

# NATURALIZING ETHICS: THE BIOLOGY AND PSYCHOLOGY OF MORAL AGENCY

by *William A. Rottschaefer*

*Abstract.* Moral agency is a central feature of both religious and secular conceptions of human beings. In this paper I outline a scientific naturalistic model of moral agency making use of current findings and theories in sociobiology, developmental psychology, and social cognitive theory. The model provides answers to four central questions about moral agency: (1) what it is, (2) how it is acquired, (3) how it is put to work, and (4) how it is justified. I suggest that this model can provide religious and secular moral theories with a basis for a common understanding of moral agency.

*Keywords:* altruism; Albert Bandura; ethics and science; evolutionary ethics; genetic selfishness; Martin Hoffman; moral agency; moral agency and biology; moral agency and psychology; moral development; naturalistic fallacy; naturalized ethics; religion, ethics, and science; social cognitive theory; sociobiology.

---

## MORAL AGENCY AS A CENTRAL FEATURE OF HUMAN PERSONS

As is well known, the major religious traditions have different views not only of the divine, but also of the human person. Their views range from the belief of the Jewish, Islamic, and Christian traditions that the human person is a substantial self, or, in some interpretations, a substantial immaterial soul, to the view in Hindu thought that the human atman is identical with the divine Brahman. They move from these views of the human person as a substantial self, whether plural or singular, to the Buddhist traditions that hold a no-self view of the human person. Despite these differences, all the major religious traditions understand human beings to be

William A. Rottschaefer is Professor of Philosophy at Lewis and Clark College. His mailing address is Department of Philosophy, Lewis and Clark College, 0615 SW Palatine Hill Road, Portland, OR 97219. His e-mail address is [rotts@lclark.edu](mailto:rotts@lclark.edu).

[*Zygon*, vol. 35, no. 2 (June 2000).]

© 2000 by the Joint Publication Board of *Zygon*. ISSN 0591-2385

moral agents, and even though some of their substantive norms of morality may differ in details, there is agreement about the capacities of persons for moral agency and about major moral values. This consensus about moral agency and moral values reflects a larger common-sense conviction that human beings are moral agents who share some basic moral values.

Moreover, secularists, whether of the humanistic or scientific naturalistic variety, agree with their religious counterparts about the moral capacities of human persons, although each party, secularist or religionist, may be skeptical of the other's ability to provide adequate grounding for their convictions. For instance, Charles Darwin took it as one of his central tasks to give an evolutionary account of moral agency, a phenomenon which he considered unique to human beings and a central feature of their nature. Thus, in his *The Descent of Man and Selection in Relation to Sex*, Darwin wrote,

I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important. This sense, as Mackintosh remarks, "has a rightful supremacy over every other principle of human action"; it is summed up in that short but imperious word *ought*, so full of high significance. It is the most noble of all the attributes of man, leading him without a moment's hesitation to risk his life for that of a fellow-creature; or after due deliberation, impelled simply by the deep feeling of right or duty, to sacrifice it in some great cause. (Darwin 1871, 70)

And a few lines later, in reply to Kant's question about duty, "whence thy original," he continues,

This great question has been discussed by many writers of consummate ability; and my sole excuse for touching on it is the impossibility of here passing it over, and because, as far as I know, no one has approached it exclusively from the side of natural history. The investigation possesses, also, some independent interest, as an attempt to see how far the study of the lower animals can throw light on one of the highest psychical faculties of man. (p. 71)

This consensus about the centrality of moral agency to human identity, though perhaps rough-hewn and confined only to broad outlines, can, I believe, provide the basis for dialogue among parties that otherwise seem to hold little in common about human persons and their nature, origin, place in the scheme of things, and ultimate destiny. As a part of an effort to promote mutual understanding and perhaps provide some commonly shared bases for whatever consensus about human beings as moral agents we may share, I want to present the outlines of a scientifically naturalistic philosophical account of moral agency.

First I describe a scientific naturalistic account of moral agency and contrast it with other approaches to understanding moral agency. I then consider these approaches within the context of the possible relationships between the sciences and ethics, distinguishing four major sorts of relationships: separatist, antagonist, compatibilist, and integrationist. Given

this contextualization of the topic and focusing on the potential contributions of biology and psychology to an account of moral agency, I then lay out six increasingly controversial hypotheses about how one should connect the results of the sciences with accounts of moral agency. In the following sections I focus on two of these connections, the explanatory and the metaethical, and show how they emerge in attempts to answer the four central questions about moral agency: (1) what it is, (2) how it is acquired, (3) how it is put to work, and (4) how it is justified. Next I propose a scientific naturalistic model for answering the question of what moral agency is. I then look at findings from sociobiology, developmental psychology, socioemotional development, and social-cognitive theory to answer questions about how moral agency is acquired and how it is put to work. That leads me in the final section to the question of how the findings of biology and psychology can be brought to bear not only to describe and explain moral agency but to justify it.

#### A SCIENTIFIC NATURALISTIC PHILOSOPHICAL ACCOUNT OF MORAL AGENCY

Besides a scientific naturalistic philosophical account of moral agency, we can distinguish some alternative approaches to understanding moral agency, including common-sense, humanistic, a priori philosophical, and theological-religious. These approaches are marked off particularly by the sources that they use in accounting for moral agency. By a common-sense understanding of moral agency, I mean one that uses both of our ordinary conceptions about the nature of moral agency, conceptions that may have an evolutionary derivation as well as a learned component. Religious-theological conceptions understand moral agency in the context of a body of religiously based and theologically developed claims that appeal to religious traditions and to some special divine revelation, religious insight, or religious experience. Humanistic accounts of moral agency are based on such refinements of ordinary human conceptions as appear, for instance, in literature and the arts. A priori philosophical accounts of moral agency appeal to various avenues of knowledge about moral agency that are independent of empirical input, findings, and theories. These accounts rely on philosophical intuitions, conceptual analyses, the demands of reason, and appeals to logical necessities and possibilities. Finally, we come to the scientific naturalistic philosophical approach, which I shall develop.

By an account of moral agency I mean a description, explanation, and understanding of moral agency that answers four central questions about moral agency: What is moral agency? How is it acquired? How is it activated? How are the central activities of moral agency, that is, moral belief formation, moral motivation, and moral action, justified? By a *philosophical* account of moral agency I mean one that addresses traditional questions in

the philosophical discipline of ethics. Although I believe that the sciences provide the best means for answering philosophical questions, I contend that these questions are substantively distinct from those asked in the sciences because of their comprehensive character and in some cases because of their normative character. By a *naturalistic* account of moral agency I mean one that makes moral agency a phenomenon of the material world. By a *scientific naturalistic* account I mean one that makes central use of the natural and social sciences to understand and explain moral agency.

Minimally, a scientific naturalistic perspective rules out an *exclusive* reliance on purely common-sense, humanistic, a priori, or religious approaches. In particular, it rules out approaches that take moral values out of the material world and place them in some nonmaterial realm, thus making access to them exclusively nonempirical. A scientific naturalistic perspective makes the further assumption that the sciences provide the best theoretical and empirical knowledge available for understanding morality. Scientific naturalists ought not to be scientistic, that is, they ought not to exclude a priori any non-scientifically based claims about moral agency. On the other hand, scientific naturalists should not accept any claims about moral agency without judging the grounds on which they are based. The track record of the various approaches to answering questions about moral agency should guide their assessments. Of course, determining the criteria for such assessments is no easy matter. Though I will not argue the matter here, I maintain that the multiple methods of the sciences have demonstrated their superiority to those of humanistic, a priori, and religious approaches with respect to epistemic criteria derivable from the ordinary perceptual and inferential capacities shared by all parties. Thus, though the scientific naturalistic perspective does not claim that nonscientific sources of justification should be excluded as justifiers for claims about moral agency, it does insist that such sources be as reliable as scientific sources.

Although in this paper I am primarily concerned with outlining a positive account of my own scientific naturalistic model of moral agency, it should be clear that the success of any model of moral agency depends on its adequacy relative to its competitors, whether scientific naturalistic, common-sense, humanistic, religious-theological, or a priori philosophical. I take some criteria of adequacy to be internal consistency and coherence, empirical support, coherence with common-sense views, fit with the best current scientific findings and theories, explanatory power, and heuristic fruitfulness.

In order to further clarify the focus and nature of my model, I compare it with an understanding of ethics that has become common within the Anglo-American analytic philosophical tradition during this century. In that tradition, ethics is usually divided into three parts: (1) metaethics, which addresses issues of moral epistemology and ontology, the nature and

function of morality, and the nature of moral agency; (2) normative ethics, which deals with moral principles and norms; and (3) applied ethics, which is concerned with the application of moral principles and norms to the resolution of moral issues in particular areas of moral concern, for instance, biomedical ethics. Because my perspective is theoretical, I am concerned with metaethics rather than normative or applied ethics.

#### RELATING THE ACCOUNTS OF MORAL AGENCY TO THE SCIENCES

No matter what one's approach to understanding moral agency, the question arises of how to relate one's efforts to the relevant sciences. Consider now the stances that a study of moral agency might take toward the sciences. There are two sorts of general positions about the relationships between the sciences and ethics. First, there are the *separatists*, who argue for a fundamental distinction between the sciences and ethics. If there are any links, the separatists believe they are relatively few and insubstantial. For instance, many religious believers hold that morality is based on theology or sacred books, so they believe that religion and ethics are intimately connected. At the same time they believe religion and the sciences to be fundamentally distinct. For those believers, separatism would come naturally. There could be another kind of separatist. Imagine humanists who are nonreligious. They do not connect religion and ethics, but they might believe that ethics derives from experience and disciplined philosophical reflection on that experience. If they also think philosophical reflection is independent of the sciences, then they too would probably join the separatist camp.

If you are not a separatist, then you are an *interactionist*. Interactionists contend that there are many, substantial connections between the sciences and ethics. Among the interactionists, we can discern three sorts: *antagonists*, *compatibilists*, and *integrationists*. Antagonists see the relationships between the sciences and ethics as primarily negative, whereas integrationists view them in a mostly positive light. The compatibilists, on the other hand, although they agree with the other two positions that the connections between the sciences and ethics are many and substantial, believe that in the end the effect of the two on each other is relatively neutral. Integrationists, however, contend that the connections between the sciences and ethics are multiple and substantive. Scientific naturalistic accounts of moral agency are integrationist.

For instance, consider something like the problem of freedom versus determinism. The separatist might argue that what scientists say about how our actions are determined by causes is one thing and what ethicists say about our moral actions being free is another thing; we should not confuse the two by thinking that one contradicts or supports the other.

The antagonist believes that freedom and determinism are antithetical and that scientific and moral points of view are incompatible. The compatibilist believes that freedom and determinism are compatible and that what the sciences might tell us about the determinants of our moral agency does not detract from our freedom. However, compatibilists do not find that the sciences provide significant support for the existence of the freedom required by moral agency or that they increase one's understanding of it. The integrationist, however, claims that knowledge of biology and psychology, for instance, supports our belief that as moral agents we are free and increases our understanding of what that freedom is and how it is that we are free.

In general the antagonist believes the same sort of thing about such problems as those of fact versus value and information versus meaningfulness. The scientist's facts and information have negative import for moral values and meaningfulness. So, whereas separatists avoid conflicts by not finding significant connections between the sciences and ethics, antagonists cannot avoid finding significant clashes between the sciences and ethics, because they believe that there are significant negative links between the two. Compatibilists find no significant conflicts between the sciences and ethics with respect to the problems of freedom and determinism, fact and value, and information and meaningfulness, but they contend that overall these disciplines do not contribute significantly to each other's enterprises. Integrationists argue that knowledge of the sciences will help in the solution of these problems.

A priori, any of the alternatives to a scientific naturalistic philosophical approach to understanding moral agency could take any of the stances on the relationships of the sciences and ethics. Historically, however, it has been the case that these approaches have adopted separatist, compatibilist, or antagonist positions. Scientific naturalists have urged an integrationist perspective. I shall not pursue these abstract possibilities or their various historical instantiations further. Rather, I now turn to a positive account of my scientific naturalistic model, beginning with an examination of the areas of investigation concerning moral agency in which it makes substantive use of the sciences.

In contrast with compatibilist interactionist positions, and, a fortiori, with separatist and antagonistic interactionist positions, my integrationist approach makes substantive use of the findings and theories of the sciences. We can distinguish six increasingly controversial, substantive dimensions in which the sciences, in particular biology and psychology, might make a contribution to ethics and so to our understanding and explanation of moral agency. These are informational, explanatory, critical, normative, metaethical, and meaningful. Let us consider each.

SIX INTEGRATIONIST HYPOTHESES ON THE RELATIONSHIPS  
 BETWEEN AN ACCOUNT OF MORAL AGENCY AND THE SCIENCES  
 OF BIOLOGY AND PSYCHOLOGY

Although all of the sciences are candidates for providing relevant input to the enterprise of understanding morality, I focus on biology and psychology. The scientific naturalist does not have to assume that these sciences alone will in the long run make the most important contributions to a scientifically based morality. Indeed, it is quite likely that the social sciences, like sociology, anthropology, political science, and economics, will make much more significant contributions than biology and psychology. But the former are arguably less well developed than the latter, so their potential contributions to an understanding of morality are less clear. Whether or not this is so, the task of exploring the connections between biology and psychology and ethics is sufficiently large and important for study in itself. Nevertheless, on the scientific naturalistic approach I am taking, both the natural and social sciences have a necessary contribution to make to a naturalistic account of morality.

Focusing on the potential contributions of biology and psychology, we can discern the following sorts of connections between those disciplines and ethics:

1. *Informational Connection Hypothesis*: Because of the relevant information about the circumstances and conditions of moral action that they can offer, the biological and psychological sciences can provide us with factual information important in ethical decision making and for ethical understanding.

An informational connection provides a minimal kind of link between biology and psychology and values. We can suppose that our abilities to act morally, for instance, our abilities to form adequately justified moral beliefs and proper moral motivation, arise independently of genetically and psychologically based capacities and dispositions discovered by the biological and psychological sciences. So, for instance, we believe that we have a duty to care for our children, and we may justify that duty on the basis of our religious beliefs, and we are motivated to carry out our duties because of our religious beliefs. But, notwithstanding, we may need to inform ourselves of a nurse's biologically based medical knowledge to find out what sort of diet is best for a sick infant. This sort of need for factual knowledge to aid ethical decision making and action is generally accepted even by separatists. It is a minimal sort of connection between ethics and the sciences, and a noncontroversial one.

2. *Explanatory Connection Hypothesis*: Our moral capacities have genetic, developmental, behavioral, and cognitive components. Thus, biology

and psychology provide descriptive, explanatory, and predictive knowledge concerning the acquisition, development, maintenance, use, change, or extinction of the cognitive, affective, and behavioral capacities that are employed in ethical action.

For example, suppose developmental psychology tells us about the emergence and development of empathy in infants and children. Because the obligation to refrain from harming others is a prototypical kind of ethical duty, we can see how having an empathetic disposition may contribute to the carrying out of this duty. If children have a tendency from early years to be empathetic, one might argue that such a tendency has a genetic basis to it. With a typical social and caregiver environment during infancy, we would expect a child to react negatively if another is hurt, to be sad or cry and spontaneously seek to help the person who has been hurt. As the child grows, he or she may learn through reinforcement, modeling, and instruction how to effectively help another person who is in distress. If all this is plausible, and it does seem to be, then we would expect biology and psychology to tell us about the evolutionary, developmental, learning, and social cognitive factors that contribute to a person's having and being able to effectively use his or her empathetic capacities to achieve moral ends.

The existence of these sorts of links between ethics and the biological and psychological sciences is more controversial than the informational connection but still not of a kind that will stir up the opposition of a lot of separatists. One reason for this is that ethical theorists hold that one's obligations are limited in some way by one's capacities. People are not obliged to do what, through no fault of their own, they are unable to do. Put very generally, human nature sets limits on what our obligations are by limiting our capacities. Even separatists are usually willing to admit that biology and psychology, or at least the latter, tell us something about the nature of moral capacities as well as how they come about and are maintained, used, and lost.

3. *Critical Connection Hypothesis*: An account of the biological and psychological bases of moral agency can effectively critique the claims of common sense or of other moral theories about human moral capacities and the nature and function of morality.

Assume that sociobiologists are able to demonstrate that there are genetically based tendencies in human beings to help relatives at some cost to the helper, and that social psychologists demonstrate that many people are motivated to help strangers without reward. These findings, based on biology and psychology, could be used to criticize claims that human beings are fundamentally selfish and so should *not* be expected to act altruistically because they cannot. This sort of critical connection between biology and psychology and ethics is clearly a consequence of the explanatory connection. As such, it is as relatively controversial or uncontroversial as that connection.



However, besides criticisms of ethical claims based on contentions that the view under critique is using an incorrect, incomplete, or inadequate account of human capacities, and so requires what cannot be delivered or does not demand what can and ought to be, ethical contentions also can be criticized on the basis of normative theories that themselves are maintained to have a basis in our biology and/or psychology. This normatively based critical connection is as controversial as the claim that there is a normative connection between the sciences and ethics. It goes as follows:

4. *Normative Connection Hypothesis*: From a knowledge of our moral capacities, and on the basis of knowledge from other sources, for instance, perceptual and social, we can formulate some general prima facie normative principles about what it is good for us to do morally, what is morally permissible and impermissible, and what is morally obligatory.

So, one might argue on the basis of what we know from biology and psychology about the human family that parents have a prima facie obligation to care for their children until they are old enough to provide for themselves. *Prima facie obligation* means that the obligation is supposed to hold unless there is some good reason why it should not. Of course, we could with good reason say that parents realize they have this duty without studying any psychology or biology. And that is correct. But this helps to bring out the point that the sciences are not some esoteric mode of knowledge completely distinct from ordinary knowledge. To return to our example, what biology and psychology can tell us about the details of our social being and familial ties, as well as about the dependence of human infants and children on their parents for development of their capacities, often, but not always, confirms and refines our experiential knowledge. Nevertheless, this sort of link remains very controversial, because the kind of connection it hypothesizes bridges the so-called fact-value gap. Showing how this can be done is a major part of any complete scientific naturalistic account of moral agency.

Further support for this hypothesis comes indirectly from arguments for the metaethical connection hypothesis, which I formulate as follows:

5. *Metaethical Connection Hypothesis*: If the explanatory hypothesis is substantiated, we infer that the capacities of moral agency, identified in the explanatory connection hypothesis, provide us with relatively reliable mechanisms for achieving moral goals. So we can use them in attempting to justify moral beliefs, motivations, and actions and in trying to understand the nature and function of morality. In addition, if the explanatory hypothesis is substantiated, we can give a naturalistic ontological account of moral values.

For instance, it has been argued that we can discern from evolutionary biology that morality plays a role in helping us achieve biological fitness

and survival. From psychology we might find that morality plays a role in helping us achieve nonbiological ends that we have developed in the course of our individual lives and our sociocultural evolution. And if biological fitness and survival or these nonbiological ends are either intrinsically or instrumentally morally valuable, we can justify actions promoting them as themselves morally required or permissible. Such justifications would appeal to the fact that the actions in question result from capacities that reliably produce the right sorts of results. Just as separatists part company with scientific naturalists with respect to the possibility of normative connections between biology and psychology and ethics, so too do they with regard to the possibility of a metaethical connection.

Finally, we have the meaningfulness hypothesis:

6. *Meaningfulness Hypothesis*: An account of the biological and psychological bases of moral agency makes an important contribution to the vision of a meaningful human life, one connected in fulfilling ways to other humans, nonhumans, and the environment.

The perspective of the long history of the evolution of life and the realization of both the richness and the variety of living things and the place of human beings in that history may be a source of meaning and inspiration to human moral agents, giving significance to their attempts to lead moral lives. So, too, may a vision of an expanding network of human beings living in harmony with each other and with nature. Integrationists claim that both biology and psychology have the resources to contribute to the delineation and refinement of these meaningful aspects of human and natural history and human beings' place within it, thus making the life of moral agency meaningful.

I have thus far attempted to characterize the general contours of a scientific naturalistic approach to the study of moral agency and its relationships to the sciences, in particular to biology and psychology. I now outline my substantive account of human moral agency.

#### A BIOLOGICALLY AND PSYCHOLOGICALLY BASED PHILOSOPHICAL ACCOUNT OF MORAL AGENCY

Scientific naturalist philosophers concerned with issues of moral agency differ in their answers to the four central questions concerning moral agency: what moral agency is, how it is acquired, how it is activated, and how the central activities of moral agency—that is, moral belief formation, moral motivation, and moral action—are justified, especially the last question.<sup>1</sup> I present portions of my own brand of scientific naturalistic ethics, one that allows for a greater role for the sciences in ethics than envisioned by a substantial portion of integrationists (see Rottschaefer 1998a; 1998b).

In answering the question of what moral agency is, any model of moral agency must meet two relevance criteria.<sup>2</sup> Accounts of the nature of moral agency must be functionally morally relevant, that is, they must build in moral responsibility; and they must be substantively relevant, that is, they must concern moral matters. Moral responsibility requires sufficient knowledge and freedom. To meet the substantive criteria for what counts as moral agency, the agency must be about those sorts of actions that are considered to be within the moral realm, as opposed, for instance, to the realms of prudence, etiquette, aesthetics, and law. Actions included within the moral realm include those actions concerned with (1) the goods of human flourishing, for instance, food, shelter, clothing, safety, companionship, and the development of intellectual, creative, practical, and social capacities, and (2) the goods of the human community, for instance, social and distributive justice and moral rights.<sup>3</sup>

On one common-sense view of moral agency a person acts in a morally correct fashion when he or she acts on the basis of adequate moral beliefs correctly applied to a particular situation. On this view, then, moral agency seems to require *morally cognitively motivated agency*. The model of moral agency I am proposing includes four functional levels: (1) a *base level*, constituted by evolutionarily acquired and behaviorally learned capacities and tendencies that incline the agent to act morally in given situations; (2) a *behavioral level*, consisting of a set of moral beliefs and desires that are the immediate sources of actions and that are influenced by both base-level and higher-level components; (3) a *reflective level*, comprising higher-level beliefs and desires, including moral norms or their equivalents, that influence the behavioral-level beliefs and desires; and (4) a *self-referential level*, consisting of conceptions of the self, including the self as moral agent, that motivate the use of moral norms and, indirectly, moral action.<sup>4</sup> In the complete model that I am proposing, then, I conceive of moral agency in its fullest extent as cognitively and morally motivated agency that is both reflective and self-referential. However, I shall not argue that for an action to be functionally moral it must engage all four levels of moral agency.<sup>5</sup>

My proposed model attempts to answer the question of what moral agency is. However, here I do not show how it meets the functional and substantive relevance criteria. I merely note that I believe that the functional relevance criterion can be met by a soft-determinist account of freedom. Thus, I understand freedom not in the libertarian fashion as the absence of determinism but rather as the absence of a certain sort of determinism, roughly that of externally compelling factors and of internal compulsion. Positively, freedom requires the operation of a cognitive-evaluative system. In the discussion below, I illustrate the operation of my model of moral agency in one substantive area of morality, that of altruistic activity.

Working within the context of my proposed model of moral agency, I now sketch answers to the questions of acquisition, activation, and adequacy.

## THE ACQUISITION AND ACTIVATION OF MORAL AGENCY

Although I do not here discuss in any detail the scientific support for my account of the acquisition and activation of each of the levels of moral agency, I do indicate something about the major scientific theories and findings upon which my model is built by briefly discussing some supportive theories and results from sociobiology, socioemotional development in developmental psychology, and social cognitive theory in psychology.

SOCIOBIOLOGY AND THE BASE LEVEL OF MORAL AGENCY:  
EVOLUTIONARILY BASED MORAL CAPACITIES

The first of the four functional levels of moral agency is the *base level*. As I have indicated, it is composed of evolutionarily acquired and behaviorally learned capacities and tendencies that incline the agent to act morally in given situations. Support for the presence of evolutionarily acquired moral tendencies comes from sociobiology, developmental biology, and developmental psychology. Let us consider the support from sociobiology. To do so, we can focus on the contribution of sociobiology to the solution of what has come to be called the problem of biological altruism.

The theory of evolution by natural selection tells us that the fittest organisms in a population survive and reproduce. The fittest organisms are the ones that have traits that enable them to do better than their competitors. An altruistic phenotype by definition inclines its bearer to benefit another at some cost to itself. It seems to follow that if altruistic traits were to arise in a population, whether by mutation or immigration, the organisms that possessed them would not, given the presence of natural selection, survive to reproduce—or, if they did, they would not reproduce as well as selfish organisms. Thus, in the short or long run, altruistic traits would be eliminated from the population. But biologists have found numerous examples in human beings and animals of apparently altruistic behavior: altruism of parents toward offspring, altruism toward relatives, and reciprocal altruism between both relatives and nonrelatives. This apparent inconsistency between theory and data constitutes the biological problem of altruism. To solve this problem, it seemed that biologists must either surrender any attempt to provide a biological explanation of altruistic behavior or admit that evolution by natural selection does not provide a completely adequate account of the evolution of all traits and search for a new account.

The selfish gene theory of altruism attempts to save evolution by natural selection by making the altruism of organisms a strategy of the selfish gene for increasing its own fitness. If the altruistic behaviors of organisms, whether toward direct descendants by parenting or toward indirect descendants through what biologists call kin altruism, are to be evolutionarily

successful, they must benefit the genes of their possessor. So too for any altruistic behaviors that are directed toward nonrelatives. In organisms that have cognitive apparatus and operate on the basis of motivations and intentions, it may be the case that some motivations are consciously altruistic in the sense that they are directed toward the benefit of the other at some cost to the benefactor. But such altruism is really only a means to further the survival and reproduction of the genes that create it. Thus, according to the selfish gene theory, the ultimate explanation and solution of the problem of altruism lies in selfishness. Morality is in the last analysis a means for the promotion of self-interest. Or so the story goes.

My model is a critique of the selfish gene theory and may be called the altruistic gene theory. It postulates that moral capacities are acquired in evolution. Consider the answer of the altruistic gene theory to these four central questions: First, whose genes are the beneficiaries of biological altruistic behaviors? Answer: It is not, as the selfish gene theory claims, the individual genes of the altruist that benefit from their possessor's altruism, but copies of those genes. Second, what sorts of behaviors can evolution by natural selection explain? Answer: In opposition to the selfish gene theory, which claims that biologically based altruistic behaviors are fundamentally selfish, the altruistic gene theorist argues that, strictly speaking, only behaviors of organisms that are altruistic in the ordinary sense of the term are explainable in terms of evolution by natural selection. Third, at what levels does selection for traits occur? Answer: Contrary to the selfish gene theorist's claim that genes are the only units of selection, the altruistic gene theorist argues that there can be multiple units of selection, for instance, genes, organisms, and communities. Fourth, is there evidence for the genuine altruism of biologically selfish genes? Answer: Yes.

Let us consider each question and answer in more detail. First, whose genes are the beneficiaries of biological altruistic behaviors? The correct answer is not, as the selfish gene theorist implicitly presupposes, the genes of the individual altruist, but rather, as the altruistic gene theorist claims, copies of those genes, that is, the genes of the beneficiary. The altruistic gene theorist argues that proponents of the selfish gene view have failed to formulate and use an empirically based distinction that makes all the difference in the world for the story of altruism. Consider the fact that there is an ontological difference between parents and offspring. Even though they share some genes of the same *type*, the sets of genes that constitute the genomes of the mother and the father are not the set of genes that constitute their offspring. They are numerically distinct. That is, they differ as individuals. To put it technically, there is a distinction between gene tokens and gene types.

What evolutionary theory aims to explain is the change in gene frequencies over time. Thus, the traits and behaviors that make genes winners in the evolutionary contest do not make *numerically identical* genes winners.

Individual genes are not eternal beings. Genes make replicas of themselves in somatic and sex cells. If they have the right stuff, it is the replicas of genes in the sex cells of the parents that are the evolutionary winners. That is, successful genes are genes that survive to reproduce not themselves, but more or less accurate copies of themselves. Successful genes are the ones that *give* of their right stuff to others, their descendants, by passing on their adaptive characteristics to them (see Rolston 1999).

Look at it from the macroscopic level. I do not pass *my* genes on to my children. I pass on more or less accurate *copies* of my genes. From an ordinary everyday perspective, becoming and being parents, parenting, is a morally good and commendable activity. Caring for one's children is considered to be an ethical duty. It is altruistic in the minimal sense that it benefits someone else at some cost to oneself. But, according to the technical biological use of terms, it is selfish. It is selfish in the sense that it benefits my *direct* descendants rather than my *indirect* descendants or the descendants of nonrelated individuals. Biological selfishness, technically defined, turns out to be a major part of everyday altruism and a significant element of morality.

Turning to our second question—What does evolution by natural selection explain?—we find further support for the altruistic gene theory. Evolutionary theory does not and cannot explain selfish behavior as it is commonly defined, that is, behavior that benefits only the actor and no one else. Evolutionary theory can explain behavior that benefits the actor only insofar as it is directed toward the survival of that individual as it functions as a means for the reproduction of others. Benefits to the individual are explained only as means to reproducing offspring. Thus, evolutionary theory, in particular the theory of natural selection, cannot explain selfish traits and behaviors if selfishness is understood in its ordinary sense of a behavior or trait that benefits only the individual possessor as an end in itself. If there were or are any such biological traits, introduced into a population by mutations, for instance, they would not survive in the long run, because an organism with such traits would not reproduce at all or would reproduce less than its competitors. The theory of natural selection cannot explain ordinary, everyday selfishness!

Now for our third question: At what levels of biological organization does selection for traits occur? This is the so-called units of selection problem. The selfish gene theory has been associated with a particular solution to the question about the levels at which selection acts. Natural selection theory tells us that various factors in an environment are more or less significant for the survival and reproduction of an organism. It is the interaction of the organism with its living and nonliving environment that constitutes the selective process that determines the pattern of gene representation in subsequent generations. The question of level of selection has to do with what properties of the organisms are relevant to the selection

process in the sense of being causally active in that process. These properties can be properties of the genes themselves, or the individual organisms, or the groups of which organisms are members, or even larger groupings. Thus, to stick with these examples, there can be selection for properties at the genic, organismic, or group level.

Evolutionary theorists are coming more and more to believe that selection happens at all these levels (Sober and Wilson 1998). Thus, if selection occurs between groups, one factor that may influence this selection process is the relative cohesiveness of the groups. Those groups whose members are more altruistic, who are more ready to sacrifice individual interest for group interests, may be evolutionarily better adapted and so have a fitness advantage. Consequently, they may survive and reproduce more than groups composed of more selfish individuals. Group selection is gaining in evolutionary thinking, but I must emphasize that it is still a controversial theory. Nevertheless, it offers another theoretical route to the understanding of biologically based altruism, one based on group selection as opposed to the altruism of parental care, kin altruism, and reciprocal altruism, which can be understood as based on individual selection, that is, selection on the level of the organism.<sup>6</sup>

Thus, if we make the reasonable assumption that altruism is a part, though indeed not the whole part, of substantive morality, there is good reason from a naturalistic perspective based on the best current evolutionary theory to believe that altruism should be found in creatures with the capacity for it. We can conclude that current evolutionary theory has solved the problem of altruism; indeed, it has shown that individual selection can result in parental, kin, and reciprocal altruism toward nonrelated individuals and that group selection can also result in altruism directed toward nonrelated individuals.

This brings us to our fourth question: Are the theories of evolutionary biology about genetically based altruism supported by the evidence? The short answer is Yes. Sociobiologists and evolutionary psychologists use evidence supporting the following sorts of hypotheses to argue for the presence of evolutionarily based altruistic capacities in human beings:

1. *Interspecies Comparison Hypothesis*: (1) If some human altruistic behaviors are genetically based, then we would expect to find these behaviors shared by our closest noncultural primate relatives. (2) If all human altruistic behaviors are culturally based, then we would not expect to find similar altruistic behaviors in primate relatives that do not have culture.
2. *Cross-Cultural Comparison Hypothesis*: If some human altruistic behaviors are genetically based and not just culturally based, then we would expect to find some common altruistic behaviors in all human cultures.

3. *Social Experiment Hypothesis*: If there are genetic limits to certain altruistic behaviors, then attempts by societies to break or exceed these limits should ultimately fail.
4. *Genetic Basis Hypothesis*: If some human altruistic behaviors are genetically based, then, as technology becomes available, we should expect to find the genes that determine these behaviors.

Although I will not go into it here, there is evidence to confirm Hypotheses 1 and 2, and thus to claim that there is reason to support the existence of evolutionarily based parental altruism, kin altruism, and reciprocal altruism toward both relatives and nonrelatives (Boehm forthcoming; Flack and de Waal forthcoming; Rottschaefer 1998; Sober and Wilson 1998; de Waal 1996).

To summarize, evolutionary theory has solved the problem of altruism; indeed, it has shown that individual selection can result in parental, kin, and reciprocal altruism toward nonrelated individuals and that group selection can also result in altruism directed toward nonrelated individuals. Moreover, this solution does not make altruism a mere means for selfish gain, if that term is understood in its ordinary sense. Although altruism is not constitutive of all of morality and moral agency, it is certainly a part of morality. Thus we can conclude that there is reason to support the existence of evolutionarily based substantive morality. Our biology has provided us with some evolutionarily based capacities for doing the right thing.

If natural selection has given us some altruistic capacities, can we say the same for development?

DEVELOPMENTAL PSYCHOLOGY AND THE BASE  
AND BEHAVIORAL LEVELS OF MORAL AGENCY:  
DEVELOPMENTALLY BASED MORAL CAPACITIES

To illustrate how biology and psychology can tell us about developmentally based moral capacities, I want to focus on our capacity for empathy and empathically based helping. During a good portion of the twentieth century, the study of moral development in children was dominated by three very different traditions in psychology: the Freudian, behaviorist, and cognitive developmental. In their own way both Freudians and behaviorists emphasized the noncognitive and nonconscious bases of our moral agency. In different ways both traditions viewed morality as the imposition of an external agent. For Freud, moral agency derives from the superego, which is a socially imposed internal constraint on the selfish instincts of the id. Behaviorists argued that moral agency was the result of operant conditioning by parents, peers, and society.

Jean Piaget and Lawrence Kohlberg, the two giants of the cognitive developmental tradition, and their followers emphasized the rational and



conscious character of moral agency and argued that moral agency is the natural result of human development under the influence of proper environmental stimulation. Kohlberg's is the best known of the cognitive developmental theories. He has argued that moral development is the result of a progression through successively more adequate moral stages defined by moral cognitive structures which guide our actions. In its classical formulation it postulated six stages of moral development.<sup>7</sup> The stages are distinguished by the moral motivation that guides action. Stage 1 is egoistically motivated and based on external rewards and punishments. Stage 2 represents an enlightened egoism. Stage 3 makes social conformity within the circle of a person's immediate contacts the central motivational factor. In Stage 4 that circle is expanded to the society in which a person is located. Stage 5 is characterized by philosophical utilitarianism, in which the welfare of the greatest number of persons is primary. The highest stage, Stage 6, considers not only the greatest happiness of all persons but what is due to an individual as a person. These stages are not age dependent, nor does Kohlberg claim that all persons reach the highest stage. In fact, most people, according to Kohlberg's research, attain only stage 4. The stage features define aspects of the moral cognitive structures. The stages are *universal* insofar as they are found in all cultures. They are also *invariant* and *irreversible*, that is, there is no stage skipping and no movement backward in moral development. The stages are also *integrative* in that one stage builds on the previous one. A later stage logically contains the concerns of the previous one. Thus, for example, the conventional stages, 3 and 4, include the focus of pre-conventional stages, 1 and 2, the self, within their larger scope of concern for the social group or society. Finally, the progression through the stages reflects an advance in *moral adequacy* from a least to a more adequate moral stance. All these features of moral growth show, according to Kohlberg, the guiding hand of nature in which the role of nurture and the influences of the social environment play a facilitative but subordinate, nondeterminative role.<sup>8</sup>

The theories and approaches of all of these traditions have undergone substantial change, but it is fair to say that the cognitive developmental views had come to a position of prominence by the 1960s.<sup>9</sup> The cognitive developmental view has the advantage of corresponding more closely to our reflective common-sense views about the nature of our moral agency. In addition, it seemed to have some substantial empirical support. But despite these advantages, Kohlberg's theory suffers from some significant empirical and conceptual problems (Flanagan 1984; Mischel and Mischel 1976; Rottschaefel 1991). The main problems concern doubts about (1) the existence of distinct stages; (2) the existence of the sixth and highest state of moral development; (3) the existence of the stage properties of universality, irreversibility, integration, and adequacy; and (4) the inability to explain the connections between moral reasoning and moral action.

The problems with Kohlberg's theory are serious enough that alternatives should be considered. For an alternative I turn to a model that incorporates behaviorist features at the base level of moral agency, work from the field of socioemotional development at both the base and behavioral levels, and cognitive social learning accounts at the behavioral, reflective, and self-referential levels of agency.

Empirical findings from the study of socioemotional development have not only revealed some of the inadequacies of Kohlberg's theory but have also provided the basis for an alternative to Kohlberg's theory. Because in their view moral agency is essentially a cognitive, conscious, and rational activity, cognitive developmentalists focused their studies on children and adolescents rather than infants and toddlers. They did not expect that morality or moral agency would be found in infants and very young children, those younger than five or six years. A number of factors, however, have led developmental psychologists to the study of moral development in infants and very young children. The results have been truly surprising and support claims that we have evolutionarily based moral capacities.

These results concern the moral learning that takes place in ontogeny. We often oppose innateness to learning and genes to environment. This opposition is misleading insofar as both phylogenetic evolution and ontogenetic learning require both genes and environment. Indeed, it is the case that much learning maintains and enhances capacities that are not genetically based. But some genetically based traits—for instance, as I shall argue, the capacity for empathic distress—require environmental input, some of which clearly involves learning.<sup>10</sup> Thus, the nature of the environmental input will affect the final form of the genetically based trait. Evolutionarily stable traits are the results of stable environments. Consequently, the reliable production of the capacities for moral agency has required the presence of social learning environments in which caregivers have provided the help necessary for the development of such capacities. Normal developmental environments for humans are such that they regularly produce agents with some moral capacities.

Let us consider briefly work that has emerged in the last twenty years or so in the field of developmental psychology known as socioemotional development. Developmental psychologists have demonstrated that there are correlations between empathy, sympathy, and prosocial behavior. Building upon these results, developmental psychologist Martin Hoffman has worked out a theory of moral development that focuses on the central role of empathy in moral agency. His emphasis on empathy in moral development is especially important for several reasons. First, as Michael Bradie (1994) has demonstrated, there is a long philosophical tradition that has looked to the emotions and our cognitive/affective capacities as the originative sources of our moral agency. Second, Charles Darwin (1871) gave pride of place to empathy in his own account of the development of what

he called the moral sense. Third, there seems to be growing evidence for an evolutionary basis for basic emotions such as happiness, sadness, disgust, anger, surprise, fear, and interest (Campos, Barrett, Lamb, Goldsmith, and Stenbey 1983; Griffiths 1997). Fourth, empathy seems to be an important source for altruistic behavior. Although I will be focusing on empathy as an evolutionarily based moral capacity, we should keep in mind that our moral agency is more than likely a complex system of a number of evolutionarily based moral capacities and that the base-level system of moral agency is composed of not only evolutionarily based moral capacities but also learned nonevolutionarily based moral capacities. I concentrate specifically on empathic distress as a kind of model system for exhibiting the ontogeny of an evolutionarily based moral capacity.

Although theorists differ about how to define *empathy*, a number of developmental psychologists, including Hoffman, take empathy to be an affective response more appropriate to someone else's situation than to one's own (Barnett 1987).<sup>11</sup> Empathy is an *emotional* state of the empathizer that is the same or similar to that of the person with whom one is empathizing *because* that person has that state.<sup>12</sup> Sympathy adds to that the condition of concern for the other. Empathy and sympathy, as we shall see, often result in altruistic tendencies and actions. By altruistic I mean what psychologists often intend by the term *prosocial*. Prosocial intentions and actions are those that have the benefit of another as their object. This leaves open the degree of self-regard in the agent's actions and intentions.

Using this understanding of empathic distress, Martin Hoffman argues that there are a series of developmental levels of empathy.<sup>13</sup> He distinguishes four levels of empathic development: global, psychologically undifferentiated, immediate, and extended.<sup>14</sup>

In Hoffman's view, one of the precursors of global empathy is neonate responsive crying. This is the highly confirmed phenomenon that newborn infants cry in response to the crying of other infants. Although there is some question in the literature whether neonate responsive crying ought to be classified as a precursor of empathy, it does seem to be a distinctive response, not a startle reaction or a reaction to aversive noises generally. Other frequently cited precursors of empathy are affective synchrony in mother (or caregiver)-infant play, which begins when the infant is two to three months old, and social referencing, which begins when the infant is ten to twelve months old.<sup>15</sup>

The first level of empathic development in Hoffman's theory, global empathy, occurs in children less than one year of age. It is a response to another's pain as if the one responding were in pain. Hoffman postulates that at this stage the infant has not self-distinguished from other selves. As psychologists put it, there is no object permanence of other persons. So the painful state of the other is also the painful state of the infant. Infants also indicate behaviorally a recognition of a departure from a standard, but

what kind of cognitive state, if any, corresponds to such behavioral signs is not clear. There is no evidence of any assessment of the source of the distress in the other, and, of course, there is no helping response.

Undifferentiated empathy, the second level of empathic development, occurs in infants aged one to two years. These infants have mastered the notion of person permanence, so they discern the distinction between self and other. But they have not yet achieved an understanding that the inner states (beliefs, desires, and feelings) of the other person are independent of their own. So they respond to the other's distress by doing for the other what would be helpful to themselves. A child may, for instance, bring his or her own mother to help another child in distress even though the latter's mother is present. Thus, there is evidence that infants grasp cognitively that others are distinct from themselves and that the others have psychological states, but they do not understand that these are different from their own when in distress. They also seem to grasp in some manner that there is a harm, a departure from the standard or what ought to be, but it is not clear what this understanding amounts to. Nor do children at this age seem to discern what the source of the harm is. They evince some tendencies to help, though the helping may not be appropriate.

By the time children are three or four years old, they recognize the distinction in psychological states and begin to offer appropriate help. At this time they have reached the level of immediate empathy. In later years, empathy is extended in the sense that others are recognized to have experiences beyond their immediate situations, and these experiences too can become bases for empathic response. Indeed, another's general condition can invoke empathy. In addition, empathy can extend beyond an individual to an entire group. Thus, the ontogeny of empathy reveals patterns of generalization with respect to the conditions of an individual person and from individuals to groups of individuals.

The empirical literature provides support for Hoffman's hypothesis on the development of empathy.<sup>16</sup> But his account is certainly not established; it remains a persuasive working hypothesis about the ontogeny of empathy, sympathy, and prosocial/altruistic behavior.

I have discussed the development of the capacity for empathic distress and empathically motivated helping as an example of how developmental psychology can help us understand our learned capacities for doing the right thing. Empathy is a particularly interesting example, because it illustrates the intimate connection between our evolutionary endowments and our developmental and social learning environments. The findings about the development of empathy, like those about evolutionarily based altruistic tendencies, of which empathy may be a manifestation, show us that our moral capacities are to some extent the product of both our evolutionary history and our ordinary developmental learning history. If this is so, the scientific disciplines of biology and psychology do more than just provide

us with information about how to act in order to do the right thing. They also provide us with an understanding of the very capacities by means of which we are able to do the right thing. They confirm the explanatory connection between the sciences and ethics that I spoke of earlier.

But clearly moral agency involves more than evolutionarily based altruistic capacities and ontogenetically fashioned empathic distress. So I turn now to a brief consideration of cognitive social learning theories, in particular the social cognitive theory of Albert Bandura, to illustrate their contribution to the understanding and explanation of moral agency as constituted by the behavioral, reflective, and self-referential levels of my model of moral agency.<sup>17</sup> Cognitive social learning theories of agency provide a way of getting around some of the difficulties concerning the acquisition and activation of moral agency present in Kohlbergian cognitive developmental theories of moral agency.

#### SOCIAL COGNITIVE THEORY AND MORAL AGENCY: SOCIALLY/CULTURALLY BASED MORAL AGENCY

Any adequate scientific account of moral agency must include reference to our cognitive capacities.<sup>18</sup> The question of how we acquire and put into action our moral capacities cannot be satisfactorily answered by appealing only to evolutionarily based and learned, but noncognitive, moral capacities. In particular, the Skinnerian claim that the science of operant behavior is the science of values fails as a complete explanation of moral agency because it does not take account of the cognitive features of human agency. This conclusion rests in large part on the findings of psychologists who have shown that cognitive factors play a role in the explanation of human behavior.

This turn to the cognitive in psychology is not a phenomenon isolated to critics of behaviorist theory; it is part of what has come to be called the cognitive revolution in psychology, which has taken two forms, representational and agential. The former focuses on human knowledge-gaining capacities and achievements and has been the focus of study in such areas as perception, memory, imagery, language, thought, and problem solving. The latter is concerned with issues of human action and has emerged in the areas of learning, motivation, personality, and social and abnormal psychology. Each in its own way stresses the cognitive capacities of human beings in coming to know the world and learning to act in it. The representational revolution often draws on theories taken from computer science and information processing theories. In doing so, it has begun to move away from folk-psychological conceptions of epistemic capacities and processes. In place of talk about beliefs and desires, it has begun to substitute notions like information, schemata, input and output buffers, and central processors.

The agential revolution has to a larger degree retained folk-psychological concepts of the cognitive. One branch of the agential revolution is a set of theories often called cognitive social learning theories or cognitive behavioral theories (Dobson and Block 1988; Erwin 1978; Kazdin 1978). Originating to some extent in the behaviorist tradition, these theories have incorporated cognitive variables as independent variables into their theories of human agency. I employ the contributions of the agential revolution in psychology for understanding moral agency.<sup>19</sup> Specifically, I make use of Albert Bandura's social cognitive theory of agency to answer questions about the acquisition and employment of moral agency.

Bandura's social cognitive theory is concerned with human agency and motivation as such, but its prospects for explaining moral agency are good, I believe, because a theory of moral agency, including moral motivation, stands a better chance of being adequate if it fits within a more general theory of agency and motivation.<sup>20</sup> Moreover, Bandura's theory of agency is an apt candidate for understanding moral agency because it is empirically well supported and theoretically well articulated. Its prospects for extension, therefore, appear brighter than those of less well established and elaborated theories. If one of the best of the cognitive social learning theories shows itself to be a poor candidate for solving the problems of moral agency, then the prospects for cognitive social learning theories generally and the agential revolution in psychology are significantly diminished. I shall briefly describe Bandura's social cognitive theory of agency and the way it accounts for how moral agency is acquired and put to work.

Like other cognitive social learning theories, and in contrast to behaviorist or neobehaviorist social learning theories, Bandura's social cognitive theory emphasizes the role of cognition in bringing about behavior (Bandura 1977b; 1986). Two further features characterize his theory—the doctrine of reciprocal determinism and the theory of the self-system. Reciprocal determinism is the view that environmental, behavioral, and cognitive factors interact with each other in effecting behavior (Bandura 1978; 1986). Philosophically, this feature of the self-system supports a soft-determinist account of freedom. The theory of the self-system postulates four major informational and motivational capacities and processes that play a central role in explaining human behavior: (1) the formulation and implementation of individual goals and standards of performance, (2) self-monitoring, (3) self-reactive influences, and (4) self-efficacy judgments (Bandura 1978; 1982; 1986). The theory maintains that to explain successful and unsuccessful performance, including the regulation of thoughts and feelings, one must postulate not only generically cognitive causes of behavior but also a set of cognitive processes which are self-referential.

Bandura's social cognitive theory emphasizes the role of cognition in both the learning and the bringing about of behaviors (1977b, 17–22, 67–72; 1986, 12–22, 116–22). He contends, on the basis of an accumulation

of research studies, that classical and operant conditioning are mediated by cognitive factors. Moreover, he shows that learning takes place not only by practice but also by modeling and symbolically. Bandura's account of learning through modeling (1977b, 22–55; 1986, 51–80) requires a rich assortment of cognitive factors: perceptual, attentional, and retentional. These include symbolic coding, cognitive organization, and symbolic rehearsal. For example, Bandura invokes a number of cognitive mechanisms to explain how people can disengage their moral principles and their application from behavior to either excuse or justify what would otherwise be an immoral action. One can seek to justify the performance of an apparent immoral action by using different or altered moral standards, comparing it favorably to other reprehensible alternatives, or trying to change the understanding of the action by reclassifying it in a benign fashion, for instance, by euphemistic labeling. Immoral activity is also justified by blaming or attributing dehumanizing characteristics to the victim. Exoneration is sought by ignoring, minimizing, or misconstruing the effects of one's actions. Finally, Bandura contends that one can find ways to diminish personal responsibility for an action by claiming to have been acting under orders or protesting that the decision was not his or hers but the group's (1977b, 155–58; 1986, 377–89).

Besides the role that cognition plays generally in human behavior, Bandura hypothesizes that the self-system plays a major cognitive and motivational role in human behavior (1977b, 193–208; 1978, 344–58; 1986, 22–46, chaps. 8 and 9). The self-system employs higher-level cognitive and motivational factors to regulate lower-level cognitive and motivational capacities. In a number of different studies concerning a wide range of complex activities such as eliminating phobias, inhibiting behaviors, regulating refractory behaviors, developing coping skills, reducing physiological stress, overcoming resignation and despondency, learning achievement strategies, increasing intrinsic interest, and choosing a career, Bandura and his colleagues have shown that the self-system plays an important role in the acquisition, maintenance, regulation, enhancement, and change of the behaviors, thoughts, and feelings necessary for accomplishing these activities and goals (1977a; 1977b, 79–93, 128–58; 1982, 122–47; 1986, chaps. 8 and 9).

In assessing the relevance of Bandura's social cognitive theory for explaining moral agency, it will help to consider the self-system in more detail. The self-system comprises two interrelated but distinct motivational subsystems, the *self-evaluative system*, consisting of competencies for forming self-standards and reacting evaluatively to efforts in attaining these standards, and the *self-efficacy system*, formed by competencies for self-efficacy judgments. (The *self-monitoring* subsystem plays a necessary informational role in the operation of each of these subsystems.) By means of the self-evaluative system, a person establishes goals and standards, both moral

and nonmoral (1977b, 128–58; 1986, 340–89). People acquire goals and standards by practice, modeling, and instruction. Their contents reflect not merely general, cultural, or social norms but also specific social influences and a person's own individual adaptation of general norms. Through evaluative self-reactions individuals acquire, maintain, regulate, and enhance their own behaviors by positively or negatively assessing their success in meeting self-prescribed standards (1977b, 130; 1986, 350–62). These reactions take the form of not only tangible or symbolic rewards or punishments but also self-approval or disapproval. The level of the self-motivation generated by self-reactions is a function of the type and value of the incentives and the nature of the performance standards. Although in Bandura's view self-administered tangible rewards or punishments do play a role in self-reaction, self-approval or disapproval can play a relatively independent motivating role. Thus, studies show that in areas that are particularly important for a person, a mere calculation of the cost-benefit ratios of the external consequences of behavior will not suffice to explain a behavior. Self-esteem effects must be included (1977b, 143–45; 1976, 145; 1986, 230, 240–41, 254–55, 348–57, 366, 374–75).<sup>21</sup>

In order to understand the role of self-efficacy judgments, we need to distinguish between perceived self-efficacy and outcome expectations. The former refers to a person's judgments about his or her capacity to perform the behaviors necessary for the achievement of certain outcomes. The latter concerns expectations about whether a given behavior will achieve the outcome toward which it is aimed. Thus, outcome expectation is dependent on factors external to, and sometimes beyond the control of, the agent, whereas self-efficacy judgments concern factors within the agent's control. Moreover, self-efficacy judgments are not global claims about oneself and one's capabilities but rather judgments about specific behaviors (Bandura 1977b, 84; 1986, 422–49; 1997).<sup>22</sup> Bandura and his colleagues have found in a number of studies dealing with a range of different behaviors that perceived self-efficacy better predicts actual performance than does previous performance. It predicts the kinds of tasks that will be undertaken and in what situations, and the effort that will be expended in preparation for, performance of, and persistence in a task. Thus, perceived self-efficacy is a major motivator of behavior.<sup>23</sup>

Given this account of the representational and motivational factors involved in human agency, we can distinguish several stages of agency. The first stage is that of the *self as a cognitive agent*. At this stage, to put it in folk-psychological terms, the self is conceived as an agent who operates on the basis of beliefs and desires whose actions are directed toward himself or herself as well as others.<sup>24</sup> This first stage of agency (which parallels the behavioral level of my model of moral agency) captures three functional criteria of moral agency, those of being a *cognitively motivated actor*. This stage reflects the emphasis on cognitive factors in social learning theories



generally and distinguishes them from behaviorist positions in which the self is merely a locus of forces.

The second stage of agency is that of the *self as a reflective cognitive agent* (corresponding to the reflective level of my model of moral agency). This second stage of agency includes a metalevel motivational system. One has higher-level beliefs and desires about one's lower-level beliefs and desires. The former influence behavior by influencing the latter. The higher-level desires take the form of self-evaluative standards. These standards include but are not confined to moral standards. The self-evaluative subsystem of the self-system is a metalevel motivational source in Bandura's account. Its standards extend beyond the moral realm to other areas of competence, but they serve to guide both beliefs and desires and, thereby, behaviors. I conclude that Bandura's social cognitive theory provides an account of moral agency that fulfills a fourth functional criterion of moral agency, that of a *morally motivated cognitive agency*.

In addition, both the self-evaluative and self-efficacy subsystems suggest a third stage of agency, that of the *self as a self-referentially reflective cognitive agent* (corresponding to the fourth level of agency in my model of moral agency). Principles and standards of self-evaluation are self-referential in several senses. First, they are self-referential to the extent that they are meant to govern and motivate the actions, beliefs, and desires of the person whose self they are. The principles and standards are also self-referential insofar as the self is active in their acquisition, maintenance, and application. But self-referentiality plays a further distinctive role in Bandura's theory of motivation and agency. For the goals and standards of behavior are not merely applied *by* agents *to* themselves but also become *personal* goals and standards. They embody ideals, ways in which a person aspires to be morally, aesthetically, cognitively, and socially. Self-satisfaction or dissatisfaction are motivating precisely because of their connection with these personal ideals. This aspect of self-referentiality differs from both the self-referentiality of applicability to the self and that of application by the self to itself insofar as moral standards, for instance, can motivate in an impersonal way, moralistically, or legalistically, or in a personal way, authentically. In the former cases the moral or legal system can apply to someone, and he or she can even self-apply it. But in the latter case the rules are the agent's in some fuller sense.<sup>25</sup> This case also may require an evaluation of the standards themselves and thus implies, in accordance with Bandura's views, that there are various ways of acquiring, maintaining, and applying standards. However, the self-referential character of self-evaluative standards refers not to a global self but to the ideal self in a specific behavioral area. The standards concern action, competencies, thoughts, desires, and feelings in some single area, for instance, playing a piano. As a result, the self of the ideal self is in the first instance an achieved, dynamic unity in a specific behavioral area. Most of Bandura's empirical

work has focused on such specific areas. However, one can plausibly extend Bandura's concept of the self to fit both diachronically and synchronically more extensive personal unifications, including that of the moral self.<sup>26</sup>

The self-efficacy system confirms the presence of the self-referential stage of agency but from a somewhat different perspective. I have been using simplified folk-psychological conceptions of the factors involved in human agency referring only to beliefs and desires. A more adequate model must include capacities and competencies. If Bandura is correct, judgments about one's capacities and competencies are also motivating. Judgments of self-efficacy motivate not only behaviors, the objects of second-level motivators, and desires and beliefs, the objects of third-level motivators, but also the acquisition, maintenance, and application of self-standards, what we have designated as one object of fourth-level motivators. For instance, negative self-efficacy judgments lead not only to behavioral failures but to dependency and self-devaluation (Bandura 1977b, 140–43; 1982, 140–41; 1986, 225–27, 359–60, 445–47). And self-efficacy judgments play a role in the selection of such self-standards as career choice and related competencies (Bandura 1982, 135–36; 1986, 430–35; 1997). Self-efficacy judgments, then, demonstrate reflexivity about both capacities and competencies and reveal self-referentiality. The self involved here is the potential self in contrast to the ideal self of the self-standards. Bandura's theory suggests, then, that human agency involves not only cognitive and reflective capacities but also self-referential capacities.

If I am correct about the relative superiority of cognitive social learning theories of moral agency to cognitive developmental theories, we can conclude that they provide the best current answers to three of the four major questions about moral agency, those of acquisition, action, and nature. What remains is the crucial question about how moral agency is justified.

#### RELIABLE MECHANISMS FOR MORAL AGENCY: A NATURALISTIC JUSTIFICATION OF MORALITY

Suppose that Jill is walking home from school, and as she passes by the yard of one of her neighbors she sees her friend's baby brother fall. No one seems to be around, and Jill sees that little Jimmy has cut himself badly. She immediately runs over, picks him up, and takes him in to his father. Suppose that Jill thinks that she did the morally right thing. She believes that helping people in need is the right thing to do and that she has an obligation to help people in need when she can. On what basis does Jill make these judgments? What justifies them? We also probably think that responding to Jimmy's need was the right thing to do. On what basis do *we* make these judgments? What justifies them?

These questions about justification are different from the ones about acquisition, action, and relevance. We are not asking how Jill acquired the

cognitive, emotional, and behavioral capacities she needed to discern that Jimmy was in need, to feel and believe that she ought to do something about it, to be motivated to do so, and to know what and how to do it. These are all questions about acquisition. Nor are we asking how she put these capacities to work, what moved her to act, and how it did so and why. These are questions of action. When we have answers to these questions, we can *explain* what she did. But we have not yet answered the question of whether she did what she ought to have done and on what basis she or we claim that she did the right thing. Nor when we ask whether what Jill did was in the moral realm or not, that is, when we raise the question of relevance, have we gotten to the question of justification.

From a common-sense perspective, it seems that questions about justification are not very difficult to answer. From what we've been told, Jill appears to have done the right thing. Indeed, we might even think that it is pretty clear that it is right because helping another person in need—when one is able to do so and when it doesn't cause that person more harm than good—is a moral principle that we should follow. So, if Jill acted on that basis, she had good reasons for what she believed and did. She was justified. But what connection, if any, do these *justifying reasons* have with biologically and psychologically based explanations of how Jill learned to act morally and was motivated to do so? In attempting to answer this question, I use some helpful parallels from naturalized epistemology.

Traditionally, justifications in both moral and nonmoral matters are thought of as having to do with providing adequate or good reasons for believing that something is so in the case of nonmoral beliefs, or for believing that something is right, obligatory, or good in the case of moral beliefs. These justifying reasons are distinguished from the ways in which beliefs are acquired. But, according to the scientific naturalistic account of justification that I am proposing, some causes can function as justifying reasons. In the naturalistic model of justification of moral beliefs, moral truth is considered to be the end or goal of moral belief formation. Some of the processes by which moral beliefs are formed and acquired are due to cognitive mechanisms and processes that reliably achieve that goal. These mechanisms and processes can serve as justifiers of moral beliefs. They provide good reasons for holding a belief. I extend this idea of reliable mechanisms to moral motivations and actions. The good or adequate reasons, the criteria in terms of which we make a judgment about attaining the well-functioning state, are conceived of as reliable mechanisms for achieving that end state. The ideal well-functioning of a moral agent requires reliable mechanisms for achieving accurate representations of the moral environment, proper motivation toward moral ends, and modes of behavioral enactment adequate for the successful execution of these ends on each of the levels of moral agency.

If these reliable mechanisms are operating, then we have the following picture of well-functioning. On the base level, evolutionarily based moral capacities and learned behavioral capacities function as reliable mechanisms for attaining the representations and motivations needed to achieve morally relevant actions. Turning to the behavioral level, reliable cognitive mechanisms lead to more or less accurate moral beliefs, and reliable motivational mechanisms bring about more or less proper bases for moral actions, and both together lead reliably to more or less successful execution of a moral action. The reflective level adds the processes that reliably lead to relatively sound moral rules or principles as cognitive and motivational bases for accurate execution. Finally, the self-referentially reflective level brings in processes that lead to conceiving of the self as a moral self. The combination of the conception of the self as a moral self and self-efficacy judgments is generally productive of both accurate reflective and behavioral-level moral beliefs and proper moral motivations, all leading to successful moral execution.<sup>27</sup>

For example, focusing on the base level, empathic distress is plausibly a reliable mechanism for achieving cognitive and motivational well-functioning at that level of moral agency. If we grant the plausibility of Hoffman's account of the nature, acquisition, and deployment of empathic capacities, empathic distress functions not only as a cause but also as a justifier or source of adequate moral cognition and motivation. Empathic distress reliably enables its possessor to identify a victim in need of help and moves its possessor to give the person assistance. Of course, we need not require that reliability mean unflinching success or even a high probability. It is sufficient that empathic distress is an effective cognitive and motivational mechanism in important cases, that is, that it is strategically successful.

If we think in terms of our epistemological model, empathic distress is closer to a traditional basic justifier than a nonbasic one, both cognitively and motivationally. Moreover, we can allow for both an externalist and an internalist perspective with respect to the assessing of adequacy. Using the above example, although Jill may be completely unable to answer questions about justification, we can determine from an externalist perspective that she possesses both a justified moral belief and a proper moral motivation and, on that basis, justification for her moral action. On the other hand, a reflective adult may be prepared to answer questions concerning justification by appeal to empathic distress on the basis of knowledge of its reliability deriving either from reflective common sense or the findings of Hoffman and others.<sup>28</sup> Thus, I contend that there is good reason for both the empathizer and some third party to claim that empathic distress is a reliable source for true moral beliefs and proper moral motivations and thus for successful moral actions.<sup>29</sup>

To the extent that the evolutionarily based capacities of the base level and the learned capacities of that level and those of the behavioral, reflec-

tive, and self-referential level are reliable means to the achievement of accurate moral representations, proper moral motivations, and effective moral actions, they serve in the naturalistic perspective as justifiers. Moral agents are justified in their moral beliefs, motivations, and actions insofar as these mechanisms are operative. Moreover, moral agents who are reflectively capable can appeal to the use of these mechanisms in responding to requests to justify their beliefs, motivations, and actions.

Thus, by determining those causal factors that function as reliable mechanisms for the production of accurate moral beliefs, proper moral motivations, and effective moral actions, the scientific naturalist provides answers not only to questions about the nature, acquisition, and employment of moral agency but also to those concerning justification. In answering these central questions about moral agency by using the findings and theories of biology and psychology, the scientific naturalist also shows how explanatory and metaethical connections can be made between ethics and the sciences of biology and psychology.

#### CONCLUSION

I have outlined a scientific naturalistic model of moral agency designed to answer four central questions of moral agency—those about its nature, acquisition, activation, and justification. I contend, though I have not argued for it here, that this model, based as it is on the findings and theories of the biological and psychological sciences, when supplemented with the theories and findings of the social sciences provides more satisfactory answers to these central questions than do its competitors, whether a priori philosophical, humanistic, or religious. Less controversially, it provides a means to correct and supplement other models of moral agency. And, perhaps less controversially still, it provides some understanding of one important feature of human persons—their moral agency, a feature recognized by all the major religious traditions and naturalists as central to being a human person. As such this scientifically based understanding of moral agency might be useful as a shared basis about what it means for human beings to be moral agents, how they become moral agents, and how they are able to put their moral capacities to work and do so in a reliable fashion. Such a common understanding and approach to moral agency may prove helpful in exploring further the religious aspects of moral agency, especially the relations of humans with the religious dimensions of the universe.

#### NOTES

1. I laid out and assessed some of these differing views in “The Scientific Naturalization of Ethics: The Death and Resurrection of Ethics” (Rottschaefer 1998b).

2. Thus far I have used such terms as *account*, *theory*, *hypothesis*, and *model* to characterize my scientific naturalistic proposals about moral agency. I use the term *model* in the phrase

"model of moral agency" in a way that is, to my knowledge, common among many scientists, especially biologists and psychologists, who are investigating theories or parts of theories or hypotheses. For instance, psychologists propose models of linguistic competency or short-term memory. A primary goal of such investigations is an adequate description and explanation of the phenomenon in question.

3. Clearly this traditional list must be supplemented to include the goods of the environment.

4. Although I formulate my model in the folk-psychological vocabulary of beliefs and desires, I do not consider such formulations necessary and would anticipate reformulations in scientific psychological or neuroscientific terms. See Rottschaefer 1998a, chap. 7.

5. I shall not attempt to determine what, if any, are the necessary and sufficient conditions for functional moral agency. In any case, I believe any such attempt must be empirically based. See note 24.

6. The most thorough and recent theoretical account of altruism, one that argues persuasively for the role of group selection, is that of Sober and Wilson (1998). They make both kin and reciprocal altruism forms of group altruism. Robert Brandon (1999) has recently argued that all "sensible" biologists will now admit that there is multilevel selection.

7. Over the course of his very productive career, Kohlberg refined and revised his procedures and views extensively. The classical formulation of his theory of three levels and six stages of moral development can be found in Kohlberg 1981. A good account of his revised theory appears in Kohlberg 1984.

8. Kohlberg has consistently contrasted his own interactionist views with those of both biological maturationism and environmentalism. However, in my view his conception of interaction remains vague about the degree to which moral structures are the result of the influences of nurture and nature. My interpretation makes nature the primary factor. Owen Flanagan (1982) offers a different point of view.

9. These theories appeared more plausible than either behaviorist or psychodynamic theories because the latter theories appeared to be incompatible with strongly held folk-psychological views of moral agency. Appeals to only natural and social environmental factors in the explanation of behavior or only to noncognitive intrapsychic processes appeared to critics to neglect an essential aspect of morality, its cognitive motivation. Such limited causal accounts of human agency have lost much of the empirical support or explanatory power they seemed to possess in their heydays (Hoffman 1988). Skinnerians and Freudians could appeal to eliminative materialist principles and argue that the reliance on a folk-psychological conception of moral agency is question begging. If either of these views prevailed, the folk-psychological image of our moral capacities and actions as cognitively based in moral principles would be radically false, and that image would be destined for the theoretical trash box. But both their scientific inadequacies and the strength of the folk-psychological conception of moral agency have made cognitive developmental theories seem much more plausible than these classical competitors.

10. Evolutionary and developmental biologists as well as philosophers of biology are now making persuasive arguments for the role of the environment in the ontogeny of genetically based traits and, in the case of organisms with learning capacities, for environmentally based learning, in the development of genetically based traits. For a recent, very clear account, see Brandon 1999.

11. One of the important differences concerns the cognitive-affective mix in the empathic response. Some theorists understand empathy as primarily a cognitive ability to recognize and understand the thoughts, perspective, and feelings of another individual. Others conceive of empathy as primarily an affective ability. This theoretical issue not only highlights an important problem about the nature of empathy but also raises a more fundamental question about the distinction between cognition and affect. Without attempting to decide that question at this point, I maintain that both cognitive and affective processes are representational processes but that such cognitive processes as reasoning and deciding have more representational power than do such affective processes as emotions. If that is so, then the understanding of empathy as primarily an affective process is better suited for discussions of development.

12. I take an emotional state to be a complex one with the following characteristics: (1) representation of something as desirable or undesirable; (2) feeling, interpreted as hedonic tone; (3) bodily sensations; (4) involuntary bodily responses and overt expressions; (5) tendencies to act; and sometimes (6) an upset or disturbed condition of body or mind (Alston 1967;

Campos, Barrett, Lamb, Goldsmith, and Stenbey 1983; Griffiths 1997). I omit the complication that a person in distress may not always have the appropriate psychological state for a stressful situation. He or she may, for example, have been rendered unconscious by a serious accident. Moreover, mere cognition of another's state as desirable or undesirable is not sufficient for empathic distress. As a result, what occurs in role taking or perspective taking may not be empathy, because these need not be emotional states. I consider having the same (or similar) emotional state as another because of the other's state to be the key element in empathy.

13. Hoffman does not mean levels in the strong sense used by Kohlberg or Piaget; rather, he means that a pattern in the development of empathy can be discerned and that these levels involve greater cognitive, affective, and behavioral complexity (Hoffman 1984a; 1984b; 1988).

14. I have used my own descriptive terms except in the case of global empathy.

15. By the age of ten to twelve months, infants understand that emotional expressions have emotional meaning, that they reveal the psychological state of the caregiver. The evidence for this is the striking phenomenon known as social referencing. Infants at this age actively try to obtain emotional cues from others, especially their mothers, in order to assist them in their own evaluation of uncertain situations. One important source of support for social referencing is the visual cliff studies.

16. Alternative interpretations have been suggested. Some researchers have suggested that the reactive distress of the infant may be due to a startle response or to self-concern. Others have claimed that the attentive concern of the infant can be explained as an orienting response or an effort to interpret and understand another's emotional behavior rather than empathy. Moreover, some infants ignore the distress of others, and there is a good amount of individual variation. Thus there is reason to question the claims about the presence of empathy in infants. Nevertheless, there are reasons to support the claim that infants are capable of empathic response, specifically the connection between their emotional response and prosocial initiatives and the connection between an infant's self-punishment and another's distress, whether or not the infant is the source of the other's distress (Thompson 1987). This tendency is increased by the mother showing the infant that he or she has caused distress and ought to do something to repair the damage. If the infant's response were primarily an orienting, startle, or initial avoidance response, it would not be accompanied, as it often is, by attempts to help the one in distress. Miriam Radke-Yarrow and her colleagues (1983) present a comprehensive discussion of the development of empathy and prosocial behavior in children, including infants and toddlers. Their studies and those of others indicate that infants display prosocial behaviors before three years of age.

17. The names used to identify the cluster of theories that I will be considering can be confusing. Roughly, social learning theories are often behaviorist or neobehaviorist theories that either deny the presence of cognitive factors or, while admitting their presence, deny their causal efficacy. Cognitive social learning theories, cognitive social theories, and cognitive behavioral theories all appeal to various sorts of cognitive factors to explain some human behaviors. Bandura calls his theory—reversing the qualifiers—a social cognitive theory in part to emphasize the primary role of cognitive factors in bringing about behaviors.

18. This discussion relies heavily on my book (Rottschaefer 1998a).

19. In their magisterial survey of learning theories, Bower and Hilgard (1981) find not only that "social learning theory [social cognitive learning theory] provides the best integrative summary of what modern learning theory has to contribute to the solution of practical problems" but also that it "provides a compatible framework within which to place information processing theories of language comprehension, memory, imagery and problem solving." In their view, information processing theories provide the best current scientific characterization of the intentional categories of cognition and action. Thus, they conclude that "social learning theory may provide a basis of consensus for much of the research in the next decade" (p. 472).

20. This view is not shared by all theorists. Compare cognitive developmentalists like Rest (1983; 1984) and Turiel (1983).

21. The very important role of self-censure or self-dissatisfaction is demonstrated very clearly in Bandura's rich account of the strategies of disengagement from moral responsibility, mentioned above in the text, that people use to avoid self-censure and self-dissatisfaction (1977b, 154–58; 1986, 375–89).

22. These judgments vary on several dimensions important for performance, such as grades of difficulty of the task, generality of expectations, and strength of expectations. The judgments

are complex, deriving by inference from cues about self-efficacy gained in actual performance, vicariously, persuasively, or emotively.

23. For critical discussion of Bandura's claims about self-efficacy see Eastman and Marzillier (1984a; 1984b) and Bandura (1984; 1997).

24. For the sake of simplicity and presentation, I have assimilated the conceptual components of Bandura's cognitivism to a folk-psychological account in terms of beliefs and desires. In point of fact, his theory is a mixed one using both folk-psychological and information processing conceptions of the mental.

25. This fuller sense seems to be what Hoffman (1988) means by the highest level of moral internalization. It is here that the conceptions of Bandura (1977b; 1978; 1986) converge with those of researchers, for example, Blasi (1984), Damon (1984), Hoffman (1988), and Wren (1991), who approach the question of the nature of moral agency from significantly different perspectives. I have tried to capture this point in the functional criterion for moral agency of moral integrity. However, I am not persuaded that this criterion represents a necessary condition of moral agency. Moral integrity may be closer to the maximal end of the continuum of moral agency than the kind of moral agency we normally practice. I believe that the question of the kind of moral agency we have, just like what kind of rational agents we are, is a matter to be determined empirically, and so I do not want to decide that issue in an a priori fashion. More empirical work needs to be done on the general applicability of these criteria to moral action. In addition, it is not clear to me that the criteria of moral motivation and moral integrity identify exclusively moral criteria. In both Bandura's and Alston's (1977) accounts, self-standards are not necessarily moral. Perhaps, in the end, substantive criteria must be invoked to distinguish moral agency from other types of self-referentially reflexive agency. However, I do think it is significant that five leading scholars of moral agency, working from diverse theoretical approaches, have postulated a criterion of moral integrity. Moreover, it is especially ironic that ardent critics of cognitive social learning theories, like Blasi and Wren, who champion moral integrity as a requirement for being a moral agent, can find empirical support for that requirement in the very sort of theory they find intrinsically irrelevant to moral agency.

26. It is an interesting question as to what extent personal unifications can be achieved within specific areas and between specific areas both across time and at a particular time in one's life. Such unifications, if possible, would also seem to be attainable in the moral domain.

27. I have described the functions of these two levels in relatively separated fashion. However, a more realistic account would have several or all of the levels functioning together. An account of successful functioning embracing several or all levels would require adequate descriptions and explanations of the interaction of the levels cognitively and motivationally.

28. But empathy, Hoffman tells us, has its limitations. Because it is based on a bystander model and there is a tendency to respond more empathically to those who are present than to those who are absent, and to those who are like us than to those who are not, it may well happen that empathic motivations fail in certain situations to be reliable motivators. They may fail to move us to help when we ought to or to help the right persons when we ought. Or they may move us to help when we should not. But these limitations on empathy should not surprise us. Given the evolutionary and learning history origins of empathy, we should not expect it to be a reliable mechanism for motivating moral actions in *all* circumstances. Our perceptual powers have their limitations, even though in normal circumstances and with respect to middle-sized objects they are generally reliable. That is, they work fairly well in conditions for which they have been designed to work, designed by nature and by the contingencies of our natural and social learning environments. With regard to nonmoral truths, we need to supplement our perceptual mechanisms in situations, for example, where the truth we seek concerns unobservable realities. Similarly, in situations where the stimuli that arouse empathy are not present or where the stimuli are so overpowering that they provoke motivations that do not lead to the best moral action for the situation, we can expect that our empathic motivations need to be complemented by reliable motivational mechanisms deriving from higher levels, for instance, from the reflective level. Empathic distress is an example of a kind of process working at the base (and perhaps behavioral) level of moral agency as a reliable means for achieving proper moral motivation and correct moral action.

29. See Rottschaefer 1998a for details about identifying moral ends and the mechanisms for achieving such ends as well as the specification of these mechanisms and their reliability.



## REFERENCES

- Alston, W. P. 1967. "Emotion and Feeling." In *Encyclopedia of Philosophy*, vol. 2, ed. P. Edmunds, 479–86. New York: Macmillan and Free Press.
- . 1977. "Self-Intervention and the Structure of Motivation." In *The Self: Psychological and Philosophical Issues*, ed. T. Mischel, 65–102. Oxford: Blackwell.
- Bandura, Albert. 1976. "Self-reinforcement: Theoretical and Methodological Considerations." *Behaviorism* 4:135–55.
- . 1977a. "Self-Efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84:191–215.
- . 1977b. *Social Learning Theory*. Englewood Cliffs, N.J.: Prentice Hall.
- . 1978. "The Self System in Reciprocal Determinism." *American Psychologist* 33: 344–58.
- . 1982. "Self-Efficacy Mechanism in Human Agency." *American Psychologist* 37: 122–47.
- . 1984. "Recycling Misconceptions of Perceived Self-Efficacy." *Cognitive Theory and Research* 8:231–55.
- . 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, N.J.: Prentice Hall.
- . 1997. *Self Efficacy: The Exercise of Control*. New York: W. H. Freeman.
- Barnett, M. 1987. "Empathy and Related Responses in Children." In *Empathy and Its Development*, ed. N. Eisenberg and J. Strayer, 146–62. Cambridge: Cambridge Univ. Press.
- Blasi, A. 1984. "Moral Identity: Its Role in Moral Functioning." In *Morality, Moral Behavior and Moral Development*, ed. W. Kurtines and J. Gewirtz, 128–37. New York: John Wiley and Sons.
- Boehm, C. Forthcoming. "Conflict and the Evolution of Social Control." *Journal of Consciousness Studies*.
- Bower, G. H., and E. R. Hilgard. 1981. *Theories of Learning*. 5th ed. Englewood Cliffs, N.J.: Prentice Hall.
- Bradie, Michael. 1994. *The Secret Chain: Evolution and Ethics*. Albany: State Univ. of New York Press.
- Brandon, R. 1999. "The Units of Selection Revisited: The Modules of Selection." *Biology and Philosophy* 14:167–80.
- Campos, J., K. Barrett, M. Lamb, H. Goldsmith, and C. Stenbey. 1983. "Socioemotional Development." In *Infancy and Developmental Psychology*, 4th ed., ed. M. M. Harth and J. Campos, 783–916. *Handbook of Child Psychology*, ed. Paul H. Mussen, vol. 2. New York: John Wiley and Sons.
- Damon, W. 1984. "Self-understanding and Moral Development from Childhood to Adolescence." In *Morality, Moral Behavior and Moral Development*, ed. W. Kurtines and J. Gewirtz, 109–27. New York: John Wiley and Sons.
- Darwin, Charles. 1871. *The Descent of Man and Selection in Relation to Sex*. London: J. Murray.
- Dobson, K. S., and L. Block. 1988. "Historical and Philosophical Bases of the Cognitive Behavioral Therapies." In *Handbook of Cognitive-Behavioral Therapies*, ed. K. S. Dobson, 3–38. New York: Guilford Press.
- Eastman, C., and J. S. Marzillier. 1984a. "Continuing Problems with Self-efficacy Theory: A Reply to Bandura." *Cognitive Therapy and Research* 8:257–62.
- . 1984b. "Theoretical and Empirical Difficulties in Bandura's Self-efficacy Theory." *Cognitive Therapy and Research* 8:213–29.
- Erwin, E. 1978. *Behavior Therapy: Scientific, Philosophical and Moral Foundations*. New York: Cambridge Univ. Press.
- Flack, J., and de Waal, F. Forthcoming. "Any Animal Whatever: Darwinian Building Blocks of Morality in Monkeys and Apes." *Journal of Consciousness Studies*.
- Flanagan, Owen J. 1982. "Moral Structures." *Philosophy of the Social Sciences* 12:255–70.
- . 1984. *The Science of Mind*. Cambridge: MIT Press.
- Griffiths, P. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: Univ. of Chicago Press.

- Hoffman, Martin J. 1984a. "Interaction of Affect and Cognition in Empathy." In *Emotion, Cognition, and Behavior*, ed. C. Izard, J. Kagan, and R. Zajonc, 103–31. New York: Cambridge Univ. Press.
- . 1984b. "Empathy, Its Limitations, and Its Role in a Comprehensive Moral Theory." In *Morality, Moral Behavior, and Moral Development*, ed. W. Kurtines and J. Gewirtz, 283–302. New York: John Wiley and Sons.
- . 1988. "Moral Development." In *Developmental Psychology: An Advanced Textbook*, ed. M. Bornstein and M. Lamb, 497–548. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Kazdin, A. E. 1978. *History of Behavior Modification: Experimental Foundations of Contemporary Research*. Baltimore: University Park Press.
- Kohlberg, Lawrence. 1981. *The Philosophy of Moral Development. Essays in Moral Development*, vol. 1. San Francisco: Harper and Row.
- . 1984. *The Psychology of Moral Development: The Nature and Validity of Moral Stages. Essays in Moral Development*, vol. 2. San Francisco: Harper and Row.
- Mischel, W., and H. N. Mischel. 1976. "A Cognitive Social-Learning Approach to Morality and Self Regulation." In *Moral Development and Behavior: Theory, Research and Social Issues*, ed. T. Lickona, 84–107. New York: Holt, Reinhart and Winston.
- Radke-Yarrow, Miriam, C. Zahn-Waxler, and M. Chapman. 1983. "Children's Prosocial Dispositions and Behavior." In *Socialization, Personality and Social Development*, ed. M. E. Hetherington, 469–546. Vol. 4 in *Handbook of Child Psychology*, ed. Paul H. Mussen. 4th ed. New York: John Wiley and Sons.
- Rest, J. R. 1983. "Morality." In *Cognitive Development*, ed. J. Flavell and E. Markmen, 556–629. Vol. 3 in *Handbook of Child Development*, ed. Paul H. Mussen. 4th ed. New York: John Wiley and Sons.
- . 1984. "The Major Component of Morality." In *Morality, Moral Behavior, and Moral Development*, ed. W. Kurtines and J. Gewirtz, 24–40. New York: John Wiley and Sons.
- Rolston, Holmes, III. 1999. *Genes, Genesis and God: Values and Their Origins in Natural and Human History*. Cambridge: Cambridge Univ. Press.
- Rottschaefer, William A. 1991. "Social Learning Theories of Moral Agency." *Behavior and Philosophy* 19:61–76.
- . 1998a. *The Biology and Psychology of Moral Agency*. New York: Cambridge Univ. Press.
- . 1998b. "The Scientific Naturalization of Ethics: The Death and Resurrection of Ethics." Paper presented at the Northwest Conference on Philosophy, University of Oregon.
- Sober, E., and D. S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard Univ. Press.
- . Forthcoming. "Target Article—Unto Others: The Evolution and Psychology of Unselfish Behavior." *Journal of Consciousness Studies*.
- Thompson, R. A. 1987. "Empathy and Emotional Understanding: The Early Development of Empathy." In *Empathy and Its Development*, ed. N. Eisenberg and J. Strayer, 119–45. New York: Cambridge Univ. Press.
- Turiel, E. 1983. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge Univ. Press.
- Waal, F. de. 1996. *Good-Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard Univ. Press.
- Wren, T. 1991. *Caring about Morality: Philosophical Perspectives in Moral Psychology*. Cambridge: MIT Press.