A pre-publication draft: In R. Burgess & K. MacDonald (Eds.) (2005). *Evolutionary perspectives on human development*. CA: Sage Publications

An Evolutionary Reconceptualization of Kohlberg's Model of Moral Development

Dennis Krebs

Simon Fraser University

An Evolutionary Reconceptualization of Kohlberg's Model of Moral Development

At the end of a program of research my colleagues and I conducted on Kohlberg's (1984) model of moral development, I concluded that it is poorly equipped to account for the kinds of moral judgments and moral behaviors people emit in their everyday lives. My search for a more ecologically valid model of morality led me to what many people would consider an unlikely source, the modern synthetic theory of evolution. In this chapter I will briefly describe Kohlberg's model of moral development, identify its deficiencies, reconceptualize Kohlbergian structures of moral judgment as evolved decision-making mechanisms, and explain why I believe this reconceptualization gives rise to a more valid model of moral development.

Kohlberg's Model of Moral Development

The two central assumptions of Kohlberg's (1984) model of moral development are that the primary source of morality is moral reasoning and that people become more moral because they acquire increasingly sophisticated structures of moral reasoning. To assess moral reasoning, Kohlberg and his colleagues developed a test containing nine hypothetical moral dilemmas followed by sets of questions (Colby & Kohlberg, 1987). In the most often cited dilemma, a character named Heinz must decide whether or not to steal an over-priced drug to save his dying wife. Kohlbergians assume that the deontic decisions people make—whether Heinz should or should not steal the drug—stem from structures of moral reasoning. To map these structures, Kohlbergians ask moral decision-makers "why" questions designed to induce them to explicate the principles from which they deduced their moral choices.

The main source of support for Kohlberg's model stems from a longitudinal study conducted by Kohlberg and his colleagues. In 1958 Kohlberg gave a cohort of boys an early version of his test, and with the help of colleagues, retested them every three or four years for

more than two decades (Colby & Kohlberg, 1987). This study produced three main conclusions.

First, moral judgments made by individuals are structurally consistent across varying content.

That is to say, individuals invoke the same forms of moral reasoning to resolve different kinds of

moral dilemma; moral judgment is organized in "structures of the whole." Second, as children

grow older, their structures of moral reasoning normally undergo qualitative changes. According

to Colby and Kohlberg (1987), new structures "transform and displace" older structures, giving

rise to the five stages of moral development summarized in Table 1. Kohlberg and his colleagues

found that the participants in his longitudinal study went through these stages in an invariant

sequence. Third, because the structures of moral reasoning that define higher stages of moral

development are more cognitively sophisticated than those that define lower stages—that is to

say, more differentiated, integrated, and logical—and based in more sophisticated perspective-

taking abilities, they give rise to more moral decisions (that is to say, moral decisions that are

more prescriptive, universal, impartial, and objective).

Insert Table 1 about here

In 1984, Kohlberg and Candee published a model linking moral reasoning to moral

conduct. Kohlberg and Candee (1984) argued that moral judgment is necessary but not sufficient

for moral behavior. To behave morally, people must decide what the most moral course of

action is, but this does not necessarily compel them to behave morally. To implement moral

decisions, people must experience an obligation to carry them out, which entails making follow-

up judgments of responsibility. In addition, people must possess sufficient ego-strength to do

what they think they should.

An Evaluation of Kohlberg's Model of Moral Development

Several years ago, my colleagues and I noted that most of the empirical support for Kohlberg's

model stemmed from studies conducted in academic contexts in which trained interviewers or

experimenters asked university students to make and to justify decisions about how the

characters in Kohlberg's dilemmas should resolve hypothetical moral conflicts. We set out to

vary parameters of this prototypic situation and determine whether they affected the forms of

moral judgment people invoked. To this end, we asked people who differed in a variety of ways

to make moral judgments in a variety of non-academic contexts, to audiences other than

experimenters, about moral dilemmas other than those on Kohlberg test, and about people other

than the characters in Kohlbergian dilemmas.

The findings from this program of research are reviewed in several publications (see

Krebs, Denton & Wark, 1997; Krebs, Vermeulen, Carpendale & Denton, 1991; Krebs, Wark &

Krebs, 1995, and Wark & Krebs, 1997). To summarize, we found, as Kohlberg would expect,

that people tend to make structurally similar—that is to say same-stage—moral judgments to

hypothetical moral dilemmas similar to those on Kohlberg's test. However, we also found that

people tend to make lower-stage moral judgments about more real-life-like dilemmas. For

example, compared to moral judgments on Kohlberg's test, participants in our studies made

relatively low-stage moral judgments to moral dilemmas about prosocial behavior and drinking

and driving (Krebs, Denton, Vermeulen, Carpendale, & Bush, 1991), business deals and free

trade (Carpendale & Krebs, 1995), prostitution (Bartek, Krebs & Taylor, 1993), and a variety of

real-life dilemmas (Krebs, Denton, Wark, Couch, Racine, & Krebs, 2002; Wark & Krebs, 1997).

Young adults made lower-stage moral judgments about Kohlbergian dilemmas and a dilemma

about drinking and driving when they were drinking at bars and parties than they did in academic

contexts (Denton & Krebs, 1990). University students made lower-stage moral judgments to a Professor of Business Administration than to a Professor of Philosophy (Carpendale & Krebs, 1992). Parents and children made relatively low-stage moral judgments about moral dilemmas they experienced in their families (Krebs, Vermeulen & Denton, 1991). Incarcerated female juvenile delinquents, especially those who scored high on a test of defensiveness, made relatively low-stage moral judgments about moral dilemmas involving prostitution (Bartek, Krebs & Taylor, 1993). People tended to make different kinds of moral judgments about themselves than they did about others (Krebs et al., 2002; Krebs & Laird, 1998).

Considered as a whole, the findings from our research program did not support Kohlberg's model of moral development. Our data were not consistent with the assumption that moral judgments stem from structures of the whole that have transformed and displaced their predecessors. Instead, our data suggested that people retain old structures of moral judgment after they develop new ones. Moral development is characterized more by an expansion of the range of possibilities than by changes in the overriding structure of the mind. The types of moral judgment people make are not determined solely by the structures of moral reasoning they have acquired; moral judgments are the product of complex interactions between internal qualities of people and the external situational variables they encounter.

Findings from our studies also suggested that the relation between moral judgment and moral behavior is more complex than Kohlberg and Candee (1984) assumed. As postulated by Kohlberg and Candee, in some contexts people appear to decide what is right, then act on their decisions. However, in other contexts people appear to act without engaging in moral reasoning, then invoke moral judgments to justify their decisions (Denton & Krebs, 1990; Krebs et al., 2002; Krebs & Laird, 1998).

In the end I concluded that Kohlberg's model pertains to only one of many functions of moral judgment, namely to make the most rational moral decision of which one is capable. The methods employed by Kohlbergians are designed to induce people to perform this function. Making moral judgments about Kohlbergian dilemmas in academic contexts is akin to solving problems in logic. As acknowledged by Colby and Kohlberg (1987), Kohlberg's model is a model of "moral competence."

To adequately account for moral development, we must attend to other functions of moral judgment and moral behavior. In real-life people use moral judgments for more social and pragmatic purposes than they do when they make moral judgments on Kohlberg's test. When people tell other people they "should" behave in particular ways, they are rarely trying to explicate their ideal conceptions of morality; rather, they are usually trying to induce the recipients of such judgments to behave in accordance with the moral prescriptions they are pronouncing. They are trying to exert social influence. Recipients of such moral judgments often respond by arguing with those who pronounce them, attempting to persuade them that the prescriptions are unjustified. Negotiations ensue in which parties adduce arguments to support their points of view (Krebs et al., 2002). In real life, people also use moral judgments to approve and disapprove of the behavior of others, to praise, condemn, and blame. In addition, people use moral judgments to guide, justify, and excuse their own behavior to approve and disapprove of themselves, to enhance their social image and to foster self-esteem (Denton & Krebs, 1990; Krebs et al., 2002; Krebs & Laird, 1998).

A Functional Reconceptualization of Kohlberg's Elements of Morality

As I have explained, when people make decisions about what the characters in Kohlbergian dilemmas should do, interviewers ask them why, and after interviewees proffer a

reason, interviewers ask them why again until interviewees reach the limit of their explanations.

For example, an interviewee might say, "Heinz should steal the drug to save his wife's life", and

the interviewer might ask, "why should Heinz save his wife's life?" In response to this probe,

the interviewee might answer, "because he owes it to her." The interviewer might then ask, "why

is it right for Heinz to fulfill his obligations to his wife?" and the person might say, "because she

has done a lot for him" or "because he made a marriage contract," and so on. Colby and

Kohlberg (1987) have classified the reasons people adduce in support of their decisions and

labeled them the "elements" of morality (see Table 2).

From Kohlberg's perspective, "why" questions induce people to explicate the principles

from which they derived their moral decisions. From a more functional perspective, such

questions induce people to identify the goals they are advising moral decision-makers to pursue,

goals such as doing their duty, paying their debts, reciprocating, obtaining approval, avoiding

punishment, enhancing their reputations, supporting their groups and so on (see Table 2). From a

Kohlbergian perspective, the purpose of repeatedly asking "why" questions is to induce people to

invoke increasingly abstract principles of morality. From a functional perspective, the purpose of

repeatedly asking "why" questions is to get at increasingly ultimate goals and purposes. If you

keep asking people functional "why" questions, you will ultimately end up in one of two places,

with God or in the theory of evolution. I ended up in the latter.

Insert Table 2 about here

An Earlier Examination of Relations Between Kohlbergian and Evolutionary

Approaches to Moral Development

I was not the first psychologist to adopt a biological approach to morality and view

Kohlberg's model from an evolutionary perspective. Kevin MacDonald (1988) was. MacDonald (1988) reviewed evidence on moral judgment and moral behavior and concluded that much of it was more consistent with a sociobiological viewpoint than with Kohlberg's cognitive-developmental model. In particular, MacDonald (1988) adduced evidence supporting the following eight conclusions. First, children's moral behavior is more self-interested than it appears from the conclusions reached by researchers. Second, people invoke different forms of moral reasoning in different contexts. Third, people perform cost-benefit analyses when they make real-life moral decisions. Fourth, "reasoning about oneself, one's relatives, and significant others is done with a different calculus than is reasoning about hypothetical situations" (p. 148). Fifth, people sometimes fail to behave in accordance with their moral judgments, especially when they conflict with their self-interest. The link between moral reasoning and moral behavior is weak except when the decisions in question involve little cost. Sixth, people use moral judgments to create false impressions and to justify their behavior: "Individuals who are adept at moral reasoning (i.e., perform at the higher stages) are better able to provide reasons which rationalize their self-interested actions in a manner that would justify their behavior to other individuals" (p. 143). Seven, only the first three stages in Kohlberg's sequence are cross-culturally universal. And finally, emotional reactions play an important role in real-life moral decision-making.

In the remainder of this paper, I build upon these ideas. I review the basic propositions of evolutionary theory, reconceptualize the structures of moral judgment described by Kohlbergians as evolved mental mechanisms, offer an account of the evolution of moral judgment and moral behavior, and explain why I believe that this account gives rise to a more ecologically valid model of moral development than the one advanced by Kohlberg and his colleagues.

Toward an Evolutionary Model of Morality

Evolutionary theorists assume that the ultimate goal of most (but not all) behavior is to enhance actors' inclusive fitness, defined in terms of the number of replicas of their genes they contribute to future generations. More precisely, evolutionary theorists assume that individuals inherit genes that were selected in ancestral environments because they guided the development of mechanisms that enabled their ancestors to survive, to reproduce and to pass the genes and mechanisms on to future generations. The function of evolved mechanisms is to help individuals solve adaptive problems pertaining to survival, reproduction, care of offspring, and so on.

There are two basic ways in which genes, or mechanisms, can be selected and increase in frequency, or evolve, over generations. In the first, the vehicles housing the genes or mechanisms—that is to say, individuals—survive until sexual maturity, pass replicas of their genes or mechanisms on to others through asexual or sexual reproduction, then take whatever measures are necessary to insure that the offspring in whom they have deposited replicas of their genes survive and reproduce. In the second, individuals behave in ways that foster the survival and reproduction of individuals other than their offspring who possess replicas of their genes— that is to say, individuals who are similar to them genetically. This process is usually referred to as kin selection. The main difference between the two forms of natural selection is that in the first individuals create the vehicles that transport replicas of 50 % of their genes to future generations, whereas in the second individuals assist individuals created by others, who may possess varying proportions of their genes.

Evolved Psychological Mechanisms

Evolutionary theorists may attend to different products of evolution. For example, some

theorists attend to the frequency of genes in a population, whereas others attend to the number of offspring produced by individuals. Evolutionary psychologists attend to the mental mechanisms that individuals inherit and the behavioral strategies such mechanisms contain. Buss (1999, pp. 48-49) has identified six properties of evolved psychological mechanisms. First, they are shaped by the ways in which they recurrently solved specific adaptive problems over the evolutionary history of the species in which they evolved. Second, they are activated by relatively "narrow slices of information." Third, the information that triggers such mechanisms—the input— pertains to the particular adaptive problems the mechanisms were designed to solve. Fourth, the information processed by such mechanisms is transformed by a set of "if-then" decision rules into output. Fifth, "the output of an evolved psychological mechanism can be physiological activity, information to other psychological mechanisms, or manifest behavior." Finally, "the output…is directed toward the solution to a specific adaptive problem." Although it is safe to assume that the strategies implicit in such mechanisms were on average winning strategies in ancestral environments, they need not necessarily produce optimal solutions to the problems that activate them in modern environments.

Kohlberg's Structures of Moral Judgment Reconceptualized as Evolved Psychological Mechanisms.

In contrast to learning theorists who posit general purpose mechanisms, both Kohlbergians and evolutionary psychologists believe the mind is composed of specialized structures or mechanisms designed for particular purposes. Kohlbergians and evolutionary theorists assume that the mental structures that give rise to moral judgments and moral behaviors are activated by particular kinds of problem and that their function is to process information in ways that give rise to particular kinds of decision. Both types of theorist assume that the form or

pattern of the output produced by mental structures reflects the ways in which they are designed, and that the primary goal of social scientists is to map the design of the mechanisms by deciphering the operating principles or decision-rules implicit in their output. Kohlbergian and evolutionary psychologists also are attentive to the origin and development of mental structures and mechanisms.

Such similarities notwithstanding, Kohlbergians make different assumptions from evolutionary theorists about how mental structures originate and the functions for which they were designed. Adopting an ontogenetic perspective, Kohlbergians assume that structures of moral judgment develop through the cognitive processes of assimilation and accommodation and that their function is to enable people to deduce the most moral solution to moral problems. Adopting a phylogenetic perspective, evolutionary theorists assume that evolved mechanisms originate from genetic variations selected in ancestral environments, and that their function is to help individuals solve real-life adaptive problems. Kohlbergians focus mainly on the cognitive functions of making moral judgments; evolutionary theorists focus mainly on the adaptive functions of moral behaviors. The central thesis of this paper is that the conception of mental mechanisms derived from evolutionary theory offers a better basis than Kohlberg's conception of structures of moral judgment for explaining the ways in which people make moral decisions and for accounting for moral development. Stated in stronger terms, I believe that most of the moral judgments people make in their everyday lives and most of the moral behaviors they emit are products of the kinds of evolved psychological mechanisms described by evolutionary psychologists, as opposed to the structures of moral reasoning described by Kohlbergians.

The Evolution of Moral Mechanisms

If we assume that people derive moral judgments from evolved mental mechanisms, the

first question that arises is, what adaptive functions did such mechanisms serve in ancestral

environments; what kinds of adaptive problems did such mechanisms help our hominid ancestors

solve? I believe the answer to this question is that the mechanisms that give rise to morality

evolved to solve the social problems that occurred when our ancestors banded together to foster

their biological interests. Among the many species in the animal kingdom, Homo Sapiens are

among the most social. From the perspective of evolution, mechanisms that dispose animals to

aggregate and interact with other members of their species—and for that matter with members of

other species—evolve when the mechanisms help animals foster their biological interests. Such

mechanisms may help animals enhance their inclusive fitness in several ways. As examples,

individuals who band together may be less susceptible than more solitary individuals to

predators, and groups may be able to hunt larger game than individuals could kill on their own.

Many of the benefits of sociality stem from cooperative exchanges. Individuals working

together can often obtain more for themselves than they could by working alone. Individuals can

often trade items of relatively little value for items of greater value. Individuals can render

valuable assistance to others at little cost to themselves in return for low-cost assistance when

they are in need. In such exchanges both (or all) parties can come out ahead through what

economists call gains in trade.

According to most experts on human evolution, cooperation was instrumental in the

evolution of the human species. To quote Leakey and Lewin (1977), "Throughout our recent

evolutionary history, particularly since the rise of a hunting way of life, there must have been

extreme selective pressure in favor of our ability to cooperate as a group… The degree of

selective pressure toward cooperation… was so strong, and the period over which it operated so

extended, that it can hardly have failed to have become embedded to some degree in our genetic

makeup" (p. 45).

In view of the benefits of cooperation, it might seem that mechanisms disposing individuals to cooperate should evolve without impediment, but unfortunately this is not the case. In his book, *Theory of Justice*, the philosopher John Rawls (1971) explains why:

> Although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits of their collaboration are distributed, for in order to pursue their ends, each prefers a larger to a lesser share. (p. 4)

Selfish preferences pose a serious problem for the evolution of cooperative mechanisms, because it is in everyone's biological interest to give less than his or her share and to induce others to give more than their share. Inasmuch as the resources in question foster the inclusive fitness of those who possess them, individuals who inherit mechanisms that dispose them to behave selfishly could fare better than those who inherit mechanisms that dispose them to behave cooperatively, mediating the evolution of selfishness and the extinction of cooperation. However, fortunately for the evolution of morality, it is not in individuals' interest to let others treat them selfishly, so we would expect mechanisms designed to counteract selfishness in others to evolve.

Most people consider selfish behaviors, defined as fostering one's own interests at the expense of others, immoral; and most people consider cooperative behaviors, defined as fostering one's interests in ways that foster the interests of others, moral. I believe that the biological function of morality is to uphold fitness-enhancing systems of cooperation. Viewing morality in these terms helps us understand its nature. Morality boils down to giving one's share (doing one's duties) and taking one's share (exercising one's rights), cooperating with others by treating

them fairly, as one would like to be treated, and resisting the temptation to maximize one's gains

at the expense of others. This conception of morality has been espoused in non-biological terms

by many scholars. As one example, Rest (1983) asserts that morality consists in "standards or

guidelines that govern human cooperation—in particular how rights, duties, and benefits are [to

be] allocated… Moralities are proposals for a system of mutual coordination of activities and

cooperation among people" (p. 558).

The Social Functions of Morality

In contrast to the purpose of making moral judgments about the hypothetical characters in

Kohlbergian dilemmas, people make moral judgments in their everyday lives for two primary

purposes: (a) to influence the behavior of others and (b) to guide their own behavior. Let us

consider each in turn.

Moral Judgments as Forms of Social Influence

Viewed biologically, moral judgments are a form of communication. Biological analyses

of communication assume that animals are evolved to send signals that induce recipients to

behave in ways that foster the senders' interests, or to manipulate them. Such signals are often

deceptive (Dawkins, 1989; Mitchell & Thompson, 1986). Humans' relatively large brains and

their capacity for language enable them to employ a significantly larger range of manipulative

communication strategies than those available to other species. Senders are able to take the

perspective of recipients (referred to as "mind-reading" by some psychologists), and plan long

into the future. Recipients' reactions to senders' signals are less a function of the physical

properties of the signals themselves and more a function of how recipients represent them

mentally.

From a biological perspective, when people send second-person moral judgments to

others such as, "you should help me," "you should keep your promises," and "you are not being fair to me," they are tying to induce recipients to foster their interests by persuading them to behave in accordance with the prescriptions. When people express more abstract judgments such as "honesty is the best policy" and "people should obey the law," they are attempting to induce recipients to uphold the systems of cooperation from which they benefit. When senders support such judgments with reasons such as, "because you will cultivate a good reputation" and "everyone will benefit in the end," they are attempting to induce recipients to form cognitive representations of the "if" conditions that activate the behaviors they are prescribing. Such judgments are tools of social influence, tactics of persuasion designed to induce recipients to behave in ways that directly or indirectly benefit those making the judgments.

The natural selection of moral judgments. Of all the moral judgments people could make, why do they make those observed and classified by Kohlberg and his colleagues? Why do only a few of the many possible moral judgments become normative in all cultures? Kohlbergian and evolutionary psychologists would agree that people select the moral judgments that are best equipped to solve the moral problems they face, that such problems are social in nature, and that they involve conflicts of interest. However, Kohlbergians would focus on ideal solutions to hypothetical versions of such problems, whereas evolutionary theorists would focus on actual solutions to real problems in which the individuals involved are attempting to foster their biological interests.

Although it might seem that individuals motivated to foster their biological interests would make moral judgments such as, "you should sacrifice your interests for my sake," such judgments would not be effective, because recipients would not be receptive to them. In effect recipients of moral judgments are agents of selection, determining what kinds of judgment work.

Although those who make moral judgments may well attempt to use them to manipulate and exploit recipients, successful manipulations would have to be subtle, leading recipients to believe that abiding by them would foster their interests. In general, the most effective moral judgments should prescribe behaviors that foster the interests of senders and recipients. Thus, I would expect moral judgments to assume the form, "you should foster my interests in ways that foster your interests," which is exactly what most of the moral judgments classified by Kohlbergians do. To summarize, senders do not exhort recipients to behave selfishly because it is not in senders' interest, and senders do not exhort recipients to behave altruistically because recipients are unreceptive to such exhortations. Senders exhort recipients to adopt the kinds of cooperative strategies prescribed by Kohlbergian moral judgments because the mental mechanisms possessed by recipients are designed in ways that make them receptive to such judgments.

     The activation of stage-based moral judgments. As discussed, we found that in certain contexts adults invoke moral judgments made by people at all of the stages described by Kohlberg. From an evolutionary perspective, we would expect people to make the kinds of moral judgment that contain the greatest potential to foster their biological interests, which will depend on the "if" conditions implicit in the problems they encounter. So, for example, relatively powerful members of groups should make Stage 1 judgments that exhort weaker members to obey authority in order to avoid punishment. Those who are relatively equal in power should be more inclined to make Stage 2 judgments to each other that uphold mutually beneficial deals. Friends and relatives should make Stage 3 moral judgments that uphold their long-term relationships; and those who have a vested interest in maintaining the social order should make Stage 4 judgments. In general, the benefits implicitly promised to those who conform to high-stage moral judgments are more general and delayed than the benefits implicitly

promised to those who conform to low-stage moral judgments. Therefore, in general, the "if"

conditions invoked to activate higher-stage moral behavior are more tenuous than the "if"

conditions invoked to activate lower-stage moral behavior.

Moral Judgments as Forms of Self-guidance

In Kohlberg's model, which assumes that all the moral judgments people make stem from

the same structure of moral judgment, there is no reason to expect any differences between the

moral judgments people make to and about themselves and the moral judgments they make to

and about others. In contrast, from an evolutionary perspective, we would expect the form of the

two types of judgment to differ in significant ways. We would expect the judgments people

make about their rights and duties to be more selfish than the judgments they make about the

rights and duties of others for two reasons. First, we would expect people to be motivated to

advance their own interests. Second, we would expect the authors of self-serving moral

judgments to constitute more receptive audiences than most other recipients. Thus, we would

expect people to send messages to themselves such as, "you deserve more than your share,"

"your contributions are worth more than those of others," and "you should look after your own

interests," then graciously receive them with favor.

The magnitude of self-serving biases in moral judgment should, however, vary across a

variety of conditions. First, when people express moral judgments about themselves to others,

recipients should exert a constraining effect on the selfishness of the judgments in much the same

way they exert a constraining effect on the selfishness of second-person moral judgments. We

would expect audiences to call people on their selfishness and point out inconsistencies in the

standards they are applying to themselves and to others. For this reason, in general, we would

expect people to make less selfish moral judgments in public than in private. People also should

be selective in the audiences to whom they express their moral judgments, choosing those they anticipate will be most receptive to their points of view (Krebs & Denton, 1997).

Second, the extent to which people apply to themselves the standards they preach to others may depend on the persuasiveness of their communications. When individuals adduce arguments in support of the moral judgments they make to others, they may in effect persuade themselves that the arguments are right and buy into them (see Brown, 1986; Taylor, 1989). As suggested by Trivers (2000), believing in the standards one preaches to others may enhance the persuasive power of one's communications. Alternatively, people may compartmentalize moral judgments about themselves and others.

Third, the mechanisms through which people make moral judgments may be designed in ways that contain cognitive representations of others. In effect, internalized images of others may send and receive moral messages. Freud has suggested that introjected images of parents, or more exactly their superegos, tell people what they ought to do and sit in judgment of them when they fail to live up to their standards. The idea that the psychological mechanisms through which people make moral judgments to themselves, often called their conscience, contain internalized representations of others, especially their parents, has been advanced by scholars from a wide variety of theoretical traditions (e.g., Aronfreed, 1968; Higgins & Eccles-Parson, 1983). However, in the same way that people may direct their moral judgments to audiences who are receptive to their biases, they may invoke internalized audiences who exhort them to advance their own interests and help them justify their self-serving behaviors.

In the end, we must rely on empirical research to map the design of structures that give rise to moral judgments about selves and others. Although a host of studies have reported self-serving biases in social and moral judgments (see Krebs, Denton & Higgins, 1988; Krebs &

Denton, 1997; MacDonald, 1988; Taylor, 1989, for reviews), people differ in the extent to which

they harbor such biases, and the biases tend to diminish, disappear, and even reverse themselves

when people make judgments about their friends and relatives (Krebs & Laird, 1998).

The Evolution of Moral Behavioral Dispositions

Implicit in the analysis of the evolution of moral judgment I have been advancing is a

profoundly important implication, namely that humans are biologically endowed with the

capacity to behave in moral ways. If, as I have argued, people use moral judgments such as those

classified by Kohlberg and his colleagues to activate evolved mechanisms that induce recipients

to behave in accordance with their prescriptions, recipients must possess such evolved

mechanisms. Indeed, I would expect the behavior-producing mechanisms to have evolved before

the judgment-producing mechanisms, else the latter would have nothing to activate. But could

mechanisms that dispose individuals to behave morally have evolved?

Some very eminent evolutionary theorists think not. Consider for example the following

conclusions:

> Nothing resembling the Golden Rule or other widely preached ethical principles seems to
> be operating in living nature. It could scarcely be otherwise, when evolution is guided by
> a force that maximizes genetic selfishness (Williams, 1989, p. 195)

> Be warned that if you wish, as I do, to build a society in which individuals cooperate
> generously and unselfishly toward a common good, you can expect little help from
> biological nature. Let us try to teach generosity and altruism, because we are born selfish.
> (Dawkins, 1976, p. 3)

As much as I admire the work of these theorists, I believe their conclusions about the

evolution of morality are incorrect, or at least misleading. When evolutionary biologists talk

about selfish behaviors, they mean behaviors that help individuals propagate their genes. Such

theorists are, in effect, saying that the mechanisms that give rise to all evolved behaviors (that is

to say, behaviors that helped ancestors propagate their genes) are designed in ways that help individuals propagate their genes in modern environments. But this is not necessarily true, because, among other reasons, behaviors that were adaptive, or genetically selfish, in ancestral environments may not be adaptive in current environments. Further, genetically selfish behaviors are not necessarily individually selfish, which is the criterion most people use for morality. As examples, it may be genetically selfish to sacrifice oneself for the sake of one's offspring or other kin, but such behaviors would not be considered individually selfish or immoral. Although cooperative behaviors may foster the genetic interests of cooperators (rendering cooperation genetically selfish), such behaviors may nevertheless qualify as moral.

In my view, there is no necessary inconsistency between behaving morally and fostering one's biological interests. It is not immoral to attempt to survive or to reproduce; indeed some think people have a moral obligation to preserve their lives and to bear children. The criterion for morality pertains to how we go about fostering our biological interests. In general, we consider it moral to foster our biological interests in ways that foster the interests of others—that is to say in cooperative ways—and we consider it immoral to foster our interests at the expensive of others—that is to say in selfish ways. I believe there is little question that mechanisms giving rise to the cooperative strategies prescribed by Kohlbergian Stage 1, 2 and 3 moral judgments have evolved in the human species, and in some other species. I also believe that such mechanisms can, under certain conditions, give rise to Stage 4 and Stage 5 moral strategies. I have reviewed evidence supporting this claim in several publications (Krebs, 1998; Krebs, 2000a, b, c; Krebs & Janicki, 2003). I will briefly summarize it here.

The Evolution of Stage 1 Strategies

Stage 1 moral judgments prescribe obeying those who are "older, bigger, or more

powerful" than you and the rules they enforce in order to avoid getting "beaten up," "hit," "punished," "put in jail," or "killed." It is easy to see how the strategy implicit in such judgments could pay off better than more selfish attempts to maximize one's gains. Members of groups who are relatively weak face a Hobson's choice: either defer to those who are more powerful, or suffer the consequences. Faced with such choices, it is more adaptive for relatively subordinate members of groups to submit to the authority of the more dominant members in order to make the best of a bad situation and live to fight another day.

There is a spate of evidence that mental mechanisms giving rise to submissive and subordinate behaviors have evolved in many species (see Alcock, 1998; Boehm, 2000; Krebs, 1998, 2000b; Sloman & Gilbert, 2000), including humans (Milgram, 1964; Strayer & Strayer,1976). In some species, such mechanisms produce dominance hierarchies or pecking orders. Deferential mechanisms are rooted in fear systems, combined perhaps with a touch of awe.

The Evolution of Stage 2 Strategies

Stage 2 moral judgments prescribe helping other people, "because you may need them to do something for you one day," keeping promises so the "other person will keep promises to you or give you something in return, " and so on. In short, you scratch my back and I'll scratch your back. As pointed out by Trivers (1971) and others, individuals who reciprocate resources may gain more than individuals who do not through gains in trade. In some contexts individuals are able to bestow huge benefits on others at very little cost to themselves. Trivers (1971) gives the example of someone throwing a drowning person a life saver.

To evolve, Stage 2 cooperative strategies must contain antidotes to exploitation. One of the strategies prescribed by Stage 2 moral judgments is commonly called Tit for Tat. Tit for Tat

is based in the decision-rule, "make a cooperative overture, then copy the response of your partner in subsequent exchanges." If your partner cooperates, you glean the benefits of the cooperative exchange. If your partner behaves selfishly, you cut your losses (and his or hers!) by behaving selfishly until he or she relents and behaves cooperatively. This aspect of the strategy is prescribed by Stage 2 moral judgments such as "an eye for an eye" and "don't get mad, get even." Computer simulations of the evolution of social strategies have found that Tit for Tat can defeat more selfish strategies if it enters populations in sufficiently large numbers (Axelrod & Hamilton, 1981). Although Tit for Tat always loses to selfish strategies one on one, it ends up winning evolutionary contests by cutting its losses in exchanges with selfish strategies and reaping the benefits of drawing with cooperative strategies, including itself. (When strategies evolve, they become increasingly prevalent in populations; thus, evolved social strategies are usually designed in ways that enable them to do well in exchanges with replicas of themselves.).

Trivers (1971), Dugatkin (1997) and others have adduced evidence demonstrating that mechanisms giving rise to reciprocity have evolved in several species. The exchange of blood in vampire bats is a particularly interesting case in point (Wilkinson, 1990). Gouldner (1960) and others have suggested that reciprocity is a universal moral norm in the human species. As suggested by Trivers (1971) reciprocity is rooted in human psychological systems giving rise to a sense of deserving, a sense of gratitude and indebtedness, and senses of righteous indignation, retribution, revenge, and vindictiveness as well as systems instilling a sense of fairness and justice.

The Evolution of Stage 2/3 Strategies

Stage 2/3 strategies prescribe giving those who have transgressed against you a second chance if they have "suffered enough" or if they "feel bad or sorry about cheating you." Such

strategies are reflected in aphorisms such as "forgive and forget" and "everyone makes mistakes."

Following the publication of Axelrod and Hamilton's (1981) findings, investigators changed parameters in Axelrod and Hamilton's computer simulations and examined the fecundity of other strategies. In general, the closer the environments approximated the actual conditions of evolution, the greater were the benefits from conditionally altruistic strategies. For example, games that allowed for the inevitable errors that occur in social exchanges found that strategies such as "Two Tits for a Tat," "Generous Tit for Tat," "Contrite Tit for Tat," and "Forgiving Tit for Tat" fared better than Tit for Tat because they were equipped to break the self-defeating blood feuds precipitated by one selfish mistake (see Krebs, 2000; Ridley, 1996, for reviews of relevant research). Simulations that enabled players to observe other players and keep track of their strategies (called "image scoring") favored the evolution of altruism through indirect reciprocity (Nowak & Sigmund, 1998).

The Evolution of Stage 3 Strategies

Among the Stage 3 moral judgments classified in Kohlberg's system, there are at least two distinguishable types of strategy. Translated into biological terms, the first one prescribes enhancing one's inclusive fitness by helping members of one's group. The second prescribes conforming to moral norms. The adaptive benefits of the first type of strategy vary in accordance with the relationship between the helper and the recipient of help.

Helping relatives. Many of the moral judgments contained in Kohlberg's scoring manual prescribe upholding families and helping relatives. It is easy to account for the evolution of mechanisms giving rise to such strategies. It is quite common among mammals for parents to sacrifice their somatic interests for the sake of their offspring. Such self-sacrificial behaviors

help the parents propagate their genes. In a classic paper, Hamilton (1964) pointed out that the biological value of parental investment can be extended to relatives other than offspring. The probability of individuals sharing genes varies in proportion to their degree of relatedness. For example, the probability of parents and their offspring sharing genes is .5. Same for siblings. The probability of grandparents and their grandchildren sharing genes is .25; and so on. Hamilton (1964) explained how a decision-rule could evolve that induced individuals to help others when the coefficient of their relatedness is greater than the cost to the helper of helping, divided by the benefits to the recipient (r>c/b). Alcock (1998) and others have reviewed evidence that members of many species behave in accordance with the strategy implicit in Hamilton's equation.

Helping individuals who seem similar to relatives. Mechanisms that direct individuals to help their relatives are limited morally because they give rise to nepotistic and discriminatory behaviors, inducing individuals to favor relatives over non-relatives, close relatives over distant relatives, and more fecund relatives over less fecund relatives. This helps explain the imbalance of helping between parents and children, and grandparents and grandchildren. However, perhaps fortunately, animals may not be able to detect the proportion of genes they share with others directly or exactly (Dawkins, 1989, but see Pfennig & Sherman, 1995 and Rushton, 1999). Animals may have to rely on kin recognition cues such as phenotypic similarity, familiarity, and proximity (Porter, 1987). Although such cues may have been highly predictive of genetic relatedness in ancestral environments, they may activate mechanisms that induce individuals to help non-relatives in modern environments. Inasmuch as the genetic payoffs to individuals from helping non-relatives are less than those from helping their relatives, helping non-relatives can be considered more altruistic than helping relatives (see Krebs, 1987).

Helping friends. Many Kohlbergian moral judgments prescribe helping friends. In part, the evolution of mechanisms that dispose individuals to help their friends can be accounted for in terms of the "misfiring" of kin-selected strategies. Friends usually resemble kin; friends tend to be similar to one another, familiar, and live in relatively close proximity. In addition, such mechanisms probably reaped adaptive benefits through more complex forms of reciprocity than those prescribed by Stage 2 moral judgments.

Tooby and Cosmides (1996) have argued that exchanges between friends do not usually conform to decision-rules such as those that define Tit for Tat strategies. Friends do not pay each other back for every favor they bestow on one another, and they do not typically seek revenge when their partners fail to repay every debt. The give and take between friends is more general, averaged over many kinds of resource and over long periods of time. Tooby and Cosmides (1996) discuss the biological significance of a phenomenon they call the "banker's paradox": banks are least likely to lend people money when they most need it and are least likely to be able to pay it back. In Tooby and Cosmides' view, investing in friends is a strategy aimed at increasing the probability that people will have "bankers" who are willing to lend them "money" when they are in need. Accumulating resources counts for nothing in evolution if they do not enhance individuals' chances of surviving, reproducing, and propagating their genes. Bestowing relatively low cost favors on friends over long periods of time can be a winning strategy if it induces them to help save your life, help you find a mate, or foster the fitness of your relatives. In this sense, helping friends is akin to investing in insurance.

Helping In-group members. Kohlbergian moral judgments also prescribe helping members of one's in-group. Such strategies could stem from mechanisms of social control, which I discuss later, from extensions of kin-selected mechanisms and from extensions of

mechanisms designed to reap benefits from reciprocity. With respect to extensions of kin-selected mechanisms, Sober and Wilson (1998) explain how, under certain (admittedly rare) circumstances, mechanisms disposing individuals to behave in genetically altruistic ways—that is to say, in ways that enhance the fitness of members of their group at the expense of their own inclusive fitness—can evolve when the group as a whole benefits more than groups containing less altruistic members. The evolutionary process involved is similar to kin selection inasmuch as it is based in propagating genes that dispose individuals to behave altruistically by helping other individuals who possess replicas of such genes. However, it differs from kin selection, because the mechanisms that mediate altruism are not calibrated in ways that favor those who are genetically-similar to the altruist. In group selection, the costs of helping those who do not possess the genes that give rise to altruism are outweighed by the benefits to the group as a whole, and selfish members are reduced in proportion when groups recombine (see Sober & Wilson, 1998).

Mechanisms that induce individuals to help members of their groups also could have evolved through the benefits of complex systems of reciprocity. Alexander (1987) has argued that the biological function of moral dispositions is to uphold systems of indirect reciprocity. Such systems may assume two forms. In the first, person A helps person B; person B helps person C; and person C helps person A. In the second, all members of a group contribute resources to a central distributor who redistributes the resources to those who have contributed. Systems of indirect reciprocity have significantly more potential than systems of direct exchange to maximize benefits for all participating members, because they enable each individual to obtain resources of maximum value to them and minimum cost to those who produced them in exchange for resources of minimum cost to them and maximum benefit to recipients. Such

systems encourage individuals to specialize in the cultivation of particular resources, thus reducing the costs of producing them.

The potential gains of systems of indirect reciprocity notwithstanding, mechanisms giving rise to the behaviors necessary to support them could not have evolved without safeguards equipped to protect participating members from being exploited by selfish individuals disposed to give less than their share and to take more than their share. Such safeguards require the detection and punishment of cheaters. Alexander (1987) has outlined three conditions that foster the evolution of systems of indirect reciprocity: (a) members of groups show a preference for givers over takers as exchange partners, (b) members of groups reward altruists and their relatives by bestowing honors on them, and (c) the success of the groups to which altruistic individuals belong enhances their fitness and the fitness of their relatives. Conversely, cheaters must be punished by rejection, ostracism, losses in prestige, and negative effects on the group that filter back to the cheater and his or her relatives.

As mentioned earlier, the game theorists Nowak and Sigmund (1998) have demonstrated that altruism can evolve through the benefits of indirect reciprocity. In the model they created, they controlled for the benefits of direct reciprocity by minimizing the probability that members of groups would interact with each other more than once. Behaving altruistically enhanced an individual's reputation or "image" and behaving selfishly degraded it. Nowak and Sigmund found that if members of groups showed a preference for those with a good reputation, altruism could evolve and become evolutionarily stable. The process that gives rise to the adaptive benefits of indirect reciprocity is similar to the process that gives rise to the adaptive benefits of Tit for Tat: individuals reap the benefits of cooperating with cooperative members of their groups and avoid the costs of interacting with selfish members of their groups. The difference between

the two processes is that in systems of indirect reciprocity, individuals are able to detect selfish members of their groups by observing them and learning about their selfish behavior from others, thus avoiding the costs of being exploited by them directly. The strategy of helping members of one's group publicly in order to cultivate a good reputation is explicitly prescribe by Stage 3 moral judgments such as, [help others] "in order to leave a good impression on the community" (Colby & Kohlberg, 1987).

Conforming to moral norms. In addition to moral judgments that encourage people to enhance their inclusive fitness by helping members of their groups, Stage 3 moral judgments encourage people to conform to moral norms. Moral norms are defined as widely-practiced types of behavior that members of groups consider right and obligatory. Some moral norms, such as keeping promises and paying debts, appear to be universal (Brown, 1991; Gouldner, 1960; Krebs & Janicki, 2003). Other moral norms, such as those that pertain to food prohibitions, are specific to particular cultures. In my earlier discussion of the evolution of moral judgment, I offered an explanation for the origin of universal moral norms. Here I make the point that humans are naturally-inclined to conform to existing moral norms because conformity pays off. Social learning theorists have demonstrated that people tend to imitate the behavior of those who are successful, powerful, and of high status (Alexander, 1987; Boyd & Richerson, 1985; Burton & Kunce, 1995). Viewed biologically, observing how others behave and copying the behavior of those who fare well would tend to be an adaptive strategy because it would enable individuals to modify their behavior without suffering the potentially adverse consequences of trial and error. The more frequent a type of behavior in a population, the more likely it is to be adaptive. In addition to the benefits of vicarious learning, conforming to moral norms may increase individuals' security by reinforcing their social identity, helping them fit in, and enabling them to

avoid the costs of the social sanctions inflicted on those who violate the norms.

The Evolution of Stage 4 Strategies

Stage 4 moral judgments exhort people to obey the law and to conform to moral norms in order to maintain social institutions that promote the common good and "provide benefits and protection to all members of society." Stage 4 moral judgments uphold more complex systems of cooperation than Stage 3 moral judgments do. In Stage 4 systems, individuals do not know most of other members of their societies, even by reputation. Stage 4 systems are based on complex divisions of labor in which individuals give to and receive from central distributors. A common currency such as money enhances the efficiency of such systems.

It is easy to see that how Stage 4 systems of cooperation could produce more benefits than less complex systems to all participating individuals. However, it is equally easy to see that such systems are more susceptible to cheating than the systems upheld by lower-stage moral judgments. Without the antidotes inherent in Stage 2 and Stage 3 systems, Stage 4 systems need policing and legal institutions capable of detecting cheating and punishing cheaters.

If as most evolutionary theorists believe, our hominid ancestors lived in relatively small groups (Dunbar, 1966), they would not have experienced the adaptive problems or opportunities necessary for the selection of mechanisms specifically-designed to uphold Stage 4 systems of cooperation. I believe that the mechanisms that dispose individuals to behave in accordance with Stage 4 moral judgments are extensions of lower-stage mechanisms such as avoiding punishment (Stage 1), making one's best deal (Stage 2), cultivating a good reputation, conforming, and upholding systems of indirect reciprocity (Stage 3).

Could Structures Mediating Stage 5-6 Strategies Have Evolved?

Virtually all ultimate moral principles espoused by philosophers of ethics, including those

that define Kohlberg's Stages 5 and 6, are based in two prescriptions: (a) maximize benefits to humankind and (b) allocate these benefits in a non-discriminatory way. Indiscriminate cooperation and indiscriminate altruism meet these criteria. Although such strategies could maximize the benefits for everyone if everyone practiced them, I do not believe they have evolved, because they are vulnerable to cheating, nepotism, and discrimination against outgroups.

The vast majority of theorists who have examined the evolution of morality have concluded that mechanisms designed to induce individuals to adopt the strategies prescribed by Stage 5 and Stage 6 moral judgments could not evolve. Alexander (1987), who concluded that a "modicum" of indiscriminate beneficence could have evolved through indirect reciprocity seems to be an exception, but the evidence he adduces supports only the evolution of Stage 3 in-group and Stage 4 national systems of indirect reciprocity, with no extension to outgroups or to all of humanity.

Kohlberg's Stages 5 and 6 are different from his first four stages in that they are "colder" and more logical. There is virtually no mention of affect in any of Kohlberg's Stage 5 moral judgments. At least one of Kohlberg's collaborators, John Gibbs (see Gibbs, Basinger, & Fuller, 1992) has concluded that Stages 5 and 6 are "metatheoretical" forms of reasoning, quite different from the forms of reasoning in earlier stages. Researchers have failed to find any evidence of Stage 5 or 6 moral judgments about hypothetical dilemmas in non-Western cultures, and in the many studies my colleagues and I have conducted, we have yet to observe any participants making Stage 5-6 moral judgments about real-life moral conflicts that have consequences for the parties involved.

Implications for Moral Development

Kohlberg's model is a model of moral development; that is to say, how people become increasingly moral over time. What implications does an evolutionary reconceptualization of Kohlberg's model have for ontogenetic moral development? In the animal kingdom, different species develop in different ways. The new-born of some species are essentially the same as adults; they are fully formed and ready to go. In other species, individuals undergo qualitative transformations as they develop. For example, caterpillars change into butterflies. In the human species, females develop breasts at puberty. Viewed from the perspective of evolution, the reason why members of some species change as they develop is because they face different kinds of adaptive problems at different phases in their life cycles. As a generalization, the early lives of mammals are dominated by survival problems; when they reach sexual maturity, their lives become dominated by reproductive problems, and in old age, they face problems associated with parental investment.

With respect to morality, we would expect humans to acquire mental mechanisms giving rise to different social strategies in a sequence determined by the types of social problems their hominid ancestors faced at different phases in their lives. In terms of this expectation, the reason why young children acquire Stage 1 strategies prescribing obedience to authority is because obeying authority is an adaptive strategy for relatively small, weak, and vulnerable people. The reason why older children acquire Stage 2 instrumental exchange strategies is because such strategies reap greater gains in relations with peers.

In Piaget's (1932) pioneering work on the development of moral judgment, he concluded that young children tended to view morality in terms of obedience to adults for two reasons. First, as emphasized by Kohlberg, young children do not possess the cognitive sophistication

necessary to understand reciprocity. Second, neglected by Kohlberg but emphasized in evolutionary models, young children's social relations are dominated by adults. According to Piaget, the reason why older children acquire a new moral orientation based in cooperation is not only because they develop the capacity to understand cooperation, but also because cooperative exchanges are more adaptive than obedience in relations with peers. Contemporary developmental psychologists such as Damon and Hart (1992) and Youniss (1986) have concluded that research evidence supports the more Piagetian interpretation.

During adolescence, children enter new social worlds dominated by relations with the opposite sex and long-term friendships, which activate mechanisms that give rise to Stage 3 moral judgments. During adolescence, social image, reputation, and fear of ostracism become salient (Brown, Lohr & McClenahan, 1986; Krebs & Van Hesteren, 1994). And as adolescents grow into adults, they enter other social worlds, governed by other "moral orders" (Harre, 1984) upheld by Stage 4 moral judgments. As put by Alexander (1987),

> I see Kohlberg's Stage 4 as representing a transition from being primarily a rule-follower to being also concerned with rule-enforcement. This interpretation is consistent with the idea that after having learned and followed the rules one's self, having invested in the system, and having produced and instructed relatives with whose welfare one is ultimately concerned, there is reproductive value in ensuring that one's investment is safe, i.e., that the rules do not change" (p. 134).

Research evidence is consistent with the conclusion that adults who have invested in their social system and who stand to gain from it make Stage 4 judgments, but those who do not stand to gain make lower-stage judgments upholding their own interests and the interests of their ingroups (Rest, 1983).

Additive stage acquisition. If structures upholding particular systems of cooperation evolved because they were adaptive with respect to particular types of social relations, we would

expect people to retain the structures as long as they engaged in the types of social relations the structures evolved to support. Therefore, as discussed, in place of Kohlberg's assumption that new stage-structures transform and displace old stage-structures, we would expect new stage-structures to be added to older stage-structures, and older stage-structures to be retained and activated in the types of context in which they were selected in ancestral environments. Krebs and Van Hesteren (1994) have reviewed research supporting this expectation. As examples, adults who make high-stage moral judgments on Kohlberg's test sometimes make Stage 1 judgments in real-life contexts involving relations with powerful authorities (Newitt & Krebs, 1999). Adults behave in the ways prescribed by Stage 1 moral judgments in military contexts, cults, and the contexts created by Milgram (1964) in his classic studies on obedience to authority. Adults invoke Stage 2 strategies in the context of business transactions (Carpendale & Krebs, 1992), Stage 3 strategies when interacting with members of their family and other ingroups (Krebs & Van Hesteren, 1994), and Stage 4 strategies in societal contexts.

On the basis of findings from many studies (see Wark & Krebs, 1997 for a review) we concluded that the moral judgments people make and the moral behaviors they display in their everyday lives are the product of an interaction between the structures of moral judgment they have acquired ontogenetically and the environmental, situational, or contextual factors that govern the activation of these structures. It is misguided to assume that people are "in" stages of moral development, except perhaps young children who do not have the cognitive sophistication to make high-stage moral judgments. It is more correct to assume that people acquire an increasingly broad range of strategies that enable them to adapt to an increasingly broad range of social contexts. It is misguided to assume that people develop general "structures of the whole;" and more correct to assume that they acquire domain-specific structures that evolved in ancestral

environments (Buss, 1999).

Inclusiveness. Kohlberg (1984) has argued that new stage-structures incorporate or integrate older stage-structures within them. Viewed from an evolutionary perspective, this argument is valid inasmuch as more complex systems of cooperation are built upon and include less complex systems of cooperation. For example, it seems quite likely that Stage 4 systems of indirect reciprocity could not have evolved and cannot be maintained without support from Stage 1 obedience strategies, Stage 2 instrumental exchange strategies and Stage 3 impression-management strategies. An evolutionary perspective also offers a clear and simple explanation for why high-stage structures are "better" than low-stage structures—they prescribe strategies that uphold systems of cooperation equipped to produce greater benefits for all contributors.

The Pinnacle of Moral Development

As I have argued, the evidence does not suggest that we are evolved to adopt strategies such as, "give to everyone according to his need," "do unto others as you would have them do unto you," "behave in a way that maximizes the greatest good for the greatest number", though we may well be disposed to preach such strategies to others. This is profoundly tragic, because if we were evolved to uphold the systems of cooperation supported by such strategies, we would be much better off than we are. If we all upheld such systems, social relations would be harmonious; everyone would help and support each other, regardless of race, creed, or color. There would not be any crime or wars. And we could invest the money we saved from the arms race, police, and jails in enhancing the quality of our lives. The problem with systems of cooperation upheld by Stage 5 and Stage 6 moral judgments is that they do not contain any antidotes to selfish and discriminatory strategies, and thus are destined to fail. We may well be evolved to uphold systems of indirect reciprocity in relatively small groups of people who know

one another by reputation. In such groups, it is possible that low-stage strategies could support some indiscriminate altruism (Alexander, 1987), but the greater the proportion of indiscriminate altruists, the greater the vulnerability of this strategy to more selfish and discriminatory strategies.

Summary and Conclusions

In this chapter, I outlined Kohlberg's model of moral development, reviewed research that called it into question, concluded that Kohlberg's model pertains to only one of many functions of making moral judgments, and argued that reconceptualizing Kohlbergian structures of moral judgment as evolved decision-making mechanisms gives rise to a more ecologically-valid model. I proposed that the mechanisms that give rise to moral decisions evolved to help our hominid ancestors resolve the inevitable conflicts of interest that arose when they banded together and established systems of cooperation to foster their biological interests. I argued that people use moral judgments to induce others to behave in ways that foster their interests and that recipients of moral judgments are agents of selection. Of the many possible kinds of moral judgment, those that define Kohlberg's stages of moral development were selected because they upheld behavioral strategies that fostered the interests of those who sent them and those who received them. In addition to using moral judgments to influence the behavior of others, people use moral judgments to guide their own behavior. Although moral judgments directed toward the self are more self-serving than moral judgments directed toward others, there are several constraints on such self-serving biases.

If people are biologically disposed to behave in accordance with the kinds of moral judgments classified by Kohlberg, then contrary to the conclusions reached by several eminent evolutionary theorists, behaving morally need not be maladaptive and humans may be moral by

nature. In the second half of this chapter, I explained how mental mechanisms disposing people to adopt the social strategies prescribed by Stage 1, Stage 2 and Stage 3 moral judgments could have evolved and how the strategies prescribed by Stage 4 moral judgments follow naturally from those that define the earlier stages. I concluded that although it may pay off for people to preach Stage 5 and Stage 6 ethical principles, humans are not biologically disposed to behave in accordance with them.

I ended by considering the implications of the evolutionary model I outlined for moral development. Adopting a life-history perspective, I suggested that people acquire structures of moral judgment in the sequence that defines Kohlberg's stages of moral development because such structures helped their hominid ancestors solve adaptive problems that occurred at different phases of their lives. New structures do not transform and displace older structures, as Kohlbergians assume, because people continue to face the adaptive problems the earlier structures were designed to solve. I noted two general implications of my analysis of the evolution of moral dispositions. On the one hand, we are moral by nature because we are disposed to behave in accordance with the prescriptions of Stage 1, Stage 2, Stage 3 and Stage 4 moral judgments. On the other hand, we are not completely moral, because we are not disposed to behave in accordance with Stage 5 and Stage 6 moral principles. And there is a tragic irony to this, because we all could be better off than we are if we were.

References

Alcock, J. (1998). *Animal behavior: An evolutionary approach.* (6th ed.). Sunderland, MA: Sinauer Associates.

Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.

Aronfreed, J. (1968). *Conduct and conscience.* New York: Academic Press.

Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211,* 1390-1396.

Bartek, S., Krebs, D. L. & Taylor, M. (1993). Coping, defending, and the relations between moral judgment and moral behavior in prostitutes and other female juvenile delinquents. *Journal of Abnormal Psychology*, *102*, 65-73.

Boehm, C. (2000). Conflict and the evolution of social control. In L. D. Katz (Ed.), *Evolutionary Origins of Morality* (pp. 79-101). UK: Imprint Academic.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Brown. B. B., Lohr, M. J, & McClenahan, E. L. (1986). Early adolescents' perceptions of peer pressure. *Journal of Early Adolescence*, *6*, 139-154.

Brown, D. E. (1991). *Human universals.* NY: McGraw-Hill.

Brown, R. (1986). *Social psychology (*2nd ed). New York: Free Press.

Burton, R. V. & Kunce, L. (1995). Behavioral models of moral development: A brief history and integration. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Moral development: An introduction* (pp. 141- 172). Boston: Allyn and Bacon.

Buss, D. (1999). *Evolutionary psychology: The new science of the mind.* Boston: Allyn and Bacon.

Carpendale, J. and Krebs, D. (1992). Situational variation in moral judgment: In a stage or on a

stage? *Journal of Youth and Adolescence*, *21*, 203-224.

Carpendale, J. & Krebs, D. L. (1995). Variations in moral judgment as a function of type of

dilemma and moral choice. *Journal of Personality*, *63*, 289-313.

Colby, A., & Kohlberg, L. (Eds.) (1987*). The measurement of moral judgment* (Vols. 1-2).

Cambridge: Cambridge University Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans

reason? Studies with the Wason selection task. *Cognition*, *31*, 187-276.

Damon, W. (1977). *The social world of the child*. San Francisco:Jossey-Bass.

Damon, W. & Hart, D. (1992). Self understanding and its role in social and moral development.

In M. H. Bornstein & E. M. Lamb (Eds.), *Developmental psychology: An advanced*

textbook (2nd ed., pp. 421-465). Hillsdale, NJ: Erlbaum.

Dawkins, R. (1989). *The selfish gene.* Oxford: Oxford University Press.

Denton, K. and Krebs, D. (1990). From the scene to the crime: The effect of alcohol and social

context on moral judgment. *Journal of Personality and Social Psychology, 59,* 242-248.

Dugatkin, L. A. (1997). *Cooperation among animals: An evolutionary perspective.* New York:

Oxford University Press.

Dunbar, R. (1966). *Grooming, gossip, and the evolution of language*. London: Faber and Faber.

Gibbs, J., Basinger, K. S. & Fuller, D. (1992). *Moral maturity: Measuring the development of*

sociomoral reasoning. Hillsdale NJ: Lawrence Erlbaum.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American*

Sociological Review, *25*, 161-178.

Hamilton, W. D. (1964). The evolution of social behavior. *Journal of Theoretical Biology*, *7*, 1-

52.

Harre, R. 1984. *Personal being: A theory for individual psychology*. Cambridge MA: Harvard University Press.

Higgins, A. & Eccles-Parson, J. E. (1983). Social cognition and the social life of the child: Stages as subcultures. In E. T. Higgins, D. N. Ruble & W. W. Hartup (Eds.), *Social cognition and social development: A sociocultural perspective* (pp. 137-151). New York: Cambridge University Press.

Janicki, M. & Krebs, D. L. (1998). Evolutionary approaches to culture. In C. Crawford, C. & D. L. Krebs (Eds.) *Handbook of Evolutionary Psychology: Ideas, Issues, and Applications*, pp. 163-207. Hillsdale, NJ: Erlbaum.

Kohlberg, L. (1984). *Essays in moral development*: *Vol 2. The psychology of moral development*. New York: Harper & Row.

Kohlberg, L., & Candee, D. (1984). The relationship of moral judgment to moral action. In L. Kohlberg (Ed.), *Essays in moral development: Vol 2. The psychology of moral development*. New York: Harper & Row.

Krebs, D. (1988) The challenge of altruism in biology and psychology. In C. Crawford, M. Smith and D. Krebs (Eds.), *Sociobiology and psychology: Ideas, issues and applications.* Hillsdale, N.J.: Erlbaum, pp. 81-118.

Krebs, D. L. (1998). The evolution of moral behavior. In C. Crawford & D. L. Krebs (Eds.), *Handbook of Evolutionary Psychology: Ideas, Issues, and Applications* (pp.337-368). Hillsdale, NJ: Erlbaum.

Krebs, D. L. (2000a). The evolution of moral dispositions in the human species. In D. LeCroy & P. Moller (Eds.) Evolutionary Perspectives on Human Reproductive Behavior. *Annals of the New York Academy of Science, Vol. 907,* pp. 1-17.

Krebs, D. L. (2000b). Evolutionary games and morality. In L. D. Katz (Ed.) *Evolutionary Origins of Morality: Cross-disciplinary approaches* (pp. 313-321). UK: Imprint Academic.

Krebs, D. L. (2000c). As moral as we need to be. In L. D. Katz (Ed.) *Evolutionary Origins of Morality: Cross-disciplinary approaches* (pp. 139-143). UK: Imprint Academic.

Krebs, D. L. (2003). How to make silk purses from sows' ears: Cultivating morality and constructing moral systems. In C. Crawford, & C. Salmon (Eds.) *Evolutionary psychology: Public policy and personal decisions.* Mahwah NJ: Erlbaum.

Krebs, D. L. & Denton, K. (1997). Social illusions and self-deception: The evolution of biases in person perception. In J. A. Simpson & D. T. Kenrick (Eds.) *Evolutionary social psychology,* pp. 21-47. Hillsdale, NJ: Erlbaum.

Krebs, D. L. & Denton, K. (1999). On the relations between moral judgment and moral behavior. In D. Garz, F. Oser, & W. Althof (Eds.) *The context of morality*. Frankfurt: M. Suhrlcamp.

Krebs, D. L, Denton, K., & Higgins, N. (1988). On the evolution of self-knowledge and self-deception. In McDonald, K. (Ed.), *Sociobiological perspectives on human behavior,* pp. 103-139. Springer-Verlag.

Krebs, D. L., Denton, K., Vermeulen, S.C., Carpendale, J.I., and Bush, A. (1991). The structural flexibility of moral judgment. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences, 61*, 1012-1023.

Krebs, D. L., Denton, K., & Wark, G. (1997). The forms and functions of real-life moral decision-making. *Journal of Moral Education, 20,* 131-145.

Krebs, D. L., Denton, K., Wark, G., Couch, R., Racine, T. P., & Krebs, D. L. (2002).

Interpersonal moral conflicts between couples: Effects of type of dilemma, role, and partner's judgments on level of moral reasoning and probability of resolution. *Journal of Adult Development, 9*, 307-316.

Krebs, D. L. & Laird, P. (1998). Judging yourself as you judge others: perspective-taking, moral development, and exculpation. *Journal of Adult Development*, *5,* 1-12.

Krebs, D. L. & Janicki, M. (2003) The biological foundations of moral norms. In M. Schaller & C. Crandall (Eds.), *Psychological Foundations of Culture.* Hillsdale, NJ: Erlbaum.

Krebs, D. L. & Van Hesteren. (1994). The development of altruism: Toward an integrative model. *Developmental Review*, *14*, 1-56.

Krebs, D.L., Vermeulen, S.C., Carpendale, J.I., and Denton, K. (1991). Structural and situational influences on moral judgment: The interaction between stage and dilemma. In W. Kurtines and J. Gewirtz (Eds.). *Handbook of moral behavior and development: Theory, research, and application*, NJ: Erlbaum, 139-169.

Krebs, D., Vermeulen, S., & Denton, K.(1991). Competence and performance in moral judgment: From the ideal to the real. *Moral Education Forum*, *16*, 7-22

Krebs, D. L., Wark, G. & Krebs, D. L. (1995). Lessons from life: Toward a functional model of morality. *MoralEducation Forum*, *20*, 22-29.

Leaky, R. E., & Lewin, R. (1977). *Origins*. New York: Dutton.

MacDonald, K. B. (1988). Sociobiology and the cognitive-developmental tradition on moral development research. In K. B. MacDonald (Ed.), *Sociobiological perspectives on human development*, pp. 140-167.

Milgram, S. (1974). *Obedience to authority.* New York: Harper.

Miller, D. T. & Porter, C. A. (1988). Errors and biases in the attribution process. In L. Y.

Abramson (Ed.), *Social Cognition and Clinical Psychology: A synthesis (*pp. 3-32). New York: The Guilford Press.

Mitchell, R. W. & Thompson, N. S. (Eds.) (1986). *Deception: Perspectives on human and nonhuman deceit*. NY: State University of New York Press.

Newitt, C & Krebs, D. L. (1999). Structural and contextual sources of moral judgment. in preparation.

Nowak, M. A. & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573-577.

Pfennig, D. W., & Sherman, P. W. (1995). Kin recognition. *Scientific American, 272*, 98-103.

Piaget, J. (1932). *The moral judgment of the child*. London: Routledge & Kegan Paul.

Porter, R. H. (1987). Kin recognition: Functions and mediating mechanisms. In C. B. Crawford & D. L. Krebs (Eds.), *Sociobiology and psychology: Ideas, issues and applications (*pp. 175-205*)*. Hillsdale, NJ: Erlbaum.

Rawls, J. (1971). *A theory of justice.* Cambridge, MA: Harvard University press.

Rest, J. F. (1983). Morality. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed.), pp. 556-629. New York: Wiley.

Ridley, M. (1996). *The origins or virtue: Human instincts and the evolution of cooperation.* New York: Viking.

Rushton, J. P. (1999). Genetic similarity theory and the nature of ethnocentrism. In K. Thienpont & R. Cliquet (Eds.) *In-group/Out-group behavior in modern societies: An evolutionary perspective*, pp. 75-107. The Netherlands: Vlaamse Gemeeschap/CBGC.

Sloman, L. & Gilbert, P. (Eds.) (2000). *Subordination and defeat: An evolutionary approach to mood disorders and their therapy.* Mahwah, NJ: Erlbaum.

Sober, E. & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior.* Cambridge MA: Harvard University Press.

Strayer, F. F. & Strayer, J. (1976). An ethological analysis of social agonism and dominance relations among preschool children. *Child Development*, *47*, 980-989.

Tajfel, H. (1982). *Social identify and intergroup relations*. Cambridge, England: Cambridge University Press.

Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. New York: Basic Books.

Tooby, J. & Cosmides, L. (1996). Friendship and the banker's paradox: Other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy, 88,* 119-143.

Tooby, J., & Devore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. G. Kinzey (Ed.), *The evolution of human behavior: Primate models*. Albany, NY: Suny Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35-57.

Trivers, R. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.

Trivers, R. (2000). The elements of a scientific theory of self-deception. In D. LeCroy & P. Moller (Eds). *Evolutionary perspectives on human reproductive behavior*. *Annals of the New York Academy of Sciences, 907*, 114-131,

Wark, G. & Krebs, D. L. (1997). Sources of variation in real-life moral judgment: Toward a model of real-life morality. *Journal of Adult Development, 4,* 163-178.

Wilkinson, G. S. (1990). Food sharing in vampire bats. *Scientific American, Feb.,* 76-82.

Williams, G. C. (1989). A sociobiological expansion of Evolution and Ethics. In J. Paradis & G.

C. Williams (eds.), *Evolution and Ethics* (pp. 179-214). Princeton: Princeton University

Press.

Youniss, J. (1986). Development in reciprocity through friendship. In C. Zahn-Waxler, M.

Cummings, & R. Ianottie. Eds. *Altruism and aggression: Biological and social origins*

(pp. 88-106). Cambridge: Cambridge University Press.

Table 1

Kohlberg's Stages of Moral Development

Stage 1

Morality is defined in terms of avoiding punishment, respecting the "superior power of authorities," "obedience for its own sake," and "avoiding damage to persons and property."

Stage 2

Morality is defined in terms of instrumental exchange; "acting to meet one's own interests and needs and letting others do the same," making deals, and engaging in equal exchanges.

Stage 3

Morality is defined in terms of upholding mutual relationships, fulfilling role expectations, being viewed as a good person, sustaining a good reputation, showing concern for and caring for others, and interpersonal conformity. Trust, loyalty, respect, and gratitude are important moral values.

Stage 4

Morality is defined in terms of maintaining the social systems from which one benefits, obeying their rules and laws, and "contributing to society." Morality involves doing one's share to uphold society and to prevent it from breaking down.

Stage 5

Morality is defined in terms of fulfilling the social obligations implicit in social contracts that are "freely agreed upon", and a "rational calculation of overall utility, 'the greatest good for the greatest number'." Morality involves orienting to the welfare of all and the protection of

everyone's rights.

Stage 6

Morality is defined in terms of following "self-chosen universal ethical principles of justice" that uphold "the equality of human rights and respect for the dignity of human beings as individual persons." Morality involves treating individuals as ends in themselves. (Colby & Kohlberg, 1987, p. 18-19)

Table 2

Kohlberg's Elements of Morality

Modal Elements

1. Obeying persons or deity

2. Blaming and approving

3. Retributing and exonerating

4. Having a right

5. Having a duty

Value Elements

Egoistic Consequences
6. Maintaining reputation

7. Seeking reward and avoiding punishment

8. Promoting good individual consequences

9. Promoting good group consequences

Ideal or Harmony-serving Consequences
10. Upholding character

11. Upholding self-respect

12. Serving social ideals or harmony

13. Serving human dignity and autonomy

Fairness
14. Role-taking; balancing perspectives

15. Reciprocity or positive desert

16. Maintaining equity

17. Maintaining social contract or freely agreeing