Prepublication draft: In Buss, D. (Ed.) *Evolutionary Psychology Handbook*, John Wiley & Sons.

The Evolution of Morality

Dennis Krebs
Simon Fraser University

Many evolutionary theorists have doubted whether moral dispositions can evolve through natural selection (Campbell, 1978; Darwin, 1871; Dawkins, 1989; Huxley, 1893). For example, according to Williams (1989), "There is no encouragement for any belief that an organism can be designed for any purpose other than the most effective pursuit of [its] self-interest….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". (pp. 195-197). In this chapter I will argue that the idea that all organisms are inherently selfish and immoral by nature is wrong, or more exactly only half right. I will explain how mechanisms that give rise to moral and immoral behaviors can evolve, and I will adduce evidence that they have evolved in the human species and in other species as well.

WHAT IS MORALITY?

In large part, the conclusions scholars reach about the evolution of morality are determined by the standards they believe an act must meet to qualify as moral. If scholars insist that a behavior must be genetically unselfish to qualify as moral, they will almost certainly infer that moral dispositions cannot evolve. If, on the other hand, they define morality in terms of individual unselfishness, they will almost certainly reach a more positive conclusion. It is, therefore, important to be clear about what we mean by morality. Everyone makes moral judgments about the goodness and badness of people, the rightness and wrongness of behaviors, and the rights and duties of members of groups. At a phenotypic level, most people agree about which kinds of behavior are moral and immoral. For example, virtually everyone considers helping others, keeping promises, and being faithful to one's spouse moral, and virtually everyone considers murder, rape, lying, and cheating immoral. However, if you ask people what makes such behaviors moral or immoral, they may well give different reasons, exposing significant differences in their underlying conceptions of morality.

Cognitive-developmental psychologists such as Colby and Kohlberg (1987), Damon and Hart (1992), and Piaget, (1932) have found that the conceptions of morality harbored by children and adults from a wide array of cultures tend to change systematically as people develop, in the stage-like ways outlined in Table 1. (See Krebs & Van Hesteren, 1994, for a comparison between Kohlberg's stages of moral development and the stages derived by other theorists). Based on empirical evidence supporting his developmental sequence and philosophical criteria of morality such as universality, prescriptiveness, and impartiality, Kohlberg (1984) concluded that the conceptions of morality that define higher stages in his sequence are more adequate than the conceptions that define lower-stages.

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)

Although Kohlberg's model of moral development is limited in several ways (see Gilligan, 1984; Krebs, 2004b; Krebs, Denton, Vermeulen, Carpendale, & Bush, 1991), there is strong and consistent support for his contention that most people view morality in the ways outlined in Table 1 and that most people believe that the conceptions that define higher stages in Kohlberg's sequence are more adequate than the conceptions that define lower stages. I will use these conceptions as working definitions of morality, re-interpreting them in more biological terms.

Selfishness and Morality

If we define unselfish as refraining from fostering one's interests at the expense of the interests of others, the assumption that an act must be unselfish to qualify as moral seems reasonable. Most moral behaviors seem unselfish; most immoral behaviors seem selfish; and behaviors prescribed by moral judgments that define relatively high stages in Kohlberg's sequence seem more unselfish than behaviors prescribed by lower-stage judgments.

It is, however, important to note that the type of unselfishness that most people assume is necessary for morality is different from the type of unselfishness that evolutionary theorists believe defies the laws of evolution (Sober & Wilson, 2000). The interests that people have in mind when they make attributions about morality are the proximate physical, material, and hedonic interests of people making moral decisions. The interests that evolutionary theorists have in mind are ultimate genetic interests. Although the two types of interest may covary, they need not necessarily correspond. Indeed, as recognized by Dawkins (1989), "there are special circumstances in which a gene can achieve its own selfish goals by fostering a limited form of altruism at the level of individual animals" (p. 6).

It is unreasonable to set genetic unselfishness as a criterion for morality. It is not immoral to propagate your genes. Morality pertains to how you go about accomplishing this task. Attempting to propagate your genes in individually selfish ways, at the expense of the physical, material, or psychological welfare of others, is immoral, but attempting to propagate your genes in individually cooperative or altruistic ways that foster the welfare of others is moral. Even if all evolved mechanisms disposed people to behave in genetically-selfish ways, need not necessarily be the case, it would not render such behaviors immoral. With this conception of morality in mind, let us turn to the central questions addressed in this chapter, can mechanisms that give rise to moral behaviors evolve and if so have they evolved in the human species?

THE ORIGIN OF MORALITY

The central contribution evolutionary psychology brings to the understanding of morality is to encourage us to ask what adaptive problems it was selected to solve. What functions did morality serve in ancestral environments? I submit that the mechanisms that give rise to moral behaviors evolved to solve the social problems that inevitably arise when individuals band together to foster their interests. When individuals are able to satisfy their needs, to survive, to reproduce and to rear their offspring on their own, there is no need for them to interact with other members of their species, and therefore no need for morality. Mechanisms that induce individuals to form groups and socialize with others were selected because such social behaviors were adaptive in ancestral environments.

The Significance of Cooperation and Conflicts of Interest

Social behaviors may help animals adapt to their environments in many ways. As examples, aggregating and mutual defense may reduce the risk of predation, and group hunting may enhance the probability of obtaining food. Most benefits of sociality stem from cooperative exchanges. However, as explained by the philosopher John Rawls (1971, p. 4) in his widely-cited book, *Theory of Justice*,

> 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.

Selfish preferences pose a problem for the evolution of cooperation, because they tempt individuals to invoke selfish strategies that, if successful, can drive cooperative strategies into extinction. However, when selfish strategies are successful, they tend to increase in frequency, which elevates the probability of them encountering other selfish strategies and engaging in low pay-off "me-me" exchanges. Individuals bent on doing less than their share and taking more than their share may end up fighting, failing to obtain resources, failing to defend themselves, and failing to rear fecund offspring. This, I believe, is the adaptive problem that the mechanisms that give rise to morality evolved to solve. The biological function of morality is to uphold fitness-enhancing systems of cooperation by inducing members of groups to contribute their share and to resist the temptation to take more than their share; to do their duties and to exercise their rights in ways that do not infringe on the rights of others; to resolve conflicts of interest in mutually-beneficial ways.

A BIOLOGICAL CONCEPTION OF MORALITY

Viewing morality in this way helps elucidate its nature. Morality boils down to meeting one's needs and advancing one's interests in cooperative ways. Morality consists in "standards or guidelines that govern human cooperation—in particular how rights, duties, and benefits are … allocated… Moralities are proposals for a system of mutual coordination of activities and cooperation among people" (Rest, 1983, p. 558). In this conception, acts such as murder, rape, and infidelity are immoral for the same reason as acts such as lying and cheating: They advance individuals' interests at the expense of the interests of others and undermine systems of cooperation. The moral judgments that define different stages of moral development (Table 1) uphold different systems of cooperation. The higher the stage, the greater the system's potential to maximize benefits for everyone involved—to produce the greatest good for the greatest number. The question is, could mechanisms that dispose individuals to behave in the ways prescribed by the moral judgments that define each of Kohlberg's stages have evolved?

THE NATURAL SELECTION OF SOCIAL STRATEGIES: EVOLUTIONARY GAMES

Imagine a group of early humans living in an ancestral environment. Assume that all members of the group inherit genes that guide the creation of mechanisms that give rise to strategies designed to maximize their biological benefits from interacting with others. Although winning strategies become more frequent in the population, the process of natural selection need not necessarily drive competing strategies into extinction. As a strategy increases in frequency, it induces changes in the social environment. In particular, it becomes increasingly likely to encounter replicas of itself, and this may affect its adaptiveness. For example, a "hawk" strategy that fares well against "dove" strategies may become increasingly costly as the proportion of "hawks" increases in the population. In addition to selecting only one strategy, the process of natural selection may induce strategies to fluctuate in frequency over generations or it may induce two or more strategies to stabilize in some proportion, between or within individuals (Maynard-Smith,1976).

Assume that some members of an ancestral group inherit mechanisms that dispose them to adopt selfish strategies and other members of the group inherit mechanisms that dispose them to adopt cooperative strategies, and that these strategies compete against each other. Which ones would increase in frequency and evolve? To answer this question, theorists have created models of social evolution such as those derived from Prisoners' Dilemma games.

Prisoners' Dilemma Models of Social Evolution.

In the simplest form of evolutionary adaptations of classic two-person iterated Prisoners' Dilemma games, each player is programmed to play one of two strategies—to behave selfishly (to "defect") or to behave cooperatively. If both players make a cooperative choice, each produces three cooperative offspring. If both players make a selfish choice, each produces one selfish offspring. If one player makes a cooperative choice and the other makes a selfish choice, the cooperative player does not produce any offspring and the selfish player produces five selfish offspring. Strategies are played off against each other in random order in computers. After the first round, or generation, offspring who inherit the strategies compete against each other, and so on. Game theorists seek to answer two questions, which strategy or strategies will evolve, and will any strategy become evolutionarily stable—that is to say, reach an equilibrium in the population such that it cannot be defeated by any competing strategy?

Prisoners' Dilemma games model several basic principles of social evolution. Pairs and groups of cooperating individuals fare better than pairs and

groups of selfish individuals. In addition, each member of a pair or group of cooperating individuals fares better than each member of a selfish dyad or group of selfish individuals. However, within a dyad or group of cooperators, selfish individuals fare better than cooperative individuals. Note how the Prisoners' Dilemma is equipped to model individual-level selection within groups and group-level selection between groups (Dugatkin & Reeve, 1994; Sober & Wilson, 1998).

THE EVOLUTION OF SELFISHNESS

On the contingencies of simple Prisoners' Dilemma models of social evolution, selfish players end up producing twice as many offspring as cooperative players. If Prisoners' Dilemma models programmed in this way validly represented the process of evolution, selfish strategies would drive cooperative strategies into extinction, and render all species selfish by nature, as many eminent evolutionary theorists have concluded. But before we accept this conclusion, we need to realize that the social context, choices, and parameters modeled in simple Prisoners' Dilemma games differ in significant ways from the contexts, choices, and parameters in which the social strategies inherited by many species evolved. Changing the parameters of Prisoners' Dilemmas to make them approximate more closely the conditions in which social strategies were selected in human and other species can decrease the adaptive benefits of selfish strategies and increase the adaptive benefits of cooperative strategies. I will demonstrate how the strategies prescribed by moral judgments that define Kohlberg's first four stages of moral development could have defeated more selfish strategies in the ancestral environments in which our hominid ancestors evolved.

STAGE 1 MORALITY: THE EVOLUTION OF DEFERENCE

In simple Prisoners' Dilemma models of social evolution, all players are equal in power. In contrast, in the real world animals differ in power and make conditional decisions that depend on the relative power of their opponents. Adopting a selfish strategy, defined as attempting to get more than one's share, may prove costly for relatively weak members of groups.

As shown in Table 1, Kohlbergian Stage 1 moral judgments prescribe deferring to those with "superior power" and "obeying authority" in order to "avoid punishment." Clearly, the strategy implicit in such judgments could be more adaptive than more blindly selfish or aggressive strategies for relatively subordinate members of groups. When members of groups are faced with a choice between competing against more powerful members of their groups or subordinating their interests to them, discretion is often the better part of valor (Cummins, Chapter XX). Adopting a deferential strategy enables subordinate members of groups to make the best of a bad situation and live to fight another day. Deferential strategies also may benefit subordinate members of groups by enhancing the fitness of more powerful members who, in turn, intimidate predators or foes.

Deferential Strategies in Humans and other Animals

Members of species ranging from crickets (Dawkins, 1989) and crayfish (Barinaga, 1996) to chimpanzees (Boehm, 2000) have been found to adopt conditional strategies such as, "if your opponent seems more powerful than you, defer to him or her; if your opponent seems less powerful than you, intimidate him or her." Such strategies give rise to dominance hierarchies, or pecking orders. Dominant and submissive behaviors are correlated with changes in levels of testosterone and serotonin in a variety of animals (see Buss, 1999, for a review of the literature).

In his pioneering book on moral development, Piaget (1932) attributed the moral orientation of young children to "the respect felt by the small for the great" (p. 107) which "has its roots deep down in certain inborn feelings and is due to a sui generis mixture of fear and affection" (p. 375). Researchers have found that children organize themselves into dominance hierarchies as young as three years of age (Cummins, 1998).

Neglected by Kohlberg (1984) and other developmental psychologists is evidence that deferential dispositions in adults stem from the same mechanisms as deferential dispositions in children. Adults may experience the same sense of awe, unilateral respect, and intimidation as children do when they encounter powerful people of high status. Milgram's (1974) classic studies demonstrate that people are more prone to submit to authority than is commonly assumed. Members of cults such as Heaven's Gate and Jonestown have proved themselves willing to commit suicide on the commands of their leaders (Osherow, 1981). Deference also may be evoked by more abstract entities, such as God.

The Morality of Deference

In one sense, deference is unselfish, because it induces individuals to subordinate their interests to those of others. However, in another sense deferential strategies are selfish, because they enable those who employ them to avoid punishment and maximize their chances of surviving and reproducing. In general, deference is individually unselfish with respect to immediate decisions, but individually selfish in the long-term—physically, materially, psychologically, and genetically.

Inasmuch as morality involves the constraint of selfishness, relatively powerful members of groups can be viewed as exerting a moralizing effect on relatively weak members of groups. However, there are at least two problems with the morality of hierarchical social systems. First, everyone except the individual at the bottom of the totem pole behaves selfishly toward those below him or her in the hierarchy. Second, there is no one to constrain the selfishness of the most dominant member of the group. To most people, it is more moral to constrain the selfishness of dominant members of groups than to reinforce it by acting submissively. Few people consider deference, submission and obedience to authority to be moral qualities in and of themselves; few people believe that it was moral for Nazis to obey authority.

The Significance of Coalitions and Mutual Control

One way in which relatively weak and subordinate members of groups can increase their power is to form coalitions. Although coalitions can exert a moralizing effect on groups by controlling the selfishness of the most powerful members, such effects are limited in two ways. The coalition may become tyrannical and it is in the adaptive interest of each member of the coalition to gain ascendancy over the other members of the coalition and take more than his or her share. To get to an egalitarian social system, we need a social equilibrium produced either by individuals or groups controlling one another's selfishness or by members of groups constraining their own selfishness—that is to say, resisting the temptation to dominate subordinates even when it is not in their immediate interest (Boehm, 2000).

STAGE 2 MORALITY: THE EVOLUTION OF DIRECT RECIPROCITY

Deferential strategies do not offer effective ways of resolving conflicts of interest between individuals who are relatively equal in power, because neither is inclined to defer and each is able to inflict damage on the other. When resources can be divided, it may be more beneficial for peers to share them than to compete for them. When resources cannot be divided, peers may be better off taking turns than fighting. Mechanisms that give rise to sharing and turn-taking strategies will evolve when the net benefits from settling for part of a resource outweigh the net benefits of competing for the whole thing.

The Evolution of Concrete Reciprocity

In classic Prisoners' Dilemma games all cooperative players reap exactly the same payoff from exchanges with other cooperators—three offspring. In contrast, in the real world, the goods and services that people exchange may vary in value. Individuals may exchange items worth relatively little to them for items that are worth considerably more, enabling all parties to gain in trade. Inevitably, members of groups encounter others who need services that they can provide at relatively little cost to themselves. As Trivers (1971) explained, it can be in individuals' interest to help others if such helping increases the probability that the recipients will help them when they are in need. However, in order for psychological mechanisms that induce individuals to reciprocate to evolve, they must contain antidotes to cheating that prevent selfish players from taking without giving in return. One strategy equipped to accomplish this is Tit-for-Tat.

Tit-for-Tat

Tit-for-Tat is based in the decision rule, "be nice, then get even." Invite mutually-beneficial reciprocal exchanges by making low-cost giving overtures to others, then copy their response. In contrast to more unconditionally altruistic or cooperative strategies, Tit-for-Tat gives rise to iterations of reciprocal exchanges between both givers and takers after the first exchange.

At first glance, it might seem that Tit-for-Tat is destined to lose to unconditionally selfish strategies because selfish strategies reap greater benefits than Tit-for-Tat on the first exchange (5 vs. 0 offspring) then tie with them (1-1) on all subsequent moves. Although this is the case in two-person games, Tit-for-Tat can end up defeating unconditionally selfish strategies if there is a relatively large number of Tit-for-Tat strategists in the population. In computer contests sponsored by Axelrod and Hamilton (1981), Tit-for-Tat defeated unconditionally selfish strategies and emerged the winner. The principle underlying

this outcome pertains to the benefits of cooperating with cooperators, which I believe was critically important in the evolution of morality.

Note that there is a fringe benefit from the evolution of Tit-for-Tat strategies, namely that it opens the door for the evolution of more unconditionally cooperative and altruistic strategies. Indeed, in an environment saturated by Tit-for-Tat strategists, one could not tell the difference between conditionally and unconditionally cooperative strategies because they would behave in the same cooperative manner. However, ironically, opening the door for unconditionally cooperative or altruistic strategies also opens the door for the reemergence of selfish strategies, which benefit by exploiting them. Selfish strategies thrive on the unconditional generosity of do-gooders.

Concrete Reciprocity in Humans and other Animals

Biologists have found that mechanisms giving rise to systems of Tit-for-Tat reciprocity have evolved in some species, though perhaps fewer than we might expect (Trivers, 1985; Dugatkin, 1998). With respect to humans, Trivers (1985) suggested that, "During the Pleistocene, and probably before, a hominid species would have met the preconditions for the evolution of reciprocal altruism; for example, long life span, low dispersal rate, life in small, mutually dependent and stable social groups, and a long period of parental care leading to extensive contacts with close relatives over many years" (p. 386). In the list of 15 unique hominid characteristics derived by Tooby and Devore (1987), many are based in reciprocity. According to Gouldner (1960): "A norm of reciprocity is, I suspect, no less universal and important ... than the incest taboo" (p. 178). When people say things such as "you scratch my back and I'll scratch yours;" "quid pro quo;" and "don't get mad, get even" they are promoting Tit-for-Tat strategies.

Accounting for the ontogenetic emergence of morality, Piaget (1932) suggested that when young children who possess deferential moral orientations grow older and interact increasingly frequently with peers in contexts in which there are no adults to tell them what is right and wrong, they figure out themselves how to coordinate their social relations in functional ways. Aided by the growth of their ability to understand reciprocity, egalitarian peer relations usher in a new moral orientation, which Piaget characterized as "the morality of cooperation" based in "mutual respect."

The Morality of Concrete Reciprocity

Tit-for-Tat forms of reciprocity are prescribed by some codes of ethics, such as those contained in the Old Testament. However, few philosophers of ethics or lay people consider the negative form of concrete reciprocity--an eye for an eye--very moral (Newitt & Krebs, 2003). Moral judgments that prescribe Tit-for-Tat forms of reciprocity such as, [you should help people] "because you may need them to do something for you one day" and "you should get even with people who rip you off" are classified as Stage 2 in Kohlberg's system.

The Adaptive Limitations of Concrete Reciprocity

The success of Tit-for-Tat strategies in Axelrod and Hamilton's (1981) computer contests notwithstanding, Tit-for-Tat strategies are limited in three respects. First, they are not equipped to invade a population of selfish strategies unless they invade in clusters that enable them to interact predominantly with replicas of themselves. This raises the question, how could such clusters have originated in the first place, especially if we assume an original state of unconditional selfishness? Second, Tit-for-Tat strategies do not become evolutionarily stable, because they open the door for more unconditionally cooperative and altruistic strategies, which in turn open the door for more selfish strategies. Finally, one selfish defection in an exchange between two Tit-for-Tat strategists locks them into a mutually recriminating and self-defeating series of selfish exchanges—a "blood feud."

STAGE 2/3 MORALITY: THE EVOLUTION OF KINDER, GENTLER, MORE FORGIVING AND CONTRITE FORMS OF DIRECT RECIPROCITY

Following the publication of Axelrod and Hamilton's (1981) findings, investigators conducted computer contests in which they changed the ground rules of the games (Dugatkin, 1997, p. 24), which opened the door for more moral strategies. Consider first the recognition that well-meaning people sometimes make mistakes.

Consider two Tit-for-Tat strategists interacting in a mutually-beneficial way. One makes a mistake and behaves selfishly, which gives rise to a blood feud. Clearly, it is in the interest of both players to reestablish the string of mutually-beneficial cooperative exchanges, which can be accomplished either by the selfish player making up for his or her mistake or the victim giving the selfish player a

second chance. Evolutionary games that followed the publication of Axelrod and Hamilton's (1981) findings found that strategies programmed in such ways could defeat Tit-for-Tat (see Ridley, 1996, for a review of relevant research). The willingness to give potential exchange partners a second chance is implied in sayings such as "everyone makes mistakes," "forgive and forget," and "forgive those who transgress against us, for they know not what they do." In Kohlberg's classification, moral judgments prescribing such strategies are classified as Stage 2/3. Trivers (1971) and others have suggested that the function of emotions such as guilt, contrition and mercy is to repair damaged reciprocal relations.

STAGE 3 MORALITY: THE EVOLUTION OF SELECTIVE INTERACTION, FRIENDSHIP, INDIRECT RECIPROCITY AND CARE

In Axelrod and Hamilton's (1981) games, players were programmed to interact randomly with all other players. In contrast, in the real world individuals may be highly selective in their choice of partners. A strategy such as "cooperate with those who cooperate with you and shun those who treat you selfishly" is well-equipped to defeat unconditionally selfish strategies. Through it, selfish players would be relegated to interacting with other selfish players in one-offspring exchanges, or with no one at all. The costs of being shunned or ostracized are potentially devastating in species that are dependent on other members of their group for survival and reproduction. Shunned individuals are, in essence, kicked out of the game—indeed all games. The wages of selfishness is ostracism, which in many species equates to death.

Psychological mechanisms that foster mutual cooperation must be designed in ways that enable individuals to (a) distinguish between cooperators and non-cooperators, (b) maximize interactions with cooperators, and (c) minimize or avoid interactions with non-cooperators. Distinguishing between cooperators and non-cooperators is a tricky task. Individuals may base such estimates on how potential exchange partners treat them, on observations of how potential partners treat others, on what potential partners say to them, especially in the form of promises and verbal contracts (Nesse, 2001), on what potential partners say to others, and on what others say about potential partners. Nowack & Sigmund (1998) found that altruistic strategies could evolve when players were able to keep track of the number of altruistic moves made by other players and adjust the probability of interacting with them accordingly.

The Evolution of Friendship

In contrast to classic Prisoners' Dilemma games, the ultimate benefits individuals are able to obtain from social exchanges in the real world may be highly variable across partners. Because members of groups have a finite amount of time and energy to devote to cooperative exchanges, it is interest to fill their "association niches" with partners or friends who possess the potential to benefit them the most (Tooby & Cosmides, 1996).

Mutual Choice and the Paradox of Popularity

Resolving to restrict your interactions to exchanges with good guys will not do you any good unless the good guys also select you. For this reason, members of groups attempt to elevate their "association value," and make themselves "irreplaceable" (Tooby & Cosmides, 1996). Individuals association value is affected by their both their willingness and ability to help others. Nesse (2001) suggested that, endowed with language, humans induce others to believe they are willing to help by making promises, which constitute commitments to future acts.

The adaptive value of selecting good guys as exchange partners and being selected as an exchange partner may produce a pleasant paradox. Individuals can maximize their gains by sacrificing their interests for the sake of others, as long as the benefits they receive from being viewed by others as an attractive exchange partner outweigh the costs of the sacrifices they incur to make themselves attractive (Alexander, 1987). To maximize their gains, individuals should select as exchange partners those they can help at least cost. Tooby and Cosmides point out that members of groups may be able to benefit each other incidentally, as they go about their business, with little or no cost to themselves. We would expect individuals to be attentive to the extent to which the resources they have to offer complement the resources others have to offer, which boils down to compatibility.

A Friend in Need

Revisiting Axelrod and Hamilton's games again, it is notable that the costs and benefits of all exchanges were reckoned directly in terms of ultimate benefits, namely the number of offspring contributed to future generations. In contrast, most of the resources people exchange in the real world are only indirectly related to reproductive success. It could pay off biologically for an individual to do many small favors for a partner or friend in return for one big

favor—100 tits for one TAT. Tooby and Cosmides (1996) discuss a phenomenon called the "Banker's Paradox". Like customers who apply for loans from banks, individuals are least likely to receive help when they most need it, because they are least able to pay it back. Tooby and Cosmides suggest that Banker's paradoxes constituted important adaptive problems in ancestral environments and that mechanisms that induce individuals to form and uphold friendships evolved to solve them.

The Design of Psychological Mechanisms Mediating Exchanges Between Friends

Tooby and Cosmides (1996) emphasize the differences between adaptations mediating concrete reciprocity and adaptations mediating exchanges among friends. As pointed out by scholars such as Clark and Mills (1993) and Shackelford and Buss (1996), people often make significant sacrifices for their friends with no expectation of compensation. The results of several studies suggest that the mental mechanisms mediating exchanges between friends are designed in ways that induce them to underestimate their costs and overestimate their gains. For example, Janicki (2004) found that participants underestimated the value of their contributions to social exchanges with friends and overestimated the value of the contributions of their partners. In addition, participants said they were more concerned about repaying than about being repaid, and felt more upset when they failed to reciprocate than when their partners failed to reciprocate. Sprecher (2001) reported similar findings on dating couples, and Greenberg (1980) found that people are motivated to avoid becoming "indebted" to others. The payoffs from friendship are like the payoffs from stocks or life insurance; they involve investments in long-term security.

Collaborative Coordination

Tooby and Cosmides' (1996) analysis of the adaptive benefits of social exchanges between friends also applies to adaptive problems such as hunting large game, building a shelter, and defending against predators that can be solved through collaborative coordination (Hill, 2002). Such problems differ from problems stemming from resource variability and variations in need because they require the simultaneous coordination of effort from two or more individuals and the distribution of the fruits of their labor. To maximize the benefits from coordinated efforts, it is in individuals' interest to select as collaborative partners those who are motivated to solve the same kinds of problems they are motivated to solve and those who possess abilities that complement their own.

The Evolution of Indirect Reciprocity

Strategies that induce individuals to select cooperators as exchange partners can give rise to systems of indirect reciprocity. In systems of indirect reciprocity, Person A gives to Person B, who gives to Person C, who gives to Person A. "What goes around comes around." Such systems have the potential to generate more benefits than systems of direct reciprocity, because they are better equipped to maximize gains in trade; however they tend to be more susceptible to cheating. People know when someone fails to pay them back and it makes them mad, but they often don't know whether people fail to pay their debts by helping third parties, and they may not care.

To evolve, systems of indirect reciprocity must contain ways of ensuring that those who do their share gain more than those who do not. As discussed, members of groups may reward cooperators and altruists directly by selecting them as exchange partners, elevating their status, and giving them material benefits. As explained by Alexander (1987), good guys also may reap indirect benefits through the enhanced fitness of their collateral relatives and through the success of their groups. In contrast, cheaters may be punished through losses in status, rejection as partners, ostracism from the group, and negative effects on the group that filter back to the cheater and his or her relatives. It follows that members of groups practicing indirect reciprocity should be vigilant for selfishness, should gossip about the social behaviors of others, and should be concerned about their reputation (Alexander, 1987). Game theorists have demonstrated that altruistic strategies can evolve and become evolutionarily stable in systems of indirect reciprocity if they enhance individuals' reputations or "images," and if members of groups discriminate in favor of those with a good reputation (Nowak and Sigmund, 1998; Wedekind and Milinski (2000).

Impression Management

Strictly speaking, individuals do not base their decisions about social exchange on how others behave; they base them on their *beliefs* about others have and will behave. What pays off in the social world is not what you do or what you are, but what others think about you—the impressions you create, your reputation (Goffman, 1959). It is in individuals'

interest to put on displays designed to induce members of their groups to overestimate their generosity and underestimate their selfishness. In support of this idea, researchers have found that people are prone to invoke more generous principles of resource allocation in front of audiences than they are in private, especially when the audiences contain members whose opinions they value and with whom they anticipate interacting in the future (Austin, 1980).

However, the selection of strategies designed to induce others to view us as more altruistic than we actually are is constrained by at least three factors. First, such strategies tend to attract exchanges with selfish exploiters. It pays off more to be viewed as a discriminating cooperator than as a gullible giver. Second, inasmuch as it is biologically costly to be deceived and manipulated, we would expect mechanisms designed to detect deception and to guard against manipulation to evolve. Cosmides (1989) has adduced evidence that our reasoning abilities are designed in ways that render us proficient at detecting cheating in the social arena. Third, false impressions are constrained by reality. To be perceived as altruistic, one must put on displays of altruism, which inevitably entails behaving altruistically. Through the medium of language—in particular gossip—members of groups can share information about the selfishness of others, reducing the opportunity to create false impressions and exploit others with impunity.

Impression-management and deception detection and prevention mechanisms undoubtedly evolved through an arms-race type of process, with deception detection and prevention mechanisms selecting for improved impression-management mechanisms, and improved impression-management mechanisms selecting for improved deception detection and prevention mechanisms (cf. Trivers, 1985). To complicate matters, each individual is both an actor and an audience, a deceiver and a detector. Social exchanges akin to sports games. Each player make offensive moves (attempts to deceive and manipulate the other) and defensive moves (guards against being deceived and manipulated).

Deception-detection mechanisms should be calibrated in accordance with the costs and benefits of detecting deception. In general, it is more beneficial to detect deception in those whose interests conflict with ours—members of out-groups and enemies—than in those with whom we share interests (see Krebs & Denton, 1997, for a review of relevant research). Indeed, when we partake in the gains of others, it may be in our interest to support their deception and self-deception (Denton & Zarbatany, 1996).

Impression Management, Deception-detection, and Morality

Deceiving others about how good we are is not right. However, if to cultivate the appearance of goodness, people must behave in fair and generous ways, impression management may induce them to behave morally. Structures designed to detect and prevent deception may constrain people from engaging in immoral behaviors. Weak detection and prevention mechanisms–gullibility, tolerance of deviance and susceptibility to exploitation—may encourage others to behave immorality.

The Evolution Of Care

To many people altruistic, caring, and loving behaviors are more moral than deferential, cooperative and fair behaviors because they are more unselfish. Whereas behaving justly entails treating everyone—including oneself—equally or equitably, behaving altruistically entails treating others better than oneself—sacrificing one's interests for the sake of others. Mental mechanisms mediating impression-management, the formation of friendships and systems of indirect reciprocity take us some distance toward accounting for caring behaviors, but they differ in significant ways from the mental mechanisms that give rise to the kind of love and nurturance people bestow on their mates and offspring.

In Axelrod and Hamilton's (1981) models, players reproduced asexually and offspring entered new generations as self-sufficient adults. Things become considerably more complicated in species that reproduce sexually and bear offspring that need assistance after birth. In sexually-reproducing species propagating one's genes usually entails helping one's offspring. Many chapters of this Handbook are devoted to mating strategies, parental investment and kinship. In this chapter I will explain how mechanisms designed to help individuals foster their reproductive success may dispose them to engage in the types of caring and altruistic behaviors that many people consider the heart of morality.

Investing in Mates and Offspring

Propagating one's genes through sexual reproduction is an inherently cooperative enterprise. Males and females must coordinate their efforts to produce a

product in which each shares an interest. Because the complement of genes that each partner contributes is inexorably linked to the complement of the other, propagating one's own genes entails propagating the genes of one's mate. Sexual reproduction is a prime example of collaborative coordination.

It is appropriate to view mating in terms of the original social problem that I argued gave rise to morality. Men and women who want offspring share a confluence of interest. Neither can achieve this goal without the assistance of the other. But they also experience a conflict of interest. It is in the interest of each party to contribute less than his or her share and to induce the other to contribute more than his or her share to their mutual investment. As with the acquisition of more survival-oriented resources, individuals adopt strategies and engage in social games to solve this adaptive problem. Some strategies, such as rape, infidelity and cuckoldry, are selfish and immoral—they are designed to foster the interests of those that invoke them at the expense of the interests of their mates. Other strategies, such as devotion and fidelity, are unselfish and moral.

When individuals choose mates, they act as agents of selection, selecting the qualities (possessed by their mates) that will be inherited by their offspring and transmitted to future generations. Sexual selection may well have played an important role in the evolution of mental mechanisms that give rise to care-oriented behaviors (Krebs, 1998; Miller, 1998). Zahavi and Zahavi (1996) have argued that female are attracted to males who have prevailed in spite of handicaps, and that dispositions to behave altruistically may have evolved through the handicap principle. More basically, it is in the adaptive interest of members of both sexes to mate with individuals who are disposed to love and care for their partners and offspring. The greater the need for assistance, the more important these qualities become. It also is in the adaptive interest of members of both sexes to select mates who will honor their commitments to them and their offspring. In general, it is more important for men than for women to select mates who are faithful, because maternity is more certain than paternity (Buss, 1994).

Clearly, humans inherit mechanisms that induce them to fall in love with members of the opposite sex, care for their offspring, and treat their relatives in altruistic ways. However, equally clearly, people sometimes cheat on their partners and mistreat their offspring. As with other evolved mechanisms, the key to understanding why people sometimes help their relatives and sometimes hurt them is to identify the "if" conditions that activate the strategies they possess.

Sex Difference in Moral Orientation

Research supports Gilligan's (1984) claim that women tend to make more Stage 3 care-oriented moral judgments than men do about their real-life moral dilemmas, but not necessarily because women acquire care-oriented dispositions early in life, as Gilligan claims. The reason why women make more care-oriented judgments than men about their real-life moral dilemmas is because the dilemmas they report are more care-oriented in nature than the dilemmas reported by men (Wark & Krebs, 1996, 1997). If you hold the type of dilemma constant, the sex difference disappears. Interpreted in evolutionary terms, the types of adaptive problem individuals experience determine the types of strategies they invoke and types of judgments they make.

The Generalization of Caring Behaviors to Kin

In Axelrod and Hamiton's (1981) games, there was a 100% probability that "offspring" would inherit the strategies of their "parents." In contrast, among members of sexually-reproducing species, the probability of individuals sharing genes or strategies with other members of their groups varies with their degree of relatedness. In an insight that had a profound effect on our understanding of evolution and altruism, Hamilton (1964) pointed out that individuals should be disposed to help other members of their groups when the genetic cost to them of helping is less than the benefits to the recipient divided by his or her degree of relatedness. Research on humans and other animals has supported this expectation (Burnstein, Chapter XX). In effect, Hamilton's rule explicates the "if" conditions built into an evolved moral strategy.

The Generalization of Caring Behaviors to Kin-like Members of Groups

Strictly speaking, the strategy described by Hamilton induces individuals to restrict their altruism to kin. However, kin-selected mechanisms may be designed in ways that induce people to help non-relatives, because they are imprecise (Krebs, 1998). Whatever the ability of genes to identify replicas of themselves in others (Rushton, 1999 vs. Dawkins, 1989), members of many species employ cues to genetic relatedness such as phenotypic similarity, familiarity, and proximity to identify relatives (Porter, 1987). Such cues may well have been more highly

correlated with kinship in ancestral environments than they are in modern environments. The more imprecise the mechanisms of kin recognition and the more they misfire in modern environments, the greater the range of altruism to which they give rise.

Stage 3 Morality

Moral judgments that uphold relationships and prescribe care-oriented behaviors are considered virtuous in all cultures (Sober & Wilson, 1998). In Kohlberg's system, moral judgments such as (a) you should help members of your groups "in order to leave a good impression in the community," (b) you should help your friends "to show love, respect, trust, or honesty because this builds or maintains a good relationship" and "to show appreciation, gratitude, or respect for everything [they] have done for you," and (c) people should help their spouses "because they feel close to them," and "because they care about them and love them" are classified as Stage 3. However, as nice as love, care, and nurturance seem, the behaviors to which they give rise suffer a significant moral limitation. The mechanisms that govern care and commitment are designed in ways that induce people to favor their friends, spouses and offspring at the expense of other people's friends, spouses and offspring. To meet high standards of morality, love and care must be regulated by justice; people must allocate their altruism fairly (Kohlberg, 1984).

STAGE 3/4 MORALITY: THE EVOLUTION OF GROUP-UPHOLDING DISPOSITIONS

I have explained how mental mechanisms that induce individuals to defer to those who are more powerful than they are, to reciprocate with peers, to make amends, to forgive, to cooperate with cooperators, and to care for their friends, mates, offspring, kin, and kin-like members of their group could have evolved. I believe we can take two more steps up the ladder of morality: (a) to less nepotistic and discriminatory dispositions to help members of one's group, and (b) to dispositions to create and uphold social contracts and formal moral codes.

At least three evolutionary processes could have mediated the selection of mental mechanisms that dispose individuals to help members of their groups. First, inasmuch as individuals benefit from the existence of the groups of which they are a part, they have a vested interest in preserving them. Groups are like partners and coalitions: it pays for individuals to uphold them when they foster their security and other adaptive interests. Second, as Alexander (1987) has explained, systems of indirect reciprocity may give rise to a "modicum of indiscriminate altruism." Third, dispositions to help members of one's group may have evolved through group selection.

Group Selection

Sober and Wilson (1998) have advanced the most compelling case for the evolution of altruistic traits through group selection (Wilson, Ch.XX). Following Darwin (1871), Sober and Wilson (1998) have argued that it is plausible to assume that groups containing members who are genetically predisposed to behave altruistically would fare better than groups containing more selfish members, just as pairs of cooperators fare better than pairs of defectors in Prisoners' Dilemma games, and this would lead to an increase in altruistic genes in the population. However, altruism would decrease in frequency within groups. Sober and Wilson outline conditions under which between-group selection for altruism could outpace within-group selection for selfishness and suggest that such conditions may have existed in the environments in which our hominid ancestors evolved. Sober and Wilson suggest that group selection of altruism was probably augmented by the evolution of cultural norms and sanctions. Critics have taken exception to the analytic framework advanced by Sober and Wilson and have argued that the conditions necessary for group selection rarely occur in nature (e.g., see commentaries following Sober & Wilson, 2000).

The Design of Mechanisms Disposing People to Uphold In-groups

Research supports the idea that humans inherit mechanisms that induce them to identify with and favor members of in-groups, in judgment and in behavior (Linville, Fischer, & Salovey, 1989). Research on social categorization has found that simply assigning people to groups—on whatever basis—may induce such biases (Tajfel & Turner, 1985). On the other side of the coin, researchers have found that people make negative, global, and undifferentiated attributions about out-group members automatically—"They are all the same, and I don't like them."—(Hamilton, Stroessner & Driscoll, 1994). Although in-group upholding dispositions expand the range of recipients beyond relatives and family members, they are nonetheless limited morally because they are inherently ethnocentric.

STAGE 4 MORALITY: THE NATURAL SELECTION OF MORAL JUDGMENTS AND THE ORIGIN OF MORAL NORMS

In Axelrod and Hamilton's (1981) games, players did not make choices in the context of a formal moral system guided by a set of rules; indeed, players were not even able to communicate with each other. In contrast, the moral systems of all human societies are defined by sets of norms, rules and laws that members express to one another in words. Parents explain these rules to their children; teachers teach them to their students, and preachers preach them to their parishioners. To many people, the essence of morality lies in obeying these rules and regulations. How do formal systems of rules originate; why do members of groups preach them to each other, and why do people obey them?

When considering the origin of formal systems of rules, it is helpful to distinguish between behavioral norms—customs practiced by most members of groups—and verbal norms—the rules and regulations people express in words, or preach. To this point, the discussion has focused on the evolution of behavioral norms, which have evolved in many species. I turn now to a consideration of verbal norms, moral judgments and formal systems of rules and laws, which are unique to the human species.

The Evolution of Moral Judgment

Biological analyses of communication have revealed that many species are evolved to send signals designed to 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 expand the range of manipulative communication strategies available to them (MacNeilage, Chapter XX). Senders are able imaginatively to take the perspective of recipients and plan long into the future. Recipients' reactions to senders' signals is less a function of the physical properties of the signals themselves than of the ways in which they represent them cognitively.

Moral judgments can be viewed as signals designed to manipulate others. Some moral judgments, called aretaic by philosophers, label people and their behavior as good or bad. They convey approbation and disapprobation; they pass judgment. In Darwin's (1871) account of the evolution of morality, he wrote, "It is...hardly possible to exaggerate the importance during rude times of the love of praise and dread of blame" (p. 500). I suspect that the precursors to the first moral judgments in the human species were grunts and coos communicating approval and disapproval.

Denotic moral judgments prescribe or prohibit courses of action. They usually contain or imply the words "should," "ought," or "it is (or was) right or wrong to…". The moral judgments classified by Kohlberg (1984) and his colleagues are deontic in nature. The function of deontic moral judgments such as, "you should help me," and "you owe me," is to persuade recipients to behave in accordance with the prescriptions they contain. The function of more abstract deontic judgments such as "honesty is the best policy", "people should obey the law," and "do unto others…" is to induce recipients to adopt strategies that uphold social systems from which senders benefit. Viewed in this way, the reasons that define the stages of moral judgments in Kohlberg's system (Table 1) equate to persuasive arguments designed to induce recipients to behave in ways that benefit senders, directly or indirectly. Such reasons are designed to induce recipients to form cognitive representations of the "if" conditions that activate the prescriptions the reasons support.

The Selection of Verbal Moral Norms

Of all the moral judgments people could make, why do members of all known cultures tend to make those classified as Stage 1, 2, and 3 by Kohlberg and his colleagues (Colby & Kohlberg, 1987; Sober & Wilson, 1998; Wright, 1994)? What causes moral judgments to become moral norms? I believe the answer to this question is, the adaptive benefits to those who make them. Although at first glance it might seem that senders should exhort others to maximize their (the senders') gains, selfish judgments would not work, because recipients would not conform to them. In effect, recipients of moral judgments are agents of selection. They determine which kinds of judgment pay off for those who send them. In a similar vein, although we would expect recipients to be receptive to moral judgments that advance their interests at the expense of those who send them, it would not be in senders' interest to transmit such judgments. For these reasons, moral judgments that evolve into moral norms should prescribe behaviors that foster the interests of senders and recipients. They should exhort members of groups to foster their interests in ways that foster the interests of others.

The Evolution of Rules

It is a relatively short step from making deontic moral judgments buttressed by reasons to espousing more formal systems of rules and laws. Endowed with the ability to form abstract representations of reality, to deduce general principles from specific cases, and to communicate ideas to others, humans are able to identify the implicit expectations that govern the systems of cooperation that have evolved in their groups and verbalize them as rules and laws. The function of such rules is to ensure that others are clear about what is expected of them, to control the behavior of others, and to induce them to uphold the systems of cooperation from which they benefit by performing their roles and doing their duties. In extrapolating rules to the group as a whole, members bind themselves (Elser, 2000). But there is more to moral rules and laws than this, at least in the human species.

Beyond Evolved Norms and Natural Inclinations

Humans possess a unique ability to imagine possibilities that do not exist. Creative people, powerful leaders, or groups as a whole may imagine systems of cooperation that could produce greater gains than the systems that have evolved in their groups. Because different systems of cooperation guided by different rules may be adaptive in different ecological contexts, different groups may develop different customs and moral codes. Members of groups should be inclined to endorse the systems of cooperation and moral codes that contain the greatest promise of fostering their interests. People also should be more inclined to accept rules that they have had a part in creating or implementing than in those that others impose upon them, because the former are more likely to foster their interests than the latter. The further the behaviors prescribed by the new rules depart from the evolved strategies members of the groups are naturally inclined to practice, the less inclined they should be to obey them.

Reason and Social Learning

Most people assume that parents and other socializing agents teach their children to obey rules by explaining the reasons underlying them (induction), by setting good examples (modeling), and by rewarding and punishing them, either physically or through love-withdrawal (see Krebs, 2004a). There is nothing in evolutionary theory that is inconsistent with the idea that reason and social learning play important roles in the acquisition of morality. Indeed, eminent evolutionary theorists such as Boyd and Richerson (1985), Dawkins (1989), Darwin (1871) and Williams (1989) have attributed morality to reason and social learning. Mechanisms that mediate these processes evolved because they enabled our ancestors to adapt to their social and physical environments. In refinements of Axelrod and Hamilton's (1981) games, researchers found that strategies such as "Pavlov" that incorporated principles of learning were able to defeat less flexible strategies (see Ridley, 1996). Even strategies such as Tit-for-Tat can be defined in terms of principles of operant conditioning: "if a behavior is followed by punishment, change it; if a behavior is followed by reward, repeat it."

Limitations of Reason and Social Learning in the Inculcation of Morality

This said, I believe that the roles played by reason and social learning in the inculcation of morality are overrated. It is true that with the power to reason, people can create systems of rules that, if everyone abided by them, would maximize everyone's gains. It also is true that people tend to copy the behavior of others and conform to social norms. However, reason and social learning can induce people to violate rules and to behave immorally as easily as they can induce people to uphold rules and behave morally.

In game theory terms, if the goal of social interactions is to maximize one's benefits, and if everyone else—or even most people—are inclined to cooperate, the most reasonable course of action is to cheat. Selfishness is eminently reasonable if your goal is to maximize your gains. Social cognition is plagued by a host of self-serving biases (see Bandura, 1991; Haselton, Chapter XX; Krebs & Denton, 1998). Haidt (2001) has advanced a great deal of evidence in support of the conclusion that most moral judgments stem from irrational, automatic, "intuitive" cognitive and affective processes and that the primary role of moral reasoning is to generate post hoc justifications for self-interested acts. In our research on real-life moral judgment and behavior (Krebs et al., 2002), we concluded that affective reactions exert a much greater effect on moral decision-making than cognitive-developmental theorists such as Kohlberg (1984) assume.

Although social learning and conformity undoubtedly play an important role in the maintenance, spread, and transmission of moral norms (Boyd & Richerson, 1985), they are not equipped to account for the origin of moral norms (Krebs & Janicki, 2004). Attempting to account for morality through modeling and

induction leads to an infinite regress. At some point in our evolutionary history, someone had to engage in a moral behavior or preach a moral rule in order for others to copy or obey it. People are highly selective about the behaviors they model and the rules they obey. To account for such selectivity, we need to understand how the mechanisms that mediate social learning were designed in ancestral environments. Boyd and Richerson (1985) have suggested that social learning mechanisms are affected by three types of bias. Direct biases incline people to evaluate (consciously or unconsciously) the behaviors that others emit and copy those that they anticipate will best enable them to achieve their goals. Indirect biases incline people to copy the words and deeds of successful people. Frequency-dependent biases incline people to model the behaviors that are most frequent in the population (see also Flinn and Alexander, 1982). Research on social learning theory is consistent with these expectations (Burton & Kunce, 1995, pp. 151-152).

The Significance of Self-interest and Sanctions

If reason, social learning and evolved moral dispositions were enough to induce people to obey rules, there would be no need for sanctions, but this is not the case. In tandem with inventing systems of cooperation that maximize their gains, members of societies must structure their environments in ways that ensure that cooperative strategies pay off better than selfish strategies. For this reason, the moral rules and laws of all societies are supported by rewards and punishments. Such sanctions may be physical (getting whipped or stoned to death), material (fines, retributions), social (disapproval, ostracism) or psychological (shame, guilt). In Kohlberg's hierarchy, systems of cooperation upheld by Stage 4 moral judgments are supported by the kinds of physical and material punishments prescribed by Stage 1 moral judgments.

Members of groups can be induced to obey almost any system of rules and laws as long as the regulations are supported by effective sanctions (Boyd & Richerson, 1985; Janicki & Krebs, 1998; Krebs & Janicki, 2003; Sober & Wilson, 1998). In effect, members of groups induce one another—and therefore themselves—to obey moral rules and conform to moral norms by structuring and engineering their environments in ways that make obedience and conformity pay off better than disobedience and non-conformity.

But there's a catch. Although detecting and punishing those who cheat you personally may pay off better than ignoring them, the costs of taking it upon yourself to catch and punish free-riders who fail to contribute their share to society usually outweigh the gains. Better to let someone else do the dirty work. In a study entitled, "Punishment allows the evolution of cooperation (or anything else) in sizable groups," Boyd and Richerson (1992) found that this problem could be overcome by punishing members of groups who failed to punish free-riders. Price, Cosmides, and Tooby (2002) adduced experimental evidence that two motivational systems have evolved to overcome the free-rider problem. One disposes people to punish free-riders and the other disposes people to recruit cooperators by rewarding cooperation. Gintis, Bowles, Boyd, and Fehr (2003) explained how a mechanism that disposed individuals to cooperate and to punish those who failed to cooperate—called "strong reciprocity"—could have invaded a population saturated by selfish individuals and given rise to evolutionarily stable altruistic norms. To induce people to obey the rules of complex societies, members create institutions such as police forces, courts and jails designed to catch and punish cheaters.

THE EVOLUTION OF CONSCIENCE

Obeying rules in order to obtain rewards and avoid punishments doesn't seem very moral. Moral judgments explicitly prescribing this strategy are classified at the lowest stage in Kohlberg's sequence. Conformity also doesn't seem very moral in and of itself. Most people believe that to qualify as moral, a behavior must spring from an internal source. Morality involves obeying internalized rules and abiding by internalized principles when no one is watching. Exemplars of morality such as Ghandi, Martin Luther King Jr., and Christ suffered great costs to uphold their beliefs. To fully account for morality, we need to account for the development of mental mechanisms that induce people to resist the temptation to advance their interests at the expense of others when they could get away with it.

Psychologists have advanced two main explanations for the intrinsic motivation to behave morally. The first is based in learning theory. In essentially the same way that pets can be trained to resist temptation and to obey rules when no one is there to reward or punish them, people can be trained to develop moral habits through conditioning and vicarious learning (Aronfreed, 1968). The second explanation is based in identification with others and the development of perspective-taking abilities. Scholars from a wide array of theoretical traditions (e.g., Aronfreed, 1968; Freud, 1926; Higgins, 1987; Kohlberg, 1984; Mead,

1932) have advanced the idea that the mental mechanisms that give rise to moral judgments and moral behaviors—often called conscience or superego—contain cognitive representations of others. In effect, people internalize mental representations of others who direct their behavior and hold them accountable, even thought the others are not physically present.

Selman (1980) has adduced evidence that as people develop, they acquire increasingly sophisticated perspective-taking abilities. The more sophisticated such abilities, they larger the range of perspectives considered and the more abstract the cognitive representations of others' points of view. Joining many philosophers of ethics, Kohlberg (1984) has argued that sophisticated perspective-taking abilities are necessary for sophisticated moral decision-making. It is easy to see how low-stage perspective-taking that enables individuals to predict the behavior of others could be adaptive, but it is more difficult to see the adaptive value of the kinds of impartial perspective-taking abilities that philosophers believe are necessary for sophisticated moral decision-making.

HOW MORAL ARE WE, BY NATURE?

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. It takes little thought to see that even though such unconditional strategies could maximize the benefits for everyone if everyone practiced them, they could not evolve without help from sanctions, because they are vulnerable to cheating, nepotism, and discrimination against out-groups.

Although highly generalized forms of impartial perspective-taking are a theoretical possibility, there is little evidence people actually engage in them in their everyday lives (Krebs, 2000a,b,c; 2004b). Kohlberg's highest stages of moral development are different from his earlier stages. They are much "colder," more logical and reasonable; there is virtually no mention of affect in Stage 5 or Stage 6 moral judgments. Some of Kohlberg's collaborators (e.g., Gibbs, Basinger, & Fuller, 1992) have argued that the moral judgments that define Kohlberg's principled stages stem from "metatheoretical" forms of reasoning, quite different from the forms of reasoning that give rise to lower-stage moral judgments. There is no evidence that people from non-industrialized societies make Stage 5 or Stage 6 moral judgments, and it is difficult to imagine how mechanisms that induce people to behave in accordance with them could have evolved in the environments of our ancestors.

This is tragically ironic. If we all behaved in accordance with high-stage moral principles 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," we would all come out ahead. Everyone would cooperate. We wouldn't have to worry about war or crime. We could invest all the money we saved from the arms race, police, and jails in enhancing the quality of our lives. However, because the strategies prescribed by lofty principles of ethics contain no antidotes to cheating and nepotism, they are destined to fail. To create moral societies, we must make it in people's adaptive interest to cooperate with others, and the only way to accomplish this is to design environments in ways that ensure that cooperation pays off better than selfishness, cheating, free-riding and favoritism.

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.

Austin, W. (1980). Friendship and fairness: Effects of type of relationship and task performance on choice of distribution rules. *Personality and Social Psychology Bulletin, 6*, 402-408.

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

Barinaga, M. (1996). Social status sculpts activity of crayfish neurons. *Science, 271*, 290-291.

Bandura, A. (1989). Social cognitive theory. *Annals of Child Development, 6*, 1-60.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development. Vol. 1* (pp. 54-104). Hillsdale NJ: Erlbaum.

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.

Boyd, R. & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13,* 171-195.

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. M. (1994). *The evolution of desire: Strategies of human mating.* New York: Basic Books.

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

Campbell, D. T. (1978). On the genetics of altruism and the counterhedonic components in human nature. In L. Wispe (Ed.), *Altruism, sympathy, and helping: Psychological and sociological implications* (pp. 39-58).

Clark, M. S., & Mills, J. (1993). The difference between communal and exchange relationships: What it is and is not. *Personality and Social Psychology Bulletin, 19,* 684-691.

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.

Crespi, B. J. (2000). The evolution of maladaptation. *Heredity, 84,* 623-629.

Cummins,D.D. (1998). Social norms and other minds: the evolutionary roots of higher cognition. In D. D. Cummins & C Allen (Eds.), *The evolution of mind* (pp. 30-50). New York: Oxford University Press.

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.

Darwin, C. (1871). *The descent of man and selection in relation to sex.* 2 vols. NY: ., Appleton.

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

Denton, K.& Zarbatany, L. (1996). Age differences in support processes in conversations between friends. *Child Development,* 67, 1360-1373.

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

Dugatkin, L. A., & Reeve, H. K. (1994). Behavioral ecology and levels of selection: Dissolving the group selection controversy. *Advances in the Study of Behavior, 23,* 101-133).

Elster, (2000). *Ulysses unbound.* London, Cambridge University Press.

Flinn, M. V., & Alexander, R. D. (1982). Culture theory: The developing synthesis from biology. *Human Ecology, 10,* 383-400.

Freud, S. (1925). *Collected papers.* London: Hogarth Press.

Gibbs, J., Basinger, K. S. & Fuller, D. (1992). *Moral maturity: Measuring the development of sociomoral reasoning.* Hillsdale NJ: Lawrence Erlbaum

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, MA: Harvard University Press.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153-172.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25,* 161-78.

Greenberg, J. (1980). A theory of indebtedness. In K. Gergen, M. S. Greenberg, & R. H. Willis (Eds.). *Social exchange: Advances in theory and research.* New York: Plenum.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

Hamilton, W. D. (1964). The evolution of social behavior. *Journal of Theoretical Biology,* 7, 1-52.

Hamilton, D. I. Stroessner, S. J., & Driscoll, D. M. (1994). Social cognition and the study of stereotyping. In P. G. Devine, D. L. Hamilton & T. M. Ostrom (Eds.). *Social cognition: Impact on social psychology* (pp. 291-321). New York: Academic Press.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review,* 94, 319-340.

Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature, 13,* 105-128.

Huxley, T. (1893). *Evolution and ethics: The second Romanes lecture*. London: Macmillan.

Janicki M. (2004). Beyond sociobiology: A kinder and gentler evolutionary view of human nature. In Crawford, C. & Salmon, C. (Eds.) *Evolutionary psychology: Public policy and personal decisions.* Erlbaum.

Janicki, M. G. & Krebs, D. L. (1997). Evolutionary

approaches to culture. In C. Crawford & D. L. Krebs (Eds.), *Handbook of evolutionary psychology: Ideas, issues, and applications* (pp. 163-208). Hillsdale, NJ: Erlbaum.

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

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. (2004a). How to make silk purses from sows' ears: Cultivating morality and constructing moral systems. In Crawford, C. & Salmon, C. (Eds.) *Evolutionary psychology: Public policy and personal decisions, (pp. 319-342)* . Erlbaum.

Krebs, D. L. (2004b). An evolutionary reconceptualization of Kohlberg's model of moral development. In R. Burgess & K. MacDonald (Eds.) *Evolutionary perspectives on human development*. CA: Sage Publications

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., Vermeulen, S. C. Carpendale, J. I. & Bush, A. (1991). The structural flexibility of moral judgment. *Journal of Personality and Social Psychology, 61*, 1012-1023.

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. & Janicki, M. (2004) 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.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57*, 165-188.

Maynard Smith, J. (1997). Commentary. In P. Gowaty (Ed.), *Feminism and evolutionary biology*. New York: Chapman & Hall.

Mead, G. H. (1934). Mind, self and society. Chicago: University of Chicago Press.

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

Miller, G. F. (1998). The history of passion: A review of sexual selection and human evolution. In C. Crawford & D. Krebs (Eds.), *Evolution and human behavior: Ideas, issues and applications (pp. 87-130).* Hilldale, NJ: Erlbaum.

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

Nesse, R. M. (Ed.) (2001). *Evolution and the capacity for commitment.* New York: Russell Sage Foundation.

Newitt, C. & Krebs, D. L. (2003). 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.

Osherow, N. (1995). Making sense of the nonsensical: An analysis of Jonestown. In E. Aronson (Ed.), *Readings about the social animal* (7th Ed.), pp. 68-86. San Francisco: W. H. Freeman.

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.

Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior, 23*, 203-231.

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.

Selman, R. L. (1980). *The growth of interpersonal understanding*. New York: Academic Press.

Shackelford, T. K. & Buss, D. M. (1996). Betrayal in mateships, friendships, and coalitions. *Personality and Social Psychology Bulletin, 22*, 1151-1164.

Simon, H. (1990). A mechanism for social selection of successful altruism. *Science, 250*, 1665-1668.

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

Sober, E. & Wilson, D. S. (2000). In L. D. Katz (Ed.), *Evolutionary Origins of Morality* (pp. 79-101). UK: Imprint Academic.

Sprecher, S. (2001). Equity and social exchange in dating couples: Associations with satisfaction, commitment, and stability. *Journal of Marriage and Family, 63*, 599-613.

Tajfel, H., & Turner, J. C. (1985). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Chicago: Nelson-Hall.

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* (pp. 183-237). Albany, NY: SUNY Press.

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.

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.

Wark, G. & Krebs, D. L. (1996). Gender and dilemma differences in real-life moral judgment. *Developmental Psychology, 32*, 220-230.

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.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science, 288*, 850-852.

Williams, G. C. (1989). A sociobiological expansion of "Evolution and Ethics", *Evolution and Ethics* (pp. 179-214). . Princeton: Princeton University Press,

Wright, R. (1994). *The moral animal.* New York: Pantheon Books.

Zahavi, A. & Zahavi, A. (1996). *The handicap principle*. New York: Oxford University Press.