ON THE RELATIONSHIP BETWEEN SCIENCE AND ETHICS

by Massimo Pigliucci

Abstract. The relationship between ethics and science has been discussed within the framework of continuity versus discontinuity theories, each of which can take several forms. Continuity theorists claim that ethics is a science or at least that it has deep similarities with the modus operandi of science. Discontinuity theorists reject such equivalency, while at the same time many of them claim that ethics does deal with objective truths and universalizable statements, just not in the same sense as science does. I propose here a third view of *quasi-continuity* (or, equivalently, *quasi-discontinuity*) that integrates ethics and science as equal partners toward the uncovering of new knowledge. In this third way, a program envisioned by William James but made practicable only by contemporary scientific advancement, science can and must inform ethics at a deep level, and ethical theory—while going beyond science—cannot do without it. In particular, I identify four areas of ethics-science collaboration: neurobiological research into the basis of moral judgment, comparative anthropology, comparative evolutionary biology of primates, and game-theoretical modeling. I provide examples within each of these fields to show how they link to ethical theories (including prescriptive work) and questions. The essay concludes with a brief discussion of the light that a scientifically informed ethics can shed on some classical problems in moral theory, such as the relationships between rationality and selfishness, egoism and altruism, as well as the concept of social contract. A joint research program involving both philosophers and scientists is called for if we wish to move ethical theory into the twenty-first century.

Keywords: continuity theory; ethics; evolutionary biology; game theory; neuroscience.

Massimo Pigliucci is Associate Professor of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1100; e-mail pigliucci@utk.edu. He is associated with the Graduate Program in Philosophy at the same university. This work was funded in part by National Science Foundation grant IBN 9707552.

[*Zygon*, vol. 38, no. 4 (December 2003).]
 ISSN 0591-2385

Is Ethics a Science?

A recurring question in ethics concerns its relationship with science. Is ethics in any reasonable sense a science, or does it at least share some fundamental communality with the sciences? If not, do we have any hope of engaging in a quest for objectivity and universality concerning moral discourse? This debate has a long and complex history on both sides, a history that includes David Hume's famous articulation of the "naturalistic fallacy" ([1739–40] 1978), G. E. Moore's more recent discussion of the same issue (1903), and of course William James's program (1891; 1918) for a dialogue between moral philosophy and moral psychology. I do not attempt an in-depth commentary of these and other relevant historical contributions here, but a brief summary of some of the main entries may help set the stage for an examination of the contemporary views discussed in this essay and for my own contribution to the current debate.

Hume's version of the naturalistic fallacy, that one cannot automatically derive an *ought* from an *is*, was famously summarized thus:

> In every system of morality, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning . . . when of a sudden I am surpris'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought* or an *ought not*. This change is imperceptible, but it is, however, of the last consequence. For as this *ought* or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. (Hume [1739–40] 1978, *Treatise* III[i]1)

Hume was criticizing the rationalist tradition, which he saw as replete with unjustified *is* –> *ought* transitions. While Hume's criticism does appear to severely limit the relevance of the natural sciences for the solution of moral questions, one has to note two important things. First, Hume did not, and never claimed to, *demonstrate* that one cannot make a connection between natural facts and moral imperatives. He simply and reasonably stated that, should one wish to make such connection, one had to be willing to provide an argument for the acceptance of this kind of move. Second, Hume is also famous for having stated that "reason is, and ought only to be, the slave of passions" (*Treatise* III[iii]3), by which one could reasonably derive that our actions—including moral decisions—find their roots in our nature as biological beings—surely a deep connection between the natural sciences and moral philosophy, if ever there was one.

James articulated his view of the connection between science (psychology in particular) and ethical theory in his essay "The Moral Philosopher and the Moral Life" (1891) and, within a broader framework, in *The Principles of Psychology* (1918). James's argument for a tight connection between science and ethics rested on the idea that what we call *values* are

rooted in our psychology. Of particular interest here is the suggestion that our instincts (ultimately, the pains and pleasures we feel as animals—see Hume) provide the basis for our perceived "higher" morality as well as the foundations for objectivity in ethics. That is why ethics, according to James, is in fact an empirical science, and moral theory cannot be developed independently of the actual structure of the universe out there.

This sort of reasoning directly leads to one brand of ethical realism, based on a naturalistic interpretation of ethics. As is well known, however, one can be an ethical realist in the completely different fashion of Moore, who put forth a more obscure and less witty version of the naturalistic fallacy than Hume's. In discussing the definition of *good*, he says:

> It may be true that all things which are good are *also* something else, just as it is true that all things which are yellow produce a certain kind of vibration in the light. And it is a fact, that Ethics aims at discovering what are those other properties belonging to all things which are good. But far too many philosophers have thought that when they named those other properties they were actually defining good. . . . This view I propose to call the "naturalistic fallacy." (quoted in Darwall, Gibbard, and Railton 1997a, 54)

In other words, values are facts, but their nature is entirely different from that of the objects of study of the natural sciences, and therefore the study of the latter does not help the moral philosopher, even if he thinks that values are "facts" in some (rather unclear) sense of the word.

Finally, one arrives at more modern authors such as J. L. Mackie ([1977] 1997) and his "argument from queerness," which leads to the rejection of there being moral "facts" at all and, therefore, an examination of the world cannot possibly bring us to the discovery of any moral value inherent in it. The stage is then set for so-called existentialist ethics and the idea that we have radical freedom of choice because of the un-factlike nature of values.

Getting to the contemporary ethics literature, we find two basic positions concerning the ethics-science relationship, *continuity* and *discontinuity*, each of which is then subdivided into various schools of thought. Simply put, the continuity position maintains that ethics is indeed a science or at least shares some of the attributes of the natural sciences. The three general kinds of continuity theory recognized and discussed by Stephen Darwall, Allan Gibbard, and Peter Railton (1997b) are neo-Aristotelianism, postpositivist neoreductionism, and reductionism. The major branches of discontinuity theory—the idea that science and ethics are fundamentally different— summarized by the same authors include practical reasoning theories, constructivism, noncognitivism, and sensibility theories. A full discussion of these is beyond the scope of this essay, but I look briefly at each to extract what they have in common. I then present an outline of what I perceive as a third way—here labeled *quasi-continuity* (it could with equal validity be called *quasi-discontinuity*)—and expound it by means of examples. I conclude with some reflections on where a joint science-ethics

research program, as embedded in a quasi-continuous theory of the relationship between the two disciplines, might lead in the foreseeable future.

Two more preliminary points introduce our discussion. First, it is obvious that the other side of the ethics-as-science debate hinges on the answer to another fundamental question: What is science? For the purpose of this essay I refer the reader to the ample literature on the latter subject (see Popper 1968; Kuhn 1970; Lakatos 1977; Kitcher 1995; Cleland 2001). In what follows, I consider science to be a quest for approximating truths about the world by employing a combination of observational and experimental methods used in hypothesis testing.[1]

Second, regardless of one's position concerning the ethics-science continuum, it is clear that both science and ethics have made progress over the centuries, albeit in a different fashion, at different rates, and certainly not without setbacks and dead ends. The case for progress in science is clear despite the ranting of some extreme postmodernist thinkers (see Hacking 1999 for a wonderful discussion of this topic). There is little question that the Copernican theory is a better rendition of the world than its geocentric predecessor, and the same goes for the theory of relativity when compared with Newtonian mechanics. The case might be thought to be more difficult to make for ethics, and yet there is currently general agreement in the West about certain answers to moral issues, a consensus that was not present as little as fifty years ago. There are, of course, societies that do not share such a moral outline (the former Taliban in Afghanistan, for example); however, at the risk of being considered a cultural chauvinist, that does not mean that the Western consensus does not represent true progress in ethics, measurable by such admittedly fuzzy concepts as the degree of human flourishing in modern societies. The situation is not that different from the one concerning science, given that the Taliban also do not accept contemporary science. As another example, many fundamentalist Christians not only do not share the Western consensus on ethics but also reject evolutionary theory, even though we know that evolutionary biology is on the right track (Pigliucci 2002). I suggest in what follows that part of this progress in ethics can be accounted for by a better understanding of the nature of human beings and societies, as uncovered by science.

Continuity: The Basics

Let us consider the first form of continuity theory, neo-Aristotelianism. A leading exponent is Philippa Foot (1972), who sees moral rules as not different in kind from etiquette, in the sense of being more or less arbitrary within a particular society. What makes ethics more important than etiquette is that for us the promotion of human well-being is of paramount significance, while following conventions is much less so. This view is continuous with science in that, although it rejects the idea of cosmic good, it sees the good as dependent on the psychology of human interests and

desires, which is the result of biological evolution in a social environment. (James would be comfortable with such a perspective.) As we shall see, both some continuity and some discontinuity schools appeal to evolutionary biology as their underpinning. I argue that it is this reasonable and widely acknowledged link with biology that makes a quasi-continuous approach all the more palatable.

The second kind of continuity, postpositivist neoreductionism, asserts that reflective equilibrium is a method shared by ethics and the natural sciences. Both evaluate beliefs (or theories) by appealing to evidence as well as to currently held theories. An example of this can be found in the writings of Richard Boyd (1988), who suggests that moral properties may be irreducible and yet explanatorily efficacious. Boyd uses analogies with the natural sciences, claiming for example that some natural kinds in chemistry or biology, such as acid or gene, are not reducible to natural kinds of physics and yet do not pose a problem for chemists or biologists. However, it seems likely that there is a disanalogy here in that natural kinds in all the sciences are at least potentially reducible to natural kinds of physics simply because that is what they are made of (all chemicals, biomolecules, etc. are made of atoms and subatomic particles).[2] Furthermore, when such reduction is not obvious, it is because of emergent properties of complex systems that are amenable to scientific investigation through nonlinear mathematics (Perry 1995). It is very difficult to envisage something analogous to this within ethics.

The third major type of continuity theory is reductionism. Of this there are several varieties, some of which lead to relativism, others that do not. Railton (1986), for example, maintains that moral judgment can have truth value in a nonepistemic way, that moral truths are objective while being relational, that the moral properties supervene on natural properties (and could be reducible to these), and that moral and empirical inquiry are analogous. Railton goes on to suggest that, while we know quite a bit more about morality than we used to, there is plenty of reason to think that currently held moral theories can be wrong in fundamental ways, that an agent may have no reason to obey a moral imperative, and that in general it is unlikely that a single set of moral rules can be appropriately applied to all societies and all individuals. Despite Railton's claims, a scientist might be justified in feeling at a loss in identifying the similarities between such ethical theorizing and the scientific enterprise, given that Railton talks about *nonepistemic* truths. Indeed, this uneasiness is a recurring feeling whenever continuity theories are discussed from the point of view of science.

Other authors have addressed the parallel between ethics and science more directly. According to Abraham Edel, for example, "To be a science is only the extreme point along a continuum of being more or less scientific" (1980, 8), although the examples cited by Edel are hardly convincing

in this respect: Kepler did engage in astrological musings, but this does not indicate that astrology is a science, only that Kepler was wrong on that count, despite his enormous accomplishments concerning the shape of planetary orbits. Edel suggests that the social sciences show some degree of "scientificality" and that, perhaps, ethics is not far behind. While the suggestion is interesting, it falls short of any actual positive argument about the scientificality of ethics itself. Edel then attacks both the fact-value and the means-end distinctions as artificial, in both ethics and science. The fact-value dichotomy allegedly corresponds to the distinction between descriptive and prescriptive inquiry in ethics. That may well be true, but the fact pointed out by the author that moral terms have nonmoral uses and that descriptive terms can imply moral meaning is an observation about language, and its bearing on the question of the relationship between science and ethics is not at all clear.

Similarly, Edel's point that in science the gathering of information is not value free may also withstand sociological scrutiny, but it is distinct from the theory-ladenness of scientific empirical investigations. The former is something that scientists strive to avoid; the latter is an integral and recognized part of the scientific method. Edel continues by dismissing the existence of fundamental disagreements within the ethics community as evidence that the field—unlike science—is not "going" anywhere and that this may be simply because there is no "there" to go toward. Again, while the point is well taken (profound disagreements have indeed punctuated the history of science as well as that of philosophy), this by itself is no more than an interesting observation, not a positive argument for a deep similarity between ethics and science. The burden of proof is on the continuity school to positively show why we should consider the ethics-science relationship any more than a superficial, or even misleading, analogy. In fact, Edel seems to hit a wall when turning to the importance of long-range testability in science. Where would such testability of ethical theories come from? How would it be implemented? I do, however, agree with Edel that "If a moral philosopher invoked a theory of human nature . . . he risked being charged with doing psychology or sociology rather than ethics" (1980, 13), which should not be the case since theories of human nature *should* enter ethics as fundamental pillars of our moral discourse, as I argue below.

The problem of testing moral theories is taken up also by Virginia Held (1983), who argues that they can be tested by "moral experience"—an extremely vague concept in itself. Held does have a good point when arguing that moral theories should be judged on more than just coherence, in this sense similarly to scientific theories. Held's agenda is not as ambitious as that of other continuity scholars; it starts with an acknowledgment that moral theories are normative, not descriptive, and that the interest lies in exploring the *analogy* between science and ethics. A major problem in

Held's approach is the vague definition of *testing* that is adopted: "to examine how it stands up under critical scrutiny, to see how it lasts against the erosion of time and opposition, and to watch how it overcomes the trials and obstacles to which it is subjected" (1983, 168). This leads Held to compare a person's "test" of character in the face of adversity to "testing" a set of directions for making a car. The disanalogy should be evident, and no scientist would grant that the two are even remotely analogous. In fact, Held goes even further by asserting that prediction in moral theory is the ability to prescribe the future based on general moral imperatives or universal moral judgments. But a prescription is not a description of what will actually happen, only a statement about what *should* happen. Held's examples are not at all convincing: "from the general moral imperative 'do not torture human beings,' one can derive the particular imperative 'do not torture this person in this prison'" (p. 169). In what sense is this analogous to a scientific prediction? Held's argument rapidly becomes even more questionable: "If we choose to act and do perform the chosen action and if we judge that we did the right thing, we thereby disconfirm a theory recommending we not do so" (p. 172). This is no disconfirmation of any theory; it is simply ignoring a moral command because we disagree with it or find it not compelling. No such disagreement or lack of compulsion can be afforded with regard to the theory of gravity. Furthermore, Held defines the all-crucial moral experience simply as "the experience of consciously choosing" (p. 173). This is certainly *not* what scientists mean by empirical evidence.

Held, like Edel, has to address the problem of disagreement among philosophers on fundamental ethical conclusions. She does so by reminding us that no scientist would disconfirm a physical theory because of the failure of a sophomore-level lab experiment; analogously, Held maintains, we should not reject moral theories because our grandmother disagrees with them. Yet, in the case of philosophy—and contrary to what generally happens in science—deep disagreements occur among experts in the field, not just with one's own grandmother. In science, such disagreement is certainly not unheard of, but it surely is a sign of major trouble, not of normal progress (Kuhn 1970).

Ronald L. Hall (1984) takes a slightly different approach to the ethics-science continuity question, essentially proposing that science is not really what modern Western mythology takes it to be, and that is why ethics is actually similar to science. In other words, while ethics cannot stand up to an idealized version of science seen as a steady progress toward truth, it can compare favorably to the real, significantly more messy, thing as it emerges from much literature in the philosophy of science. Hall is not in the simplistic business of considering science just another social construction, but simply to point out that science is more complex and uncertain than it is depicted in textbooks does not make a positive argument for the claim

that ethics should be considered analogous to science in any meaningful way.

Overall, one gets the impression that the major failure of continuity approaches is that they are based on a simplistic conception of either ethics or (more often) science. The unquestionable, if bumpy, progress of science is difficult to parallel in ethics, which should at least give us pause before we accept any equivalency between the two enterprises. The meanings of *theory*, *testability*, and *fact* in the two disciplines are radically different and may be incommensurable. Furthermore, the fact-value dichotomy is not genuinely analogous to the distinction between data and theory in science. Even if we acknowledge a more modern, less positivist, view of the scientific enterprise, as we should, the theory-ladenness of science does not justify an equivalency with ethics—unless one adopts an extreme postmodernist attitude to the effect that science is only or primarily a cultural construct (and at the huge price of reducing ethics to that level, too). Continuity theorists have simply not shown convincingly that ethics can be considered a science at its early stages of development.

Discontinuity: The Basics

The first group of discontinuity theories falls under the heading of practical reasoning (Darwall, Gibbard, and Railton 1997). Here, objectivity is guaranteed not by the existence of an independent metaphysical order but by the allegedly universal demands imposed on the agent by practical reason (e.g., Nagel 1970; Korsgaard 1986). The reason that science and ethics are seen as distinct is the latter's practical character (it needs to answer questions such as "What do I *do*?") as opposed to science's theoretical concern with the laws of nature. The problem I see with this approach is that its renunciation of theoretical approaches puts its supposed objectivity on shaking ground. *Why* is it that the demands imposed on the agent by practical reason are universal? The source of this universality is assumed rather than discussed, leaving the reader with an unsettling sense of arbitrariness permeating the whole approach.

A second kind of discontinuity is found in constructivism, as in John Rawls (1980). Here again we find a kind of dualism in the claim of objectivity for ethics while denying that this is of the same sort as that found in science. Constructivism rejects the possibility of discovering independent moral facts by reason; instead, objectivity really means a constructed point of view that can be shared by all. Moral principles, therefore, are not so much true as reasonable for certain sections of humankind (Rawls clearly states that his approach works only if people share certain moral intuitions and habits of thought). The only moral "facts" are the findings of what kind of outcomes one might expect given certain procedures and circumstances. This view may actually be closer to science than some versions of

continuity in that one might reasonably expect some branches of science—game theory, for example—to help in figuring out the possible outcomes of hypothetical moral systems.

Noncognitivism (e.g., Blackburn 1988) maintains that there is a sharp difference between moral and scientific meaning, though again some exponents of this school stress the similarities between ethics and science. For example, Gibbard (1997) emphasizes the use of game theory and evolutionary biology to understand how moral norms come to be accepted within human groups. It is this evolutionary underpinning that helps explain the common feeling that people have about the objectivity of moral beliefs. Notice the surprising similarity here with Foot and some aspects of neo-Aristotelianism mentioned above within continuity theories. However, the discontinuity for noncognitivists arises from the conclusion that facts about the natural world have an explanatory role that is not attributable to moral concepts. This is why to say that something is "wrong" is not to state a fact, in the noncognitivist view, but to express an attitude that *feels* (possibly because of evolution, one wants to add) like a fact.

The last kind of discontinuity theory is represented by sensibility theories (e.g., McDowell 1987; Wiggins 1991). Here the basic idea is that normative judgments may be understood by an analogy with secondary qualities, such as colors: making a normative judgment would then be exercising a kind of natural human sensibility. However, the obvious question is what objects moral judgment might be an exercise in the perception of as well as exactly what kind of sensibility is one talking about. Sensibility theorists claim that the cognition they envision concerns objects or properties that are not part of the explanatory framework of the natural sciences, nor can they be reduced to them. While this would indeed account for why people who allegedly "experience" moral demands tend to see them as categorical rather than hypothetical, one cannot help but think of plenty of other human delusions in which people "know" something that turns out to be only in their heads.

I see several general problems with the discontinuity approach. First, it too often falls into metaphysical speculation. Ethics within the discontinuity framework is supposed to be objective and universal (except for the self-imposed limitations acknowledged by Rawls), and yet the bases of such objectivity and universality are considered completely outside the framework of investigation of the natural sciences.[3] This seems to require some strong justification, little of which is actually given. Second, while some discontinuity theorists do call upon certain branches of science to buttress their claims, this seems more like an afterthought, an awkward added-on feature, rather than an organic integration between science and ethics in a joint search for knowledge. It is this deeper amalgamation of science and moral theory that I am calling for.

The Third Way: A Quasi-Discontinuous Approach to Ethics and Science

The basic idea being proposed here is that ethics and science are neither continuous nor discontinuous, at least in the senses advocated so far by authors on both sides. Instead, I think that while science is necessary for our understanding of morality in humans it is not sufficient to resolve ethical questions. Ergo, a joint ethics-science research program is necessary to make headway. I explore this possibility by means of four examples of areas in which science and ethics can work together while the latter maintains a distinctive character and is not simply reduced to a branch of science—contrary to what is advocated, for example, by E. O. Wilson (1998). The examples are necessarily sketchy, but the goal is to provide the reader with a basis sufficient to judge the validity of the quasi-continuous approach and an introduction to the necessary literature, not to present a complete survey of any of these fields. I conclude by briefly discussing some of the consequences and future directions of this research program.

The Neurobiology of Morality. Perhaps one of the most fascinating areas of interaction between science and ethics is the neurobiological research on how the human brain makes moral decisions. This rather novel field is currently characterized by very tentative results and a great deal of speculation. Yet, the whole area of neurobiological science has exploded in recent years, yielding profound new insights not only on the nitty-gritty details of neuronal action but also on general questions such as conceptual learning, language, and consciousness itself (e.g., Anderson et al. 1999; Dolan 1999; Adams 2000; de Oliveira-Souza 2000; Cardinal et al. 2001; Greene et al. 2001; Miller et al. 2001; Wallis, Anderson, and Miller 2001; Heekeren et al. 2002). For this as well as for the next few examples I cannot possibly provide a comprehensive review (for a philosophically oriented review see Casebeer and Churchland 2003) or even attempt to do justice to the mass of literature accumulated over even the past decade. Instead, I simply summarize representative case studies to provide a general understanding of my view for a joint science-ethics research program.

Damage to sectors of the prefrontal cortex is known to be associated with severe impairment of the ability to make decisions and disruption of social behavior, even though factual knowledge of social conventions and moral rules is maintained. Steven W. Anderson and colleagues (1999) conducted a study to compare patients who received the damage as adults with those who received it when very young (less than sixteen months old). The patients with early damage shared several characteristics, including being unresponsive to physical or verbal punishment, a propensity for petty thievery and lying, an incapacity to formulate plans for the future or to seek or hold a job, limited interests (usually into passive activities such

as watching TV), reckless financial behavior, no expression of remorse or guilt, and no evidence of empathy toward other human beings. These early-damage subjects had normal intellectual abilities but failed to learn rules and strategies from repeated experiences or from feedback, yielding a significant impairment of social-moral reasoning.

Perhaps the most interesting finding of the study was that, while most of these characteristics were shared by patients damaged in adulthood, the latter actually retained an ability to learn about moral rules and to understand ethical reasoning; they just failed to apply them. According to the researchers, in the early-damage patients "moral reasoning was conducted at a very early ('pre-conventional') stage, in which moral dilemmas were approached largely from the egocentric perspective of avoiding punishment" (1999, 1033). These patients also failed to make long-term moral choices over ethically questionable ones that led to immediate gain, even—and this is most important—when they could plainly see the difference from a rational perspective. These results suggest the existence of at least two distinct systems, one in charge of the acquisition of moral knowledge (by which in this context one means the rules of the particular time and society in which the subjects happen to be living), the other responsible for implementing such knowledge. The authors also noticed that the psychological profiles of these patients are similar to those of psycho- or sociopaths, and indeed that psychopathy has been associated with dysfunction in the prefrontal regions of the brain. From an evolutionary standpoint, experimental studies in animals have yielded parallel results—for example, the demonstration of defects in social interactions in baby monkeys with lesions of the amygdala and inferotemporal cortex.

But, one could argue, while this is all very interesting, is it not a matter of descriptive science rather than prescriptive ethics? Perhaps, but my point is that the two cannot be either conflated or Solomonically separated from each other. If moral truths are not "out there" but are shaped by people's natural propensities, these behaviors are "wrong" in a very special—non-normative—sense, because they arise from damage to specific areas of the brain that normally underlie ethical reasoning. In a sense, to say that an individual with neurological damage affecting her ethical decision making is wrong would be analogous to saying that a computer known to be malfunctioning has given us the wrong answer when instructed to perform a mathematical calculation. We would surely not blame the computer for the error, even though the outcome would still be a mistake. I do not mean to oversimplify matters and say that people's behaviors are as deterministic as that of a computer, but we also need to abandon the equally simplistic idea that we can transcend our biology by sheer will power. Of course, if damaged patients do develop sociopathic tendencies, they need not only to be treated but also to be restrained from harming society. This

and other studies (referenced above) show how neurobiology and evolutionary theory can assist us in understanding *why* we observe certain behavioral patterns in human beings.

Another implication of the neurobiological literature on moral decision making is that, while humans are likely predisposed (genetically hardwired) to adopt *some* moral rules, they are just as likely not born with any particular version of them. One could reasonably conclude that morality is essential for life in a social animal but that the specific version of the moral code to be adopted varies depending, within certain limits, on the circumstances (including, but not limited to, cultural ones). Once we fully accept that the brain is just another biological organ, philosophizing on the appropriateness of certain behaviors (including prescriptive ethical theorizing) will have to be constrained by the characteristics and limits of the organ in question, in the same way that the biologist's understanding of what animals eat is closely dependent on the study of their masticatory apparatus. (This is admittedly a simplification.)

Comparative Evolutionary Biology. A second area of ethics-science cooperation is to be found in comparative evolutionary biology and the study of how moral systems actually evolved. This field is unfortunately very controversial, because it is plagued by often simplistic assertions made by sociobiologists and evolutionary psychologists (Pigliucci and Kaplan 2000; Smith, Mulder, and Hill 2001; Kaplan 2002). Nevertheless, my central tenet is that philosophical investigation can be informed by the best science available and has the capacity of reasoning beyond the limits of empirical science. Obviously, in doing so the margin of error increases, but philosophizing within the constraints of science is far better than philosophizing with no constraints whatsoever (or within the milder constraints of logic alone). Recent examples of the application of evolutionary biology to the evolution of morality have been provided by Christopher Boehm (2000) and by Elliott Sober and David Sloan Wilson (1998). I will briefly discuss the latter.

I am not here endorsing Sober and Wilson's specific position as expressed in *Unto Others: The Evolution and Psychology of Unselfish Behavior* (1998). That is irrelevant to my argument. What I want to emphasize is how a science-ethics *method* can generate fruitful hypotheses and explanations that are of relevance to moral theory. Remember that the hallmark of a "good" scientific theory is not necessarily its ultimate truth (though that is certainly one quality highly sought after) but rather the amount of useful research it generates. Sober and Wilson set out to show how evolution can lead ethicists beyond the simple dichotomy presented by egoism and altruism that is fundamental to so much philosophical discussion. They start by noting that altruism in the psychological sense need not correspond to evolutionary altruism; in fact, argue Sober and Wilson, altruistic

behavior requires no conscious thinking at all, it can be innate. "The concepts of psychological egoism and altruism concern the motives that people have for acting as they do. . . . In contrast, the evolutionary concepts concern the effects of behavior on survival and reproduction . . . regardless of how, or even whether, they think or feel about the action" (1998, 6).[4]

This distinction between proximate (psychological) and ultimate (evolutionary) causes is crucial to an understanding of how biology can interface with ethical theory. The practical consequences of this distinction can also hardly be overestimated. Again, Sober and Wilson:

> . . . psychological egoists who help because this makes them feel good may make the world a better place. And psychological altruists who are misguided, or whose efforts miscarry, can make the world worse. Although the two concepts of altruism are distinct, they often are run together. People sometimes conclude that if genuine altruism does not exist in nature, then it would be mere wishful thinking to hold that psychological altruism exists in human nature. The inference does not follow. (Wilson and Sober 2000, 186)

The first part of Sober and Wilson's book is devoted to a defense of group-selection theory in evolutionary biology. I direct the reader to the original source for the technically interesting details. It is still very debatable whether group selection can play a relevant evolutionary role (Williams 1992; Morell 1996; Getty 1999); nonetheless, what interests me is the method of starting out from a scientific theory and relevant observations and expanding on their consequences for ethics, which is what Sober and Wilson do in the second part of their book. They start out by setting the parameters for acceptance or refutation of psychological egoism:

> . . . when psychological egoism seeks to explain why one person helped another, it isn't enough to show that one of the reasons for helping was self-benefit; this is quite consistent with there being another, purely altruistic, reason that the individual had for helping. Symmetrically, to refute egoism, one needs to cite examples of helping in which only other-directed motives play a role. If people sometimes help for both egoistic and altruistic ultimate reasons, then psychological egoism is false. (Wilson and Sober 2000, 197)

According to Sober and Wilson, the key in supporting altruistic philosophical theories over egoistic ones is to understand that psychology is about proximate mechanisms, while evolution is about final causes. (Once again, James would have found his thinking quite at home here.) Thus, there is no question that all organisms are evolutionarily egotistic. Yet, this leaves plenty of room for psychologically genuine altruism *if* it furthers evolutionary goals (in this case not just of the individual, but of society—hence the reliance on group selection). In fact, Sober and Wilson argue that an organism can adopt multiple evolutionary strategies, yielding a true mix of egoism and altruism that reflects what we know or intuit about actual human nature. For example, it is possible for an essentially hedonistic animal to provide parental care as a way to maximize the propagation of

its genes. Natural selection will make sure that parental care is associated with pleasure and decoupled as much as possible from pain so that the organism will behave altruistically, *because* that is the best way to be an evolutionary egoist.

Sober and Wilson conclude by noting that ethics includes both a descriptive and a normative component and hoping that their book will contribute to the exploration of the hypothesis that morality is—as Boehm (2000) argued—a result of natural selection and in particular of group selection. The importance of this to ethics is highlighted by the fact that "Every normative theory relies on a conception of human nature. . . .[5] Human nature circumscribes what is possible. . . . If we want to understand the capacities that people now have, surely an understanding of our evolutionary past is crucial" (Wilson and Sober 2000, 205). Indeed, moral theory without evolutionary biology is an enterprise conducted in an artificial vacuum. This is not to say that moral theory is, or can be reduced to, evolutionary theory, but rather that the two need to inform each other. Realizing that something evolved also does not imply that we cannot change it—within certain limits—if we wish to do so. We need to understand just what these limits are, however, or we risk embarking on a utopian program of social reform doomed to failure by ignorance of our own nature.

Game Theory. A third potentially very fruitful area of contact between science and ethics is that of game theory. This branch of mathematics has been playing an increasingly important role in evolutionary biology as a way to model the possible outcomes of natural selection given certain situations and constraints (Maynard-Smith 1982; Williams 1992b).

A good example is provided by Martin A. Nowak, Karen M. Page, and Karl Sigmund (2000) in their analysis of the ultimatum game and how it can enlighten us on the distinction (or lack thereof, under certain circumstances) between rationality and fairness. The game is simple: two players have to agree on how to split a sum of money, and they only get one shot at it (i.e., there is no bargaining). The proponent makes an offer; if the receiver accepts it, they split the money in the proposed fashion, and if not, they both lose everything. It is easy to demonstrate that a rational responder should accept even very small amounts of money, because the alternative is to get nothing, and she has no negotiating power. Therefore, the proponent should be able to get almost the whole sum. Yet when the game is played by human beings—and even when results are compared across cultures, with different kinds of incentives being offered—the majority of proponents offer between 40 percent and 50 percent, and as many as half of the responders reject offers of less than 30 percent. It would seem that human beings tend to be more fair than rational.

But Nowak and coworkers noted that things might not be that simple. They proposed that human beings react instinctively without realizing that

there is no possibility of playing the game again. We are accustomed to repeated interactions, in which situation it makes perfect sense (i.e., it is rational) for the responder to "punish" the proponent for being unfair. The proponent, in that case, will lose not only the money but also his reputation. The responder will lose less money than the proponent, and she will have the advantage of seeing her reputation (for fairness, a social value) increase in preparation for future interactions. Nowak and colleagues simulated a continuum of situations in which the players had access to a certain amount of information about each other (their reputation), from zero to very reliable. With no information available on the outcome of previous encounters, the system might be expected to settle on a population of rational agents, not of fair players. However, the more background information is added, the more the stable strategy shifts first toward multiple potential outcomes (a mix of rational and fair) and eventually settles on fair. This is true whatever the mechanism involved, be it learning by trial and error, imitation, or genetic inheritance. In other words, taking into account reputation, adaptation favors fairness, and there is no more distinction between the latter and reason, because it becomes now irrational to behave as if reputation were not a factor. A change in the environment (from nonsocial to social) has changed what it means to act rationally. This, according to the authors, agrees well with published studies on the emergence of cooperation and bargaining in human societies. To me, this represents an excellent example of how science (game theory in this case) can inform ethical *prescriptive* theory, not just ethics as description: the circumstances (degree of sociality) define which actions are rational and—if one subscribes to the idea that ethical judgment can be arrived at rationally—moral judgment itself.

Comparative Anthropology. My fourth and last major area of ethics-science collaboration is comparative anthropology. In this case, the work bridging ethics and science can be done comparing the few remaining human hunter-gatherer societies, other "non-modern" societies, or by considering technological societies coexisting in the modern world. The literature in this field is vast and certainly not free from controversy. Some of this literature has directly involved the field of ethics, in particular bioethics. For example, L. Turner (1998) has applied the anthropological perspective to contemporary bioethical issues. I use his work as an example of how comparative anthropology can inform ethical philosophizing. Turner notes that some scholars emphasize the existence of some kind of "core morality" that cuts across human societies, while others contend that there is a degree of incommensurability among the beliefs of different cultures. If indeed there were a shared "common sense" across cultures, applied ethics would be an easier task than it appears. It would be possible to proceed from simple, paradigmatic examples to the consideration of

more sophisticated situations from the vantage point of an at least partially common ethical language. This resembles Rawls's idea of "overlapping consensus" of moral and religious traditions leading to a stable and coherent reflective equilibrium.[6] Unfortunately, remarks Turner, comparative anthropology clearly shows how these authors "do not recognize the extent to which their moral imagination is contingent upon (their own) social history, religious background, and social status" (1998, 129).

Turner discusses specific examples, such as truth telling concerning cancer patients in the U.S. as compared to Japan or Italy. Interestingly, the latter two are very different cultures, with Italy in many respects closer to the American than to the Japanese model. Italians and Japanese by and large do not share the American attitude that patients have a right to be informed of their conditions. Indeed, physicians and families who shield patients from distress in those countries are seen as carrying out their social role of fostering hope and protecting their loved ones from the results of what is perceived as devastating knowledge. This is not only an interesting anthropological observation, but it has serious consequences for medical practice in the U.S. Since the U.S. is a multicultural society, doctors cannot just assume that their version of "common" sense is shared by all their patients, lest they find themselves in bioethically troublesome waters. Similar ethical Pandora's boxes can be helped by comparative anthropological studies of problems such as euthanasia, brain death, and organ transplants. Turner concludes:

> Instead of assuming the singularity of common sense, it is possible to argue that, over time, multiple webs could develop, with correspondingly different notions of common sense. Alternatively, there may exist a single web, within which all human reflection takes place. However, this web might be so elaborate and variegated that human participants within particular communities could have only a rather limited appreciation of the claims made by participants from other communities. (1998, 133)

While comparative anthropology can readily be used in problems of applied ethics, should the meta-ethical community be concerned about it? After all, moral principles could be entirely independent of (and indeed, in direct contradiction to) the specific norms adopted by any particular human society at any point in time: again, the distinction between descriptive and prescriptive ethics is apparent. But such an attitude flies in the face of the utter lack of objective evidence for the existence of moral truths "out there" and goes against the evolutionary approach adopted in this essay. If ethics is neither universal nor entirely arbitrary but rather the result of historical contingencies and the necessities of a particular species of primates, as I maintain here, then comparative anthropological studies can provide the equivalent of natural experiments in biology. These natural experiments can be compared to the wider universe of simulation experiments afforded by approaches such as game theory to establish a solid

theoretical-empirical foundation for ethics not dissimilar from that of partially historical sciences such as evolutionary biology. All of this, of course, with the caveat that we can actually purposefully change what is to what we think ought to be, thereby maintaining the semi-discontinuity between ethics and science which is proposed as a paradigm in this essay.

Implications of the New Program for Some Long-Held Ethical Views

What game theory, comparative evolutionary biology, comparative anthropology, and neurobiology have discovered in recent years has the potential to have dramatic effects on some long-standing philosophical positions and debates. A satisfactory exploration of such implications constitutes a research program in and of itself and will not be pursued here; however, following the general spirit of this essay, I would like to provide a taste of what such program might look like, hoping to whet the appetite of other philosophers inclined to pursue this matter further.

A first example comes from the consideration of Kant's proposal of one of the most enduring systems of deontological ethics in his *Groundwork of the Metaphysics of Morals* ([1785] 1998). In it, he suggested that the moral life is a matter of rationality, not emotions (clearly, contrary to Hume's opinion discussed earlier). According to a strict reading of Kant, actions done because of compassionate inclinations have no moral worth. Rather, it is duty as discovered and implemented by rational thought that is praiseworthy.[7] This approach has always struck me as substantially at odds with everything we know or intuit about human nature, and yet it has generated a long tradition in ethics of considering the merits of rationality and its allegedly fundamental role in discovering moral truths. The problem is, some astounding discoveries in neurobiology have the potential to drive a mortal stake through the heart of any philosophical argument tightly linking rational thought and moral duty.

Perhaps the most intriguing of these findings concerns patients with extensive damage to the amygdalas, two almond-shaped structures embedded in the temporal lobes that function as gateways to the limbic system. The limbic pathways mediate emotional arousal. Some patients have been found to have completely calcified (and therefore nonfunctional) amygdalas, while other patients have lost them as a result of accident or surgery. One such patient, studied by Antonio Damasio (1999), had essentially lost any negative emotion whatsoever. She experienced no fear, regardless of the stimulus, and she was incapable of acting in socially apt ways whenever a healthy negative or cautious reaction would have been necessary. The patient did, however, retain positive emotions, which made it easy for her to make friends but also got her into awkward social situations with what normal humans would consider excessive or premature effusions. This

patient did retain the capacity to *rationally* understand under what circumstances she should have been more circumspect or fearful or should have altered her social behavior; however, this intellectual knowledge was completely useless to her, because it was not accompanied by any baseline for comparison provided by emotional responses.

V. S. Ramachandran (Ramachandran and Blakeslee 1998) discusses several cases of brain damage affecting emotional responses and moral choice, including what happens to patients without amygdalas. He compares such individuals to purebred versions of Mr. Spock, the Vulcan character in *Star Trek*, characterized by a general blunting of their emotional responses, which can be measured by their absolutely flat galvanic skin responses.[8] An extreme case is presented by individuals affected by the rare Cotard's syndrome, which essentially disconnects all sensorial inputs from the limbic system, with the result that the patients can perceive and think rationally about the world but have no way to connect any emotion whatsoever to their thinking and perceiving. Here we have human beings who are capable of rational thinking as assessed by any standard measure but who are completely incapable of emotional responses. These patients characteristically are convinced that they are dead (to the point that they assert smelling their own rotten flesh or worms crawling on their skin) (Ramachandran and Blakeslee 1998, 167). Cotard's-syndrome patients are as close as we can imagine to experiencing death. It seems that Mr. Spock is not really a believable character after all and that pure rationality without emotion does not lead to the moral life but to a nightmare existence characterized by a complete indifference toward any aspect of life, moral or otherwise. While this may be puzzling for a rationalist, it makes perfect sense to an evolutionary biologist. If behavior (moral or otherwise) is adaptive, and therefore favored by natural selection, it has to be associated (by selection) with pleasurable emotions—and vice versa, immoral behavior with unpleasurable emotions—to "persuade" people to act one way or the other.[9] This is no different from the more mundane explanation of why sex is pleasurable. If it weren't, people would not engage in it, and their line of descent would rapidly dwindle to extinction.

When one considers the pieces of the puzzle provided by neurobiology, Hume's view of morality as expressed in his *Treatise* comes closer to reality. His discussion of the relationship between emotions and reason is much more levelheaded than that of most philosophers before or since. He claimed that reason can never provide a motive for any action, because this comes from emotions and feelings. He criticized philosophers who maintained that we need to use reason to regulate our conduct while suppressing our passions. As we have seen, such Spock-like "idealized" human beings would be entirely dysfunctional in society. As Hume puts it:

> 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin,

to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledge'd lesser good to my greater, and have a more ardent affection for the former than the latter. ([1739–40] 1978, bk. 2, part 3, sect. 3, p. 415)

This resonates very closely with the cases just discussed and with our previous consideration of patients with frontal lobe damage who are unable to seek their own long-term good even though they understand their predicament from a rational point of view. This, Hume says, is because reason is a means to the end of passions and desires. It can tell us how to best satisfy the latter two, but by itself it cannot generate any action, as the victims of the horrible Cotard's syndrome experience all too well. While I would not go as far as Hume in his famous statement about the relationship between reason and passion quoted at the beginning of this essay, his view nevertheless acquires new force from research in neurobiology.

A second example of the consequences of a quasi-continuous approach between science and ethics affects the endless discussions about egoism and altruism that plague the ethical literature, as I have already mentioned. Works along the lines of Sober and Wilson's (1998) show that such dichotomy may vanish if one considers the distinction between psychological and evolutionary altruism/egoism and understands that perfectly genuine psychological altruism may be the necessary product of evolutionary egoism. Furthermore, game theory, as we have seen, actually shows that concepts such as fairness and rational egoism depend on the environment in which individuals live, so that it makes perfect sense (that is, it is advantageous for the agent) to be cooperative or fair within a society where gossip creates and destroys reputations. Simpleminded egoism can be completely irrational in such circumstances. The debate concerning altruism and egoism needs to consider such findings and other current research in the biological sciences insofar as cases that were thought of as altruism may turn out not to be and vice versa.

Another fundamental idea in moral philosophy that can be affected by a science-ethics continuum is the concept of a social contract as a basis for the acceptance of our notions of morality and justice. Brian Skyrms (2000) uses game theory to bring discussions of Hobbes-like social contracts into the realm of scientific investigation. Skyrms starts out by considering simple game-theoretical approaches that make the same assumptions of rationality in the agents as contractual philosophical theories. He then proceeds to show that theories of rationality so constructed tend to single out a unique outcome in the case of zero-sum games but not in the broader and more realistic context of nonzero-sum games. Skyrms suggests that a much more reasonable conception of the social contract, and one in line with what emerges from game-theoretical models, is found (again) in Hume, who saw the social contract as a patchwork of conventions grown up over time. In fact, Skyrms comments that modern Humeans such as Gibbard

"take inspiration as well from Darwinian dynamics. The social contract has evolved. . . . Different cultures, with their alternative social conventions, may be instances of different equilibria, each with its own basin of attraction" (2000, 272).[10] Skyrms continues: "There is a moral here for philosophers and political theorists who have attacked the theory of games on the basis of its rationality assumptions. Game theory has a far broader domain of application than that suggested by its classical foundation" (p. 275).

An important point to consider is that Darwinian game-theoretical models allow for a plurality of strategies, including suboptimal ones, to coexist at equilibrium. This has the potential to lead to much more realistic philosophical models of human societies, reflecting the observed current and historical multiplicity of "solutions" humans have adopted in order to ensure a more or less harmonious (or simply stable) social life. Information on the variety and fitness of these alternative solutions can come from comparative anthropology, as discussed above. Skyrms considers as an example the relaxation of some rigid assumptions typically embodied in game-theoretical models of situations such as the prisoner's dilemma[11] and concludes that it is actually fairly easy to see the emergence of cooperation over selfishness, and that—again—it simply does not make any sense to consider one or the other as rational because rationality depends on the context. In his words: "In conditions, such as those of correlated encounters, where the evolutionary dynamic theory is structurally at odds with the rational choice theory, the evolutionary theory provides the best account of human behavior" (2000, 282). Because of these considerations, it seems to me that speculating about social contracts without a more rigorous, game-theory-like, approach to back up one's intuitions might soon become unacceptable philosophical practice.

The Road Ahead

In this essay I have outlined a possible resolution of the ethics-science continuity/discontinuity debate. As G. W. F. Hegel proposed ([1821] 1952), many philosophical debates can be cast in terms of a thesis and antithesis, the resolution of which may forge a synthetic view that retains the best of both original positions. Indeed, some authors who would count themselves in the discontinuity camp have in practice adopted the idea of science necessarily informing ethics that I am proposing here as a new research program in ethical theory (e.g., Gibbard 1997).

The main thrust of what I am proposing is the explicit recognition of two points: (1) philosophy in general (not just ethics) cannot afford to do without a serious and in-depth consideration of the findings of science, and (2) science by itself is not capable of addressing the broad questions of interest to philosophers and, ultimately, to humanity itself. The first point

is becoming increasingly clear from the fact that philosophers are incorporating science more and more into their work (despite the danger of "scientism": see Sorell 1991), realizing that science is the best tool to provide the philosopher with boundary conditions to frame his or her theories so that they are relevant to the world as it is. We are free to speculate about possible universes, but we are ultimately interested in the real one in which we have to live. It makes much sense, therefore—indeed it would seem to be imperative—to use science to circumscribe that world and then focus our philosophizing within the realistic parameters so identified.

The second point is granted by most scientists (except, e.g., Wilson [1998]). Science is extremely powerful when it can deal with focused questions of an empirical nature, but when the scientific method is brought in as the sole arbiter in broader philosophical questions, it fails, because the empirical evidence is either not obtainable or only a partial component of a much larger picture.

The research program sketched here requires ethicists to become conversant in the language of science, especially of evolutionary biology, game theory, comparative anthropology, and neurobiology. This is a daunting task indeed, but recoiling from it might lead to endless and sterile squabbles about matters the comprehension of which we will never significantly advance without a fresh perspective. At the same time, ethicists cannot do it all. They are certainly not expected to conduct neurobiological experiments or engage in direct anthropological research. The field of ethical philosophy itself is large enough to preclude individuals' fully encompassing the whole spectrum that I have depicted here. Therefore, the real answer must lie in close cross-discipline collaborations between ethicists and biologists in a new and invigorating search for the unity of knowledge. This was only barely conceivable in James's time, but it is now a real option offered by the advancements of current science, and it might be the challenge for both science and philosophy in the twenty-first century.

NOTES

1. An argument can be made that science—as any complex concept—can best be understood as a "family-resemblance" concept in Wittgenstein's sense rather than being defined by some sharp discontinuity with other sorts of activities. Even so, the question here becomes whether the set of family-resemblance properties defining science is distinct enough from the one defining ethics.

2. I am not actually defending reductionism in science here. What I have in mind is closer to John Dupré's idea (1993) that lower hierarchical levels explain *how* rather than *what* is possible at higher levels. So, for example, while molecular genetics cannot be used to predict the response of populations to natural selection, it explains how the genetic machinery of organisms makes such response possible.

3. It could be argued that in this ethics is not different from mathematics. Mathematical truths are not investigated in the same way as scientific truths. There are intuitions that are shared by most working mathematicians, such as what counts as a valid inference and what constitutes a good starting point for reasoning, and conclusions follow from these. There are two problems with this rejoinder, as far as I can see. First, mathematics is not a science in the sense of

having the goal of understanding the physical natural world; rather, it is a self-contained logical system that would be unaltered even if the outside world were dramatically different from what it is. Second, not all moral theorists share the same kinds of intuitions and reasoning methodologies, which makes mathematics a much tighter logical discipline than ethics.

4. The gap between these two levels is filled by a number of other layers, including just what goes on in the brain that results in a particular action being carried out. This mid-layer level of causality is addressed by the neurological literature discussed above.

5. I am not as convinced of this as Sober and Wilson are. I think that every normative theory *should* be based on a conception of human nature, but see Kant's as one that clearly is not. On the other hand, I think this is exactly the reason why the Kantian system encounters so many problems when applied to real human beings.

6. However, as I already pointed out, Rawls does not think that this overlapping consensus will encompass everyone in the world or even everyone in the U.S. In fact, a significant part of his *Theory of Justice* (1971) is spent dealing with the problem of those persons within a society whom this "consensus" does not cover or apply to, who disagree with the majority about such basic things that it is impossible to engage them in moral reasoning as understood by most members of the society.

7. I realize that some neo-Kantians think that this interpretation of Kant is extreme, but see Singer 1995 in support of my understanding of moral worthiness in Kant.

8. Notice that in all of these cases the patients are otherwise normal, so that these examples cannot simply be dismissed as aberrant and uninformative. Furthermore, experimental neurobiology is devising more and more sophisticated instruments capable of studying the effects of smaller alterations in the brain, as well as methods to study the intact brain in action. In a sense, these approaches are the neurobiological equivalent of using mutations in genetics to study how living organisms are put together. Despite limitations, the latter research program became one of the most successful of twentieth-century science.

9. This principle is also applied in the form of a system of sanctions and rewards within any society that wishes to enforce a given set of rules, no matter how arbitrary (see Sober and Wilson 1998, chap. 5). The difference in the case of natural selection is that it works through long periods of time and leads to the evolution of feelings that are highly internalized, and therefore more stable and difficult to overcome. That is why biology puts a stricter limit than culture on what we can do.

10. The reference to "basins of attraction" is in the context of chaos/complexity theory, although the extent to which the latter is relevant to game-theoretical models or evolutionary ecology in general is still under dispute (Stone 1994; Solé et al 1999).

11. The prisoner's dilemma is a classic game-theoretic class of situations in which two players have to decide whether to cooperate or defect, based on a matrix of payoffs (for cooperating) and penalties or costs (for defecting). It tuns out that if the prisoner's dilemma is played in a single round, defecting is the typical outcome, even though it often costs both players; however, when the game is played iteratively (simulating repeated social interactions), cooperation emerges as the rational strategy, because it helps a given player to build and maintain a "reputation" within the group.

REFERENCES

Adams, J. 2000. Gray Matter and Gray Areas: The Neuroscience of Morality. *brain.com.*

Anderson, Steven W., Antoine Bechara, Hanna Damasio, Daniel Tranel, and Antonio R. Damasio. 1999. "Impairment of Social and Moral Behavior Related to Early Damage in Human Prefrontal Cortex." *Neuroscience* 2:1032–37.

Blackburn, Simon. 1988. "How to Be an Ethical Antirealist." *Midwest Studies in Philosophy* 12:361–75.

Boehm, Christopher. 2000. "Conflict and the Evolution of Social Control." *Journal of Consciousness* 7:79–101.

Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, ed. G. Sayre-McCord. Ithaca, N.Y.: Cornell Univ. Press.

Cardinal, Rudolf N., David R. Pennicott, C. Lakmali Sugathapala, Trevor W. Robbins, and Barry J. Everitt. 2001. "Impulsive Choice Induced in Rats by Lesions of the Nucleus Accumbens Core." *Science* 292:2499–2501.

Casebeer, William D., and Patricia S. Churchland. 2003. "The Neural Mechanisms of Moral Cognition: A Multiple-Aspect Approach to Moral Judgment and Decision-Making." *Biology and Philosophy* 18:169–94.
Cleland, Carol E. 2001. "Historical Science, Experimental Science, and the Scientific Method." *Geology* 29:987–90.
Damasio, Antonio. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.
Darwall, Stephen, Allan Gibbard, and Peter Railton, eds. 1997a. *Moral Discourse & Practice: Some Philosophical Approaches*. New York: Oxford Univ. Press.
———. 1997b. "Toward *fin de siecle* Ethics: Some Trends." In *Moral Discourse & Practice: Some Philosophical Approaches*, ed. S. Darwall, A. Gibbard, and P. Railton, 3–47. New York: Oxford Univ. Press.
de Oliveira-Souza, Ricardo. 2000. "The Moral Brain: A Functional MRI Study of Moral Judgment." *Neurology* 54:A104.
Dolan, Raymond J. 1999. "On the Neurology of Morals." *Nature Neuroscience* 2:927–29.
Dupré, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge: Harvard Univ. Press.
Edel, Abraham. 1980. "Ethics—A Modest Science?" *Zygon: Journal of Religion and Science* 15 (March): 7–19.
Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives." *Philosophical Review* 81:305–16.
Getty, Thomas. 1999. "What Do Experimental Studies Tell Us about Group Selection in Nature?" *American Naturalist* 154:596–98.
Gibbard, Allan. 1997. "Wise Choices, Apt Feelings." In *Moral Discourse & Practice: Some Philosophical Approaches*, ed. S. Darwall, A. Gibbard, and P. Railton, 179–98. New York: Oxford Univ. Press.
Greene, Joshua D., R. Brian Somerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293:2105–8.
Hacking, Ian. 1999. *The Social Construction of What?* Cambridge: Harvard Univ. Press.
Hall, Ronald L. 1984. "The Analogy between Ethics and Science." *Zygon: Journal of Religion and Science* 19 (March): 83–85.
Heekeren, H. R., I. Wartenburger, H. Schmidt, C. Denkler, H. P. Schwintowski, and A. Villringer. 2002. "Brain Moral Judgment—Towards a Neurobiology of Law." *Lili* 32:114–26.
Hegel, Georg Wilhelm Friedrich. [1821] 1952. "The Philosophy of Right; The Philosophy of History." Chicago: *Encyclopædia Britannica* (Great Books of the Western World), 369.
Held, Virginia. 1983. "The Validity of Moral Theories." *Zygon: Journal of Religion and Science* 18 (June): 167–81.
Hume, David. [1739–40] 1978. *A Treatise of Human Nature*. New York: Clarendon.
James, William. 1891. "The Moral Philosopher and the Moral Life." In *Essays in Pragmatism*, ed. A. Castell, 176. New York: Hafner.
———. 1918. *The Principles of Psychology*. New York: H. Holt.
Kant, Immanuel. [1785] 1998. *Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge Univ. Press.
Kaplan, Jonathan. 2002. "Historical Evidence and Human Adaptations." *Philosophy of Science* 69: S294–S304.
Kitcher, Philip. 1995. "Precis of the Advancement of Science." *Philosophy and Phenomenological Research* 55:611–17.
Korsgaard, Christine. 1986. "Skepticism about Practical Reason." *Journal of Philosophy* 83:5–26.
Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions*. Chicago: Univ. of Chicago Press.
Lakatos, Imre. 1977. *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge Univ. Press.
Lewontin, Richard C. 1998. "The Evolution of Cognition: Questions We Will Never Answer." In *Methods, Models, and Conceptual Issues. An Invitation to Cognitive Science*, ed. D. Scarborough and S. Sternberg, 107–32. Cambridge: MIT Press.

Mackie, John Leslie. [1977] 1997. "The Subjectivity of Values." In *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
Maynard-Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge Univ. Press.
McDowell, John. 1987. *Projection and Truth in Ethics*. Lawrence: Univ. of Kansas Press.
Miller, B. L., W. W. Seeley, P. Mychack, H. J. Rosen, I. Mena, and K. Boone. 2001. "Neuroanatomy of the Self: Evidence from Patients with Frontotemporal Dementia." *Neurology* 57:817–21.
Moore, George Edward. 1903. *Principia Ethica*. Cambridge: Cambridge Univ. Press.
Morell, Virginia. 1996. "Genes vs. Teams: Weighing Group Tactics in Evolution." *Science* 273: 739–40.
Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton Univ. Press.
Nowak, Martin. A., Karen M. Page, and Karl Sigmund. 2000. "Fairness versus Reason in the Ultimatum Game." *Science* 289:1773–75.
Perry, David A. 1995. "Self-organizing Systems across Scales." *Trends in Ecology and Evolution* 10:241–44.
———. 2002. *Denying Evolution: Creationism, Scientism and the Nature of Science*. Sunderland, Mass.: Sinauer.
Pigliucci, Massimo, and Jonathan Kaplan. 2000. "The Fall and Rise of Dr. Pangloss: Adaptationism and the Spandrels Paper 20 Years Later." *Trends in Ecology & Evolution* 15:66–70.
Popper, Karl R. 1968. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper and Row.
Railton, Peter. 1986. "Moral Realism." *Philosophical Review* 95:163–207.
Ramachandran, V. S., and Sandra Blakeslee. 1998. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: William Morrow.
Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard Univ. Press, Belknap Press.
———. 1980. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77:515–72.
Singer, Peter. 1995. *How Are We to Live? Ethics in an Age of Self-Interest*. Amherst, N.Y.: Prometheus.
Skyrms, Brian. 2000. "Game Theory, Rationality and Evolution of the Social Contract." *Journal of Consciousness* 7:269–84.
Smith, Eric A., Monique Borgerhoff Mulder, and Kim Hill. 2001. "Controversies in the Evolutionary Social Sciences: A Guide for the Perplexed." *Trends in Ecology & Evolution* 16:128–35.
Sober, Elliott, and David Sloan Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard Univ. Press.
Solé, Ricard V., Susanna C. Manrubia, Michael Benton, Stuart Kauffman, and Per Bak. 1999. "Criticality and Scaling in Evolutionary Ecology." *Trends in Ecology & Evolution* 14:156–60.
Sorell, Tom. 1991. *Scientism: Philosophy and the Infatuation with Science*. London: Routledge.
Stone, Lewi. 1994. "Ecological Chaos." *Nature* 367:418.
Turner, Leigh. 1998. "An Anthropological Exploration of Contemporary Bioethics: The Varieties of Common Sense." *Journal of Medical Ethics* 24:127–33.
Wallis, Jonathan D., Kathleen C. Anderson, and Earl K. Miller. 2001. "Single Neurons in Prefrontal Cortex Encode Abstract Rules." *Nature* 411:953–56.
Wiggins, David. 1991. "A Sensible Subjectivism?" In *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford, England: Blackwell.
Williams, George C. 1992a. *Natural Selection: Domains, Levels and Challenges*. New York: Oxford Univ. Press.
———. 1992b. "Optimization and Related Concepts." In *Natural Selection: Domains, Levels and Challenges*, 56–71. New York: Cambridge Univ. Press.
Wilson, David Sloan, and Elliott Sober. 2000. "Summary of *Unto Others: The Evolution and Psychology of Unselfish Behavior*." *Journal of Consciousness Studies* 7 (1–2): 185–206.
Wilson, Edward O. 1998. *Consilience: The Unity of Knowledge*. New York: Knopf, distributed by Random House.