Misinformation," in *Belief: Form, Content and Function*, ed. by R. <. Bogdan, Oxford: Clarendon Press.

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

Gould, S. J. and R. C. Lewontin (1978): "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Program," *Proceedings of the Royal Society of London*, 205, 581-598.

Hamilton, W. D. (1964): "The Genetical Evolution of Social Behavior I & II," *Journal of Theoretical Biology*, 7, 1-52.

Harms, W. F. (1998): "The Use of Information Theory in Epistemology," *Philosophy of Science*, 65(3), 472-501.

Horgan, T. and M. Timmons (1992): "Troubles on Moral Twin Earth: Moral Queerness Revisited," *Synthese*, 92, 221-260.

Kitcher, P. (1994): "Four Ways of "Biologicizing" Ethics," in *Conceptual Issues in Evolutionary Biology*, ed. by E. Sober, MIT Press, 439-450.

Levin, M. (1984): "Why Homosexuality is Abnormal," *The Monist*.

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

Millikan, R. G. (1984): *Language, Thought, and Other Biological Categories: new foundations for realism*.,A Bradford Book, Cambridge, MA: MIT Press.

------ (1990): "Compare and Contrast Dretske, Fodor, and Millikan on Teleosemantics," *Philosophical Topics*, 18(2), 151-161.

------ (1993): *White Queen Psychology and Other Essays for Alice*.,A Bradford Book, Cambridge, MA: MIT Press.

------ (1996): "Pushmi-Pullyu Representations," in *Mind and Morals: essays on cognitive science and ethics*, ed. by L. May, M. Friedman and Clark A., Cambridge, MA: MIT Press, 145-161.

Raff, R. A. (1996): *The Shape of Life: genes, development, and the evolution of animal form*, The University of Chicago Press.

Ruse, M. and E. O. Wilson (1986): "Moral Philosophy as Applied Science," *Philosophy*, 61, 173-192.

Skyrms (1998): "Evolution of Inference," *Manuscript*.

------ (1996): *Evolution of the Social Contract*, Cambridge University Press.

Sober, E. (1984): *The Nature of Selection: evolutionary theory in philosophical focus*, Cambridge, MA: MIT Press (368 pp).

------ (1993): *Philosophy of Biology*, Westview Press.

Trivers, R. L. (1971): "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology*, 46(3), 35-57.

regularities in normative convictions, which is, come to think of it, just the sort of thing that philosophers are good at.