

# Morality

## An Evolutionary Account

Dennis L. Krebs

Simon Fraser University, Burnaby, British Columbia, Canada

**ABSTRACT**—*Refinements in Darwin's theory of the origin of a moral sense create a framework equipped to organize and integrate contemporary theory and research on morality. Morality originated in deferential, cooperative, and altruistic "social instincts," or decision-making strategies, that enabled early humans to maximize their gains from social living and resolve their conflicts of interest in adaptive ways. Moral judgments, moral norms, and conscience originated from strategic interactions among members of groups who experienced confluences and conflicts of interest. Moral argumentation buttressed by moral reasoning is equipped to generate universal and impartial moral standards. Moral beliefs and standards are products of automatic and controlled information-processing and decision-making mechanisms. To understand how people make moral decisions, we must understand how early evolved mechanisms in the old brain and recently evolved mechanisms in the new brain are activated and how they interact. Understanding what a sense of morality is for helps us understand what it is.*

All normal people possess a sense of morality. It is species specific and culturally universal—an integral aspect of human nature. Of what does this sense consist? Where did it come from? In this article, I argue that, to account for a sense of morality, we must identify the mental mechanisms that produce it and explain how they were selected and refined in the course of human evolution.

The article is divided into fifteen sections. I begin by offering a definition of morality, then outline traditional psychological models of how people acquire a sense of morality, arguing that these models leave important questions unanswered. I go on to summarize Darwin's (1874) theory of the origin of the moral sense, which I use in the remainder of the article as a framework for organizing contemporary theory and research. In the next

sections, I explore Darwin's contention that the roots of morality consist in a set of prosocial instincts (in modern terms, decision-making strategies) that have evolved in many social species; I identify the kinds of adaptive problems that prosocial dispositions evolved to solve and dismiss the claim that all evolved strategies are selfish. I go on to explain how strategies that give rise to deferential, cooperative, and altruistic behaviors and moral virtues have evolved in humans and other species. I conclude that humans and other animals are naturally disposed to behave in selfish and unselfish ways, depending on the conditions.

Following this account of the evolution of selfish and unselfish behavioral dispositions, I refine Darwin's account of the origin of a moral sense. I argue that the proximate mechanisms that give rise to prosocial behaviors produce affective experiences that contribute to a primitive sense of morality in humans and, perhaps, in other primates. I discuss the significance of strategic social interactions among members of groups, emphasizing the adaptive value of perspective-taking and the role of punishment. I refine Darwin's account of the origin of conscience, suggesting that the moral sense originated when early humans invoked moral judgments and moral reasoning to induce members of their groups to behave in prosocial ways. I close this discussion by explaining how moral judgments become selected and copied to produce both culturally relative and universal moral norms.

In the last sections of the article, I argue that several different mechanisms may contribute to the creation of moral beliefs and standards, especially mechanisms that give rise to moral intuitions, heuristics, social schemata, social learning, self-persuasion, self-judgment, and rational deliberation. If, in fact, several different mechanisms are equipped to produce moral beliefs and moral decisions, then the challenge becomes to explain how they are activated and how they interact (when more than one is activated). I close by explaining how the explanatory framework I derived from neo-Darwinian theory is equipped to subsume, organize, and revise the more proximately based models of moral development advanced by psychologists (cf. Barkow, 2006; Buss, 2004; Gintis, 2007; E.O. Wilson, 1998) and by arguing that understanding the adaptive functions that morality evolved to serve can help elucidate its nature.

Address correspondence to Dennis L. Krebs, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada V5A-1S6; e-mail: krebs@sfu.ca.

## WHAT IS A SENSE OF MORALITY?

At an intuitive level, everyone knows what a sense of morality is. After all, we all possess one. It is a mental phenomenon that consists in thoughts and feelings about rights and duties, good and bad character traits (virtues and vices), and right and wrong motives and behaviors. It contains feelings of entitlement and obligation, as well as such evaluative feelings as pride, respect, gratitude, forgiveness, guilt, shame, disgust, and righteous indignation. It contains standards of conduct; conceptions of distributive, contractual, restorative, retributive, and procedural justice; and thoughts and feelings about purity, sanctity, and sin (Colby & Kohlberg, 1987; Gilovich, Keltner, & Nisbett, 2006; Higgins, 1987; Shweder, Much, Mahapatra, & Park, 1997). Many people's conceptions of morality are embedded in religious beliefs (Atran, 2004; Boyer, 2001).<sup>1</sup>

Some aspects of a sense of morality are universal. For example, all people believe that they have the right to foster their welfare in socially acceptable ways and that they have a duty to fulfill their social obligations (Berkowitz, 1972), to reciprocate with others (Gouldner, 1960), to support their groups (Sober & Wilson, 1998), to refrain from hurting others for no reason (Turiel, 2002), and to avoid behaving in impure and sinful ways (Shweder et al., 1997). All people believe in some form of the Golden Rule (Hauser, 2006; Wright, 1994). However, people from different cultures and individuals within the same culture may define, operationalize, and prioritize these beliefs, values, and rules in different ways (Shweder et al., 1997; Vasquez, Keltner, Ebenbach, & Banaszynski, 2001).

When one attempts to define a sense of morality more precisely, however, difficult questions arise. What do the various aspects of a sense of morality have in common? How do moral obligations differ from other obligations? How do conceptions of moral rightness and goodness differ from conceptions of logical, conventional, and aesthetic rightness and goodness? What determines which character traits people consider good, which forms of conduct they consider right, and how they prioritize these values? Why are some aspects of a sense of morality universal and others culturally relative? As I will demonstrate, attending to the functions morality evolved to serve supplies a basis for answering such questions.

## ONTOGENETIC ACCOUNTS OF THE ACQUISITION OF MORALITY

Traditional psychological theories offer accounts of how a sense of morality emerges in infancy and changes as children develop. Social learning theorists account for the acquisition of morality in essentially the same way that most laypeople do—in terms of the internalization of social values (Bandura, 1991): Socializing

agents inculcate a sense of morality in children by teaching them the norms of their cultures, by rewarding them when they obey the rules, by punishing them when they disobey them, by setting good examples, and by offering justifications for the forms of conduct they prescribe and prohibit. Eminent evolutionary theorists have endorsed social learning accounts of the acquisition of morality. For example, Dawkins (1989) has asserted, "Let us try to teach generosity and altruism, because we are born selfish" (p. 3).

Cognitive-developmental theorists account for the acquisition of a sense of morality in terms of the development of children's capacity to understand or "construct" moral issues, to take the perspective of others, to coordinate social relations, and to make rational moral decisions (Kohlberg, 1984; Piaget, 1932). At first, children possess a primitive understanding of morality, just as they possess a primitive understanding of other phenomena. With cognitive development and social experience, their conceptions of morality become increasingly broad, integrated, differentiated, logical, impartial, and reversible (Kohlberg, 1984). Thus, adults' conceptions of morality are better organized and more complex than those of children.

Traditional psychological models of the acquisition of a sense of morality have considerable explanatory power (Rest, 1983). However, each model accounts for only part of the process, leaves important questions unanswered, and is inattentive to—indeed, sometimes disparaging of—the insights of other approaches (e.g., Aronfreed, 1968; Bandura, 1991; Gibbs, 2006; Kohlberg, 1984; Saltzstein & Kasachkoff, 2004). For example, social-learning accounts do not offer overriding ultimate explanations for how moral rules and norms originated, why people preach them, why children conform to some of them but not to others, and why children resist induction and engage in moral argumentation (Krebs & Janicki, 2004). Cognitive-developmental approaches have difficulty accounting for affective influences on moral judgment (Haidt, 2001), irrational and childish thinking in adults (Krebs & Denton, 2005), the use of moral reasoning to justify immoral acts (Bandura, 1991), self-serving biases in moral judgment (Krebs & Laird, 1998), situational influences on moral cognition (Krebs, Vermeulen, Carpendale, & Denton, 1991), and the links between moral reasoning, moral judgment, and moral behavior (Aronfreed, 1968; Bandura, 1991; Haidt, 2001; Krebs & Denton, 1999, 2006).

## DARWIN'S THEORY OF THE ORIGIN OF THE MORAL SENSE, IN A NUTSHELL

In contrast to psychological theories that focus on the ontogenetic origin of a sense of morality, Darwin (1874) focused on its origin in the human species. Although some of Darwin's ideas seem risibly outdated, his overriding approach and basic ideas may well constitute the most useful existing framework for understanding the complexities of morality. Whereas modern

<sup>1</sup>Relations between religion and morality are complex and controversial. I do not address them in this article. For models of the evolution of religion, see Atran (2004) and Boyer (2001).

scholars—including contemporary evolutionary biologists—focus on parts of the project, Darwin synthesized them into a coherent whole. (For a review of controversies in contemporary evolutionary accounts of human behavior and criticisms of evolutionary psychology, see Gangestad & Simpson, 2007, and Krebs, in press).

In *The Descent of Man and Selection in Relation to Sex*, Darwin (1874) theorized that a sense of morality originated from precursors possessed by other animals that became refined and expanded in the course of human evolution. Darwin (1874) outlined four stages in the evolution of morality. First, early humans inherited social instincts, which “lead an animal to take pleasure in the society of its fellows, to feel a certain amount of sympathy with them, and to perform various services for them” (p. 95). Second, humans acquired increasingly sophisticated intellectual abilities, which endowed them with a conscience:

As soon as [early humans’] mental facilities had become highly developed, images of all past actions and motives would be incessantly passing through the brain of each individual; and [a] feeling of dissatisfaction . . . would arise, as often as it was perceived that the enduring and always present social instinct had yielded to some other instinct.

(Darwin, 1874, p. 96)

Third, “after the power of language had been acquired, and the wishes of the community could be expressed, the common opinion how each member ought to act for the public good, would naturally become in a paramount degree the guide to action” (p. 96). Finally, “the social instinct, together with sympathy, [would be] greatly strengthened by habit, and so consequently would be obedience to the wishes and judgment of the community” (p. 96).

The basic idea that morality involves fostering the public good and that a sense of morality evolved from primitive social dispositions that are enriched and expanded by cognitive development, language, and social learning is as valid today as it was when Darwin advanced it. In the ensuing sections of this article, I will outline a revision of Darwin’s model based on new knowledge of the process of evolution and new developments in research on morality (see Box 1).

## ADAPTIVE SOCIAL PROBLEMS AND THE SELECTION OF SELFISHNESS

According to Darwin, the first step in the evolution of morality involved the acquisition of prosocial instincts. As explained by Darwin, animals form groups for a reason: “With those animals which were benefited by living in close association, the individuals which took the greatest pleasure in society would best escape various dangers; whilst those that cared least for their comrades, and lived solitary, would perish in greater numbers” (Darwin, 1874, p. 102). Darwin recognized that animals may obtain benefits from group living by exchanging goods and services and by coordinating their efforts to obtain food, defend

themselves against predators, remove parasites, and build shelters. However, Darwin recognized that it is difficult to explain how mental mechanisms that dispose individuals to help other members of their groups could evolve:

It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents, or of those who were the most faithful to their comrades, would be reared in greater numbers than the children of selfish and treacherous parents belonging to the same tribe. He who was ready to sacrifice his life . . . rather than betray his comrades, would often leave no offspring to inherit his noble nature. . . . Therefore it hardly seems probable, that the number of men gifted with such virtues . . . could be increased through natural selection.

(p. 127)

It is important to note that Darwin did not assert that altruistic individuals could not produce more offspring than selfish individuals. He did not define altruistic behaviors as behaviors that enhance the fitness of recipients at a cost to the fitness of donors, as contemporary biologists do.<sup>2</sup> He simply considered it improbable that altruists would fare better than those who inherited dispositions to behave in more selfish and treacherous ways.

## Fundamental Social Dilemmas

Confluences and conflicts of interest among members of groups give rise to adaptive problems that I will label *fundamental social dilemmas*. As expressed by the contemporary philosopher, Rawls (1999):

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 produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share.

(p. 4)

Although members of groups can obtain “a better life” by cooperating with others than by living “solely by [their] own efforts,” it may be in each individual’s interest to induce others to give more and to take less than their share, while themselves giving less and taking more. Such dilemmas may assume a variety of forms, as modeled in social games such as the prisoner’s dilemma and the public goods dilemma.

Fundamental social dilemmas are exacerbated by natural selection. If those who are naturally disposed to behave selfishly contribute more than their share of offspring to future generations, the proportion of selfish members in a group will increase. As the number of selfish members increases, the benefits of

<sup>2</sup>“Altruism refers to interactions where the actor incurs a fitness cost while benefiting the recipient” (Kurland & Gaulin, 2005, p. 448).

## BOX 1

### *Phases in the Evolution of Morality*

#### **SOCIALITY AND THE FORMATION OF GROUPS**

Aggregation and affiliation

#### **ADAPTIVE SOCIAL PROBLEMS: SOCIAL DILEMMAS**

Confluences and conflicts of interest

#### **THE SELECTION OF PROSOCIAL BEHAVIORAL STRATEGIES AND SOCIAL EMOTIONS**

- Deference
  - Fear, awe, respect
- Cooperation
  - Gratitude, anger, indignation
  - Forgiveness, mercy
- Altruism
  - Love, sympathy, empathy

#### **STRATEGIC INTERACTIONS**

Manipulation, the activation of prosocial dispositions by members of groups, button-pushing, offensive and defensive tactics, social sanctions, perspective-taking, arms races

#### **THE ORIGIN OF CONSCIENCE**

Anticipatory fear of punishment, regret, remorse, perspective-taking

#### **THE ORIGIN OF MORAL JUDGMENT AND MORAL REASONING**

- The evolution of symbolic language and intelligence
- Moral judgment and moral reasoning as tactics of social influence in strategic interactions
  - Aretaic moral judgments: Approval and disapproval
  - Deontic moral judgments: Persuasion
  - Moral argumentation
  - Impression management
  - Self-justification

#### **THE SELECTION OF MORAL JUDGMENTS AND THE EVOLUTION OF MORAL NORMS**

- Moral judgments and forms of moral reasoning that produce adaptive benefits by influencing the behavior of others increase in frequency
- The co-evolution of genes and culture

#### **THE ORIGIN OF MORAL BELIEFS AND STANDARDS, AND THE REFINEMENT OF CONSCIENCE**

Sources of moral beliefs:

- Moral intuitions
- Moral heuristics
- Social schemata
- Social learning, moral habits
- Self-persuasion
- Self-reflective perspective-taking
- Rational deliberation and creative thought

selfishness will decrease, because fewer and fewer members will be disposed to behave cooperatively and, therefore, there will be progressively fewer cooperators to exploit. In the end, if everyone behaved selfishly, the cooperative system would collapse, to everyone's detriment (Brown, 1986; Hardin, 1968). The key to understanding how moral dispositions evolved is to understand how animals resolve fundamental social dilemmas.

Because Darwin did not realize that genes are the primary unit of inheritance, he believed that the process of natural selection was driven by the survival of the fittest individuals and the propagation of offspring. In contrast, contemporary evolutionary theorists focus on the selection of genes. They assume that members of social species inherit genes that, in interactions with epigenetic inputs "of mind-boggling complexity" (Barkow, 2006, p. 4), design mental mechanisms that dispose animals

to make decisions that affect the probability that the genes will be propagated (Alcock, in press; Dawkins, 1989). Animals that inherit genes containing recipes for the development of mechanisms that enable them to fare well in competitions with animals that possess alleles of these genes contribute more replicas of their genes to ensuing generations than their competitors do.<sup>3</sup>

Evolutionary psychologists seek to decipher the design of the mechanisms that have evolved through natural selection, which entails explicating their implicit "grammar," or decision-making rules. Common misconceptions notwithstanding (see Lickliter & Honeycutt, 2003, and commentaries), evolutionary

<sup>3</sup>Single genes have won such evolutionary contests in approximately two thirds of the loci in the human genome. Aspects of selection such as environmental variation, heterozygote superiority, frequency-dependent selection, and mutation may give rise to more than one allele at particular loci.



psychologists assume that most decision-making mechanisms, especially in the social domain, contain flexible, facultative, “if-then” decision rules or strategies (Buss, 2004), which dispose animals to respond differently in different social contexts (Barkow, 2006; Gangestad & Simpson, 2007; Gaulin & McBurney, 2004).<sup>4</sup>

### Are All Evolved Dispositions Selfish?

Recasting Darwin’s doubts, it might seem that genes that contain recipes for the development of mechanisms disposing animals to behave in selfish ways would fare better in evolutionary contests than genes that help produce mechanisms disposing animals to behave in cooperative and altruistic ways, rendering all species selfish by nature. Some very eminent evolutionary theorists appear to have reached this conclusion. Consider, for example, the following statement:

I think “nature red in tooth and claw” sums up our modern understanding of natural selection admirably. . . . If you look at the way natural selection works, it seems to follow that anything that has evolved by natural selection should be selfish.

(Dawkins, 1989, pp. 2–4)

There are, however, three problems with this deduction. First, the kind of selfishness favored by natural selection is different from the kind of selfishness that people view as antithetical to morality (Alexander, 1987; Barkow, 2006; Gintis, 2007; Krebs, 2006; Krebs, in press; Nesse, 2001). Evolution favors the selection of genetically selfish traits, defined as traits that help animals propagate replicas of their genes, but when people (including Darwin) characterize traits as selfish and immoral, they are referring to psychologically selfish traits, defined as traits that dispose people to advance their physical, material, and psychological welfare at the expense of others (Batson, 2000).<sup>5</sup> “That genes are selfish simply does not imply that self-interest is at the core of the human psyche; it is not” (Gintis, 2001, p. xvii). Second, it is implausible that a rigidly unconditional social strategy such as “always advance your interests at the expense of others” (modeled as “All D” in the prisoner’s dilemma) would fare better than more flexible, facultative, and conditional social strategies. Finally, dispositions that were genetically selfish in the environments in which they were selected may give rise to genetically unselfish behaviors in other environments (Crespi, 2000).

<sup>4</sup>When the environments of animals contain variations that affect their fitness during their lifetimes, natural selection favors mechanisms equipped to respond adaptively to the variations. Because social problems are highly variable and because behavioral flexibility is relatively easy to design in comparison with morphological flexibility, we would expect most social dispositions to be facultative.

<sup>5</sup>For example, *selfishness* is defined in the Oxford dictionary as, “deficient in consideration for others; concerned chiefly with one’s own personal profit or pleasure.”

As Darwin implicitly recognized, principles of natural selection do not close the door on the possibility that animals that are disposed to foster the welfare of others at their personal expense can contribute more replicas of their genes to future generations than could animals that are disposed to behave in more selfish ways. Psychologically and biologically unselfish strategies have evolved in many species, including our own, as I will now explain.

## THE EVOLUTION OF PROSOCIAL BEHAVIORS

In his discussion of morality, Darwin alluded to three types of unselfish behavior—obedience to authority, cooperation, and altruism.<sup>6</sup> Evolutionary theorists have made major strides in explaining how social strategies that give rise to these types of behavior can evolve. In this section, I will review contemporary accounts of the evolution of these strategies and argue that they constitute the roots of morality.

### The Evolution of Deference

Consider two members of a group facing a prototypic adaptive problem—each wants the same resource. Both animals have three basic choices—to try to take more than their share, to share the resource, or to take less than their share. Clearly, the net benefits of each choice depend on the choice that the other animal makes. If both animals decide to share, they each obtain part of the resource at minimal cost. If one animal decides to take less than his or her share, the other is best off taking more. However, if both animals try to take more than their share, they may end up in a costly fight.

Resolving conflicts by invoking selfish strategies that precipitate fights is virtually always a second-choice strategy in the animal kingdom. It is usually more adaptive for animals to resolve their conflicts by invoking conditional strategies such as the following: “If my opponent is more powerful or of higher status than I am, let him or her have the resource in question; if I am more powerful or of higher status, take it for myself” (Cummins, 2005; Duntley, 2005; Lorenz, 1966). Such strategies pay off for relatively dominant members of groups by enabling them to avoid the costs of coercing others, and they pay off for subordinate members by enabling them to avoid the costs of fighting a losing battle (Buss, 2004; Hartung, 1987). In addition, it may be in the long-term interest of both dominant and subordinate members of groups to foster each another’s adaptive interests because they need one another to survive and reproduce.

Dominance hierarchies produced by dominance-deference strategies have been observed in species as diverse as hens, crickets (Dawkins, 1989), crayfish (Barinaga, 1996), and pri-

<sup>6</sup>Darwin also discussed what he called *self-regarding behaviors*, such as temperance and chastity, but he considered them to be “lower” forms of morality than behaviors that stemmed from prosocial instincts.

mates, (Boehm, 2000; Cheney & Seyfarth, 1990; de Waal, 1982). In chimpanzees, “lower-ranked individuals defer, obey, show loyalty and respect, and yield preference;” however, “low-ranking individuals are not simply dominated or exploited; they typically benefit from protection, advice, leadership, and intervention in disputes” (Haslam, 1997, p. 300). Haslam also noted, “There appear to be clear continuities between humans and nonhuman primates regarding the realization and representation of dominance” (Haslam, 1997, p. 304).

### The Evolution of Cooperation

There is a great deal of evidence that members of many social species are naturally disposed to resolve fundamental social dilemmas in cooperative ways. Investigators have studied five main types of cooperative behavior: mutualism, concrete reciprocity, cooperating with cooperators, indirect reciprocity, and long-term social investments. Let us consider how strategies that give rise to each of these forms of cooperation evolved.

#### Mutualism

Consider two or more animals pursuing a goal that they are unable to achieve on their own or a goal that they can achieve more effectively by working in concert. Prevalent examples are mating, caring for offspring, grooming, killing large game, defending one’s group, and defending territory. The animals may choose to do less than their share, their share, or more than their share, and again the adaptive consequences of these choices depend on the choices made by other animals. If others do their share, it may be in each animal’s interest to do less than his or her share; however, if too many animals make this selfish choice, the group may fail to accomplish the task, to the detriment of all.

It is well-established that many species solve this type of fundamental social dilemma by working together to obtain the resource in question (see Sachs, Mueller, Wilcox, & Bull, 2004, for a review of the literature). The prevalence and significance of mutualistic strategies may be underestimated in the literature on cooperation (Nesse, 2001). The magnitude of mutualism in the human species is immense. It takes thousands of people coordinating their efforts to create a city or to build a dam. Even if the benefits that animals bestow on their partners are incidental byproducts of efforts aimed at enhancing their own fitness (Dugatkin, 1997), they nonetheless contribute to the welfare of their partners.

#### Concrete Reciprocity

Now consider two animals that possess goods or services that are of more value to each other than they are to them. Clearly, the animals could advance their interests by exchanging the resources and reaping the gains in trade. Delayed exchanges may be especially profitable, because animals can give goods and services of relatively little value to them but of great value to recipients in return for credit that they can cash when they are

in need. The problem is, delayed exchanges are more vulnerable to cheating than simultaneous exchanges are. If one animal gives, the other may gain more by, in effect, taking the money and running—another variation of the fundamental social dilemma.

#### Tit for Tat

To evolve, social exchange strategies must contain antidotes to exploitation (Trivers, 1971). Consider, for example, the conditionally cooperative strategy, “help others, then do onto them as they do onto you.” Although animals invoking this tit-for-tat strategy lose to those that invoke selfish strategies on the first exchange, they quickly cut their losses. In addition, they may gain compensatory benefits from ongoing cooperative exchanges with cooperators that are unavailable to selfish animals that get cut out of the cooperative loop.

In an award-winning study, Axelrod and Hamilton (1981) solicited a variety of strategies from game-theory experts and played them off against each other in computerized round-robin simulations of natural selection. They found that tit-for-tat strategies could defeat more selfish strategies as long as the original population contained a sufficient number of cooperative strategies. Although empirical evidence suggests that tit-for-tat strategies have evolved in a relatively small number of species (Dugatkin, 1997; Sachs et al., 2004), there is good reason to believe that humans and chimpanzees are among them (Flack & de Waal, 2000; Leakey & Lewin, 1977). The salience of tit for tat in humans is implicit in pronouncements such as “you scratch my back and I’ll scratch yours”, “quid pro quo”, and “don’t get mad, get even.”

Following the publication of Axelrod and Hamilton’s (1981) findings, game theorists changed the parameters of the prisoner’s dilemma game to render it truer to the process of evolution (Dugatkin, 1997) and pitted social strategies against one another in other games (Axelrod, 1988, Gintis, 2000). Although this research revealed that it is more difficult for cooperative strategies to evolve in large groups than in small groups (Boyd & Richerson, 1992), it also demonstrated that changes in the parameters enabled the evolution of strategies that were “nicer” than tit for tat. One selfish lapse or error between cooperating tit-for-tat players precipitates a self-defeating blood feud. Game theorists found that strategies such as “tit for two tats,” “generous tit for tat,” “contrite tit for tat,” “Pavlov” (win-stay, lose-switch), and “firm but fair” that were equipped to break self-defeating iterations of selfishness, avoid exploitation by unrepentantly selfish strategies, and exploit unconditional do-gooders paid off better than strict tit-for-tat strategies in conducive conditions (Dawkins, 1989; Ridley, 1996). The long-term benefits of forgiving a transgression or two and making it up to someone you treated selfishly may outweigh the short-term costs. As suggested by such sayings as “everyone makes mistakes,” and “forgive and forget,” humans may practice (or at least advocate) kinder and gentler forms of tit for tat in some circumstances. Precursors of these strategies have been

observed in forgiving and reconciliation behaviors in primates (Flack & de Waal, 2000).

### *Cooperating With Cooperators and Indirect Reciprocity*

In Axelrod and Hamilton's (1981) simulations of evolution, each player interacted with other players on an equally probable basis, but this rarely occurs in the real world. It is in the best interest of animals to be selective about those with whom they interact and to adopt such conditional strategies as "cooperate with cooperators" (Skyrms, 2000). To invoke these strategies, animals must distinguish between good guys and bad guys and promote exchanges with the former while avoiding exchanges with the latter. Animals may base such decisions on how potential exchange partners treat them or on their observations of how potential exchange partners treat others. Studies have found that humans are very good at identifying those who are disposed to cooperate, even after very brief encounters (Frank, 1988).

An important byproduct of selective interaction strategies is that they enhance the value of cooperating—or at least of being viewed as a cooperator.<sup>7</sup> The short-term costs of helping are outweighed by the long-term benefits of a favorable social image or reputation, which promotes beneficial social exchanges with third parties (Alexander, 1987). Game theorists have demonstrated that systems of indirect reciprocity (A helps B and C helps A), can evolve in small groups if observers accurately track the behaviors of others and interact preferentially with those who behave altruistically (Nowak & Sigmund, 1998; Wedekind & Milinski, 2000). It is, however, unclear whether the conditions necessary for the evolution of altruism via indirect reciprocity can be met in larger groups in which it is more difficult to distinguish between cooperators and cheaters (Richerson & Boyd, 2001).

The other side of cooperating with cooperators is rejecting those who fail to cooperate. If all members of a group shunned or ostracized those who behaved selfishly, selfish individuals would lose the benefits of sociality and social exchange. As Darwin (1874) noted, rejection and ostracism are potentially lethal in social species.

### *Long-Term Social Investments*

Members of groups cannot interact with everyone to the same extent. As explained by Tooby and Cosmides (1996), it is in group members' interest to fill their "association niches" with partners who are most willing and best able to foster their fitness. As a result, members of groups tend to form mutually-beneficial relationships or friendships with those who possess matching or complementary abilities and resources.

<sup>7</sup>As pointed out by Alexander (1987), the optimal strategy is to manipulate others into viewing you as more altruistic than you really are. However, as argued by other writers (e.g., Krebs, 2005b; Nesse, 2001), to be viewed as an altruist, one must display a reasonable quantity of altruism, and we would expect the mechanisms that evolved to detect altruism to be designed in ways that minimize the chances of being duped.

People rarely invoke tit-for-tat strategies in interactions with their friends (Shackelford & Buss, 1996; Tooby & Cosmides, 1996); indeed, people tend to feel offended when their friends adopt an "exchange orientation" (Clark & Mills, 1993). In accounting for the types of social strategies adopted by friends, Tooby and Cosmides (1996) alluded to a phenomenon called *the banker's paradox*—the tendency for banks to be least likely to lend people money when people need it the most. Tooby and Cosmides explained how it could be in individuals' adaptive interest to suffer the relatively small costs of helping friends over long periods of time if such acts increased the likelihood that their friends (i.e., "bankers") would be there for them when their fitness was in jeopardy. Although the investments people make in their friends may, like insurance policies, foster their long-term security, they need not be driven by instrumental proximate motives; the assistance people offer their friends could stem from genuinely altruistic motives (Nesse, 2001).

Mechanisms have evolved in many species that dispose them to form social bonds with and support those on whom their fitness is dependent, such as mates, offspring, parents, siblings, friends, exchange partners, or members of coalitions and groups (Brown & Brown, 2006). As suggested by Flack and de Waal (2000),

Inasmuch as every member [of a group] benefits from a unified, cooperative group, one expects them to care about the society they live in, and to make an effort to improve and strengthen it similar to the way the spider repairs her web, and the beaver maintains the integrity of his dam . . . In trying to improve this quality for their own purposes, they help many of their group mates at the same time.

(p. 14)

Extending this idea, Laland, Odling-Smee, & Feldman (2000) have suggested that it is in each animal's adaptive interest to support other animals that "niche construct" in ways that foster their fitness.

Brown and Brown (2006) have argued that theorists have erred in accounting for costly long-term social investments in terms of selfish motives. Rather, these theorists assert that social bonds generate the desire to give rather than to receive. In support of this assertion, investigators have found that people tend to view those with whom they have formed social bonds as extensions of themselves (Aron & Aron, 1986), people underestimate the costs of giving to their friends and overestimate the value of the benefits they receive, and people report feeling more upset when they fail to pay their friends back than they do when their friends fail to pay them back (Janicki, 2004).

### **The Evolution of Altruism**

None of the social strategies discussed so far induce animals to behave in biologically altruistic ways (i.e., in ways that increase others' fitness at a cost to their own fitness); they are all equipped



to enhance the fitness of those who adopt them. There is, however, compelling evidence that biologically altruistic strategies can evolve in at least three ways—through sexual selection, kin selection, and group selection. The principles and products of these forms of selection have been widely discussed (Gintis, Smith, & Bowles, 2001; Hamilton, 1964; Miller 1998, 2007; Sober & Wilson, 1998; Zahavi & Zahavi, 1997). I will summarize them briefly here.

### *Sexual Selection of Altruism*

In *Descent of Man*, Darwin emphasized the significance of sexual selection in human evolution (Richards, 1987). The central idea underlying the sexual selection of altruism<sup>8</sup> is that traits that are attractive to the opposite sex, and that therefore increase the probability of mating, will evolve even if they are not optimal for survival. Traits that diminish an animal's prospects of surviving (to an optimal point) but enhance its prospects of reproducing are common in the animal kingdom (Alcock, 2005). The gaudy ornamentation of peacocks and other birds are often-cited examples. Potential mates may prefer partners who possess costly altruistic traits for two reasons—because the traits signal that their bearers will treat them and their offspring altruistically and because the traits signal “good genes” that render those who possess them powerful enough to survive in spite of the costs of behaving altruistically (see Zahavi & Zahavi, 1997, for an elaboration of costly signaling theory). Notwithstanding well-documented sex differences in qualities that men and women find attractive in their mates, members of both sexes rank qualities related to altruism near the top of their lists (Buss, 2004; Miller, 2007).

### *Kin Selection of Altruism*

Darwin (1874) anticipated what has come to be called *kin selection* when he theorized that the social instincts originated in “parental and filial affections” (p. 95). The basic idea underlying kin selection is that genes may increase in frequency by guiding the development of mechanisms that dispose individuals to sacrifice their bodies (the vehicles transporting the genes) in order to preserve other bodies that share copies of their genes by descent (Dawkins, 1989). In an insight that had a profound effect on our understanding of the evolution of altruism, Hamilton (1964) demonstrated that altruistic dispositions could evolve if the fitness costs of altruistic behaviors were less than the fitness benefits they bestowed on recipients, multiplied by their degree of relatedness ( $C < rB$ ). In effect, Hamilton's rule explicates the “if” conditions built into an evolved social strategy.

There is abundant evidence that members of many species are disposed to sacrifice their personal and biological interests for the sake of their kin (Alcock, 2005; Kurland & Gaulin, 2005).

<sup>8</sup>Sexually selected behaviors tend to be altruistic in terms of their contribution to survival but selfish in terms of their contribution to reproductive success.

Ethnographic and experimental research on humans has found positive correlations between degree of relatedness and the probability of engaging in a variety of helping behaviors, especially those that are costly and life threatening to the helper (Burnstein, 2005; Kurland & Gaulin, 2005). Kruger (2003) found a significant relation between kinship and helping when empathy, self–other overlap, and expectations of reciprocity were controlled; and Segal, Hershberger, and Arad (2003) found that monozygotic twins were more prone to help one another than dizygotic twins were.

For altruism to evolve through kin selection, animals must discriminate between kin and nonkin and favor those to whom they are most closely related. The extent to which animals are able to make this distinction is of considerable theoretical significance. If animals were good at detecting degree of relatedness, their altruistic behaviors would be nepotistic, and the process of kin selection would be poorly equipped to account for altruism toward nonkin. However, if animals were poor at distinguishing between kin and nonkin, they would end up behaving altruistically toward a broader range of recipients, and they would be susceptible to manipulation by nonrelated members of their groups (Alexander, 1987; Trivers, 1985).

Humans distinguish between kin and nonkin on the basis of cues such as familiarity, proximity, and phenotypic similarity that reliably signaled kinship in ancestral environments (Burnstein, 2005). Theorists such as Johnson, Stopka, & Knights (2003) and Van den Berghe (1983) have argued that these cues induce humans to behave altruistically toward nonrelatives in modern environments. Other theorists, such as those who account for human altruism in terms of gene–culture coevolution, have argued that contemporary humans are quite good at distinguishing between kin and nonkin (Fehr & Gächter, 2002, 2003). These theorists question whether dispositions to behave altruistically toward nonkin could have evolved through kin selection. Brown and Brown (2006) have proposed that the mechanisms that regulate altruism are more sensitive to fitness interdependence than to degree of relatedness per se.

### *The Expansion of Altruism Through Group Selection*

Pessimistic about the ability of moral traits to evolve through individual selection, Darwin (1874) entertained the idea that they could evolve through group selection:<sup>9</sup>

A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most

<sup>9</sup>“The social instincts, which no doubt were acquired by man as by the lower animals for the good of the community, will from the first have given to him some wish to aid his fellows, some feeling of sympathy, and have compelled him to regard their approbation and disapprobation. Such impulses will have served him at a very early period as a rude rule of right and wrong.” (Darwin, 1874, p. 120)



other tribes [and through this process] the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase.

(p. 129)

As Darwin (1874) recognized, however, there is a problem with this explanation: “But it may be asked, how within the limits of the same tribe did a large number of members first become endowed with these social and moral qualities?” (p. 127). Translated into modern terms, although groups replete with altruists may fare better than groups replete with selfish members, we would expect selfish members to fare better than altruists within altruistic groups, producing a progressive increase in the proportion of selfish members within the groups.

As explained by Sober and Wilson (1998), several conditions must be met in order for between-group selection for altruism to outpace within-group selection for selfishness. First, the variance between groups must be relatively greater than the variance within groups (which could occur, for example if there were relatively little migration or infiltration, and if members of groups conformed to the same cultural norms). Second, the pace of between-group selection for altruism must be relatively rapid in comparison with the pace of within-group selection for selfishness (which could occur, for example, if groups engaged in frequent wars). And third, groups must bud, disperse, fission, or reform at optimal times (to disrupt the inevitable process of within-group selection for selfishness). Sober and Wilson (1998) suggested that cultural evolution has helped create the conditions necessary for group selection in the human species by mediating the rapid spread of altruistic norms within groups. Other theorists are skeptical (Dawkins, 1989; Trivers, 1985; Williams, 1966). Note that altruistic behaviors that evolve through group selection qualify as unconditionally and genetically altruistic within groups.

### The Evolution Of Moral Virtues

Sexual selection, kin selection, and group selection are equipped to guide the evolution of moral virtues other than altruism. For example, Miller (2007) has suggested that sexual selection buttresses other forms of selection for moral virtues such as kindness, sympathy, forgiveness, commitment, fidelity, trust, trustworthiness, fairness, and the willingness to suffer costs to punish “free riders” and otherwise uphold one’s group. In particular, he adduced evidence in support of the following hypotheses: (a) moral virtues are attractive to potential mates, whereas vices (such as those embodied in the Seven Deadly Sins) are unattractive; (b) moral virtues are heritable; (c) moral virtues signal good genes (i.e., genes that are correlated with physical and mental health), and vices signal bad genes; (d) moral virtues constitute costly signals that only the fittest members of groups can afford to display on an ongoing basis and are, therefore, difficult to fake; and (e) moral virtues signal

(honestly) that those who possess them are willing and able to invest in their mates and offspring.

### THE INTRUSION OF SELFISHNESS

It is important to note that although all social strategies discussed to this point may give rise to unselfish resolutions of fundamental social dilemmas in conducive “if” conditions, they also may give rise to selfish resolutions in other conditions. For example, dominant animals may exploit submissive members of their groups (Duntley, 2005), and submissive animals may cheat when they can get away with it. People may exert less effort in group tasks than they do when undertaking the same activities alone—a phenomenon called “social loafing” (Latane, Williams, & Harkins, 1979). Members of hunting and gathering societies who share large, perishable game may hide and hoard small nonperishable goods (Hill, 2002). Individuals who are disposed to help friends, family, and ingroups may be disposed to harm outsiders. The conditional design of social decision-making mechanisms has an important implication for models of human nature. It is misguided to characterize humans as either good or evil—they are naturally inclined to behave in both ways, depending on the conditions.

To summarize, contemporary theory and research supply a basis for refining Darwin’s original ideas about the evolution of “social instincts.” Modern theorists have accounted for the evolution of strategies that give rise to deferential, cooperative, and biologically altruistic behaviors in terms of the adaptive benefits of avoiding punishment, coordinating effort, exchanging goods and services, selecting fitness-enhancing mates, supporting kin and others on whom one’s fitness is dependent, and upholding groups. Researchers have adduced a plethora of evidence that such strategies have evolved in many species. Our closest phylogenetic relatives, chimpanzees, are naturally disposed to behave in deferential, cooperative, sympathetic, altruistic, reconciliatory, rule-following, rule-enforcing, and group-upholding ways (Flack & de Waal, 2000).

### THE ORIGIN OF A MORAL SENSE: MORAL EMOTIONS

Accounting for the evolution of prosocial behavioral dispositions is one thing; accounting for the evolution of the sense that considers prosocial behaviors moral is a considerably more challenging task, to which I now turn. When the evolved mechanisms that induce people to behave in prosocial ways are activated, they produce psychological states that people experience as fear, awe, righteous indignation, disgust, guilt, shame, gratitude, forgiveness, contrition, love, sympathy, empathy, loyalty, or a sense of solidarity. Neuroscientists have made considerable progress in mapping the neurological and chemical substrates of mental states associated with such “moral

emotions,”<sup>10</sup> and psychologists have made great strides in identifying the “if” conditions that activate them. Consider, briefly, some findings from psychological research on guilt, gratitude, sympathy, and solidarity.

Feelings of guilt are activated when people believe that they have violated their moral standards or have failed to live up to their moral obligations, especially with respect to the treatment of others (Baumeister, Stillwell, & Heatherton, 1994). Guilt motivates people to engage in reparative behaviors. Feelings of gratitude are evoked by the perception that a benefactor has tried to enhance one’s well-being (McCullough, Kilpatrick, Emmons, & Larson, 2001). The intensity of gratitude is affected by the value and cost of the benefit, the perceived intentions of the benefactor, and the extent to which the beneficence was expected and normative. In contrast to feelings of indebtedness, which are experienced as a negative state that motivates people to redress a perceived inequity (Greenberg, 1980), gratitude is experienced as a positive state that motivates people to help not only those who benefited them, but also third parties. Expressions of gratitude reinforce the prosocial behaviors of benefactors.

As we have seen, Darwin thought that sympathy was integral to the evolution of morality. Ahead of his time, Darwin (1874) questioned whether sympathetic behaviors are aimed at “reliev[ing] the suffering of another, in order that our own painful feelings may be at the same time relieved” (p. 103). Darwin recognized that this account fails to explain “the fact that sympathy is excited, in an immeasurably stronger degree, by a beloved, than by an indifferent person” (p. 103). Batson (1991) has adduced evidence that people may help victims not only to relieve their own vicariously experienced “personal distress,” but also out of sympathetic concern for victims, as an end in itself. Empathy is activated by perceptions of similarity (Hoffman, 2000). Empathizing with others induces people to behave in altruistic ways in contexts in which more selfish motives, such as those aimed at evoking approval and avoiding disapproval from self and others, are controlled (Batson, 1991).

Richerson and Boyd (2001) have marshaled a great deal of evidence to support the conclusion that “we are adapted to living in tribes, and the social institutions of tribes elicit strong—sometimes fanatical—commitment” (p. 215). In addition, Richerson and Boyd (2001) stated the following:

<sup>10</sup>For example, investigators have found that being in love is associated with the activation of the ventral tegmental and caudate nuclei of the brain (areas associated with experiences of pleasure), the production of neurochemicals such as dopamine, and the inhibition of neurochemicals such as serotonin (Fisher, 2004). Hormones such as oxytocin, vasopressin, prolactin, and endogenous opioids are associated with social bonding, sexual behavior (Brown & Brown, 2006), and the inhibition of brain mechanisms mediating self-preservative behaviors. In a recent review of neurological research on empathy, Decety (2005) discussed the role of mirror neurons (which fire in the identical ways when people perform acts and observe others performing the acts) and concluded that, “Shared representations [that constitute] distributed patterns of neural activation in two individuals who socially interact . . . are widely distributed in the brain. . . . cognitive processes that exert a top-down control on these shared representations are mediated by specific subregions of the prefrontal cortex . . .” (p. 153).

[Tribal instincts] evolved through a complex interaction between natural selection at an individual level and group selection at a cultural level. . . . Tribal instincts are laid on top of more ancient social instincts rooted in kin selection and reciprocal altruism. . . . We are simultaneously committed to tribes, family, and self, even though our simultaneous and conflicting commitments very often cause us . . . great anguish.

(p. 191)

Tajfel and Turner (1985) have shown that people are disposed to identify with and favor members of groups to which they have been assigned on an arbitrary basis. The intensity of emotional identification with groups is readily apparent at sporting events, where “We are Number 1,” at least when “we” are winning (Cialdini et al., 1976).

Neo-Darwinian evolutionary theory leads us to attend to the adaptive functions of moral emotions. Guilt induces people to repair fitness-enhancing social relations (Trivers, 1985). Gratitude supports systems of reciprocity (McCullough et al., 2001; Trivers, 1985). Feelings of emotional closeness, love, and empathy motivate people to help, support, and refrain from harming those on whom their fitness is dependent, or at least to help those on whom their fitness was dependent in ancestral environments (Brown & Brown, 2006). Feelings of solidarity, or “tribal instincts,” dispose people to enhance their inclusive fitness<sup>11</sup> indirectly by upholding their groups (Richerson & Boyd, 2001). Anxiety and anticipatory shame motivate people to avoid social disapproval and ostracism (Darwin, 1874). Moral emotions dispose people to uphold fitness-enhancing social commitments (Frank, 2001; Nesse, 2001).

The affective experiences that accompany prosocial behaviors produce a primitive moral sense. For example, feelings of gratitude or indebtedness instill the sense that one should return a favor, and feelings of righteous indignation instill the sense that it is wrong for people to exploit others. As explained by Krebs, Denton, and Wark (1997), there is a close conceptual association between the prudential and moral meanings of words such as “should” and “right.” For example, it is difficult to feel that one should obey authority or help a friend in need without feeling that it is right to perform these acts.

Darwin (1874) suggested that activation of social instincts engenders a rudimentary sense of duty: “The imperious word ought seems merely to imply the consciousness of the existence of a rule of conduct, however it may have originated” (p. 112). In addition, Darwin suggested that animals feel good or satisfied when they behave in ways that are consistent with their social instincts and that they feel bad when they do not. However, as Darwin recognized, people do not experience a sense of moral obligation to satisfy instincts such as those that regulate hunger and sex, and people may feel motivated to help others because

<sup>11</sup>Inclusive fitness refers to the number of genes that individuals contribute to the next generation directly (by producing offspring) and indirectly (by helping nondescendant kin).

they fear their wrath or because they feel sympathy for them without feeling that they are morally obliged to perform these acts.

In an attempt to account for the difference between moral obligations and other obligations, Darwin (1874) argued that moral obligations stem from social instincts, which differ from other instincts because they are ever present. But this explanation is, at best, incomplete. Choosing to satisfy a pressing but transient personal need that leaves a more enduring social need unsatisfied may induce people to feel that they made the wrong choice; however, it need not necessarily induce them to feel that the choice was immoral.

I believe that Darwin (1874) honed in on the factor responsible for bridging the gap between prosocial dispositions and a sense of morality when he wrote, "man would be influenced in the highest degree by the wishes, approbation, and blame of his fellow-men, as expressed by their gestures and language" (p. 106). Research has revealed that social rejection activates brain bases similar to those activated when people experience physical pain (Eisenberger, Lieberman, & Williams, 2003). de Waal (1991) accounted for the emergence of prescriptive rules in chimpanzees in terms of the "active reinforcement of others" (p. 338). The crucial difference between feelings of obligation and feelings of moral obligation is that the latter pertain to shared standards about what members of groups deserve from and owe each other that are reinforced by sanctions. Moral standards pertain to cooperative solutions to fundamental social dilemmas that affect people's welfare. Consistent with this interpretation, Trivers (1985) suggested that "a sense of fairness has evolved in the human species as the standard against which to measure the behavior of other people, so as to guard against cheating in reciprocal relationships" (p. 388).

### STRATEGIC INTERACTIONS

To understand how primitive emotional reactions shared by humans and other animals gave rise to a uniquely human moral sense, we need to attend to the dynamics of strategic interactions among members of groups. It is believed that, like other primates, early humans would have attempted to maximize their gains from group living by inducing members of their groups to behave in ways that enhanced their inclusive fitness. To accomplish this, they would have attempted to induce others to behave in deferential, cooperative, and altruistic ways. In effect, they would have attempted to activate prosocial dispositions in others (or press their prosocial buttons) by invoking tactics such as begging; administering physical and material rewards and punishments; issuing threats; and signaling their approval and disapproval through facial expressions and auditory signals that conveyed respect, gratitude, anger, disgust, and so on (cf. de Waal, 1991).

#### The Adaptive Value of Perspective-Taking Abilities

Perspective-taking or "mind-reading" abilities are invaluable tools (or weapons) in strategic interactions. They enable people

to construct cognitive representations of others; store them in their minds; view events from others' perspectives; understand what others are thinking, feeling, and planning; and imagine how others will respond to their behavior (Selman, 1980). Bloom (2004) suggested the following:

our enhanced social intelligence allows us to reason about how other people will act and react in situations that do not yet exist, so as to plan and assess the consequences of our own actions. It is adaptive to be capable of imagining hypothetical situations and of seeing these situations from another person's point of view . . . one perverse side effect of this is increased empathy.

(pp. 141–142)

Alexander (1990) has suggested that the adaptive value of the ability to "see ourselves as others see us so that we may cause competitive others to see us as we wish them to" (p. 7) played a pivotal role in the rapid expansion of the human brain.

#### Policing and Punishing Free Riders and Rule Breakers

It is easy to see the adaptive benefits of punishing those who wrong you and your relatives, and there is no question that dispositions to administer such punishments have evolved in the human species. But, as expressed by J.Q. Wilson (1993), "our sense of justice . . . involves a desire to punish wrongdoers, even when we are not the victims, and that sense is a 'spontaneous' and 'natural' sentiment" (p. 40). Many studies have found that people are willing to suffer significant costs to administer punishments to third parties who behave selfishly on group tasks (Gintis, Bowles, Boyd, & Fehr, 2003; Price, Cosmides, & Tooby, 2002).

Dispositions to punish third party cheaters are significant in the evolution of moral systems because they induce members of groups to uphold norms that promote harmony within groups, support systems of cooperation, and contribute to the welfare of the group as a whole. But how could such dispositions evolve? Revisiting the fundamental social dilemma, one finds that, although such dispositions could create a "better life for all," each member of the group could avoid the costs of punishing those who violate group norms by engaging in second-order free riding and letting others do the dirty work.

Dispositions to punish norm violators could evolve if the adaptive benefits to those who fulfilled this responsibility outweighed the costs, relative to the net adaptive benefits of less socially responsible alternatives. Those who administered sanctions could, for example, be compensated by gains in power, status, reputation, social support, and benefits to relatives. The costs of punishing cheaters could be reduced by employing sanctions that were biologically cheap to administer, such as threats, disapproval, shunning, and gossip, or by vesting responsibility in dominant members of groups, designated police, or in groups as a whole. In addition, the costs of failing to administer sanctions could be increased. As Boyd and Richerson



(1992) and Axelrod (1997) have demonstrated, virtually any behavioral norm can evolve in large groups if members punish both those who violate the norm and those who fail to accept responsibility for punishing norm violators.

Although virtually all theorists agree that humans are disposed to punish free riders at a cost to themselves, there is considerable disagreement about how these dispositions evolved. Some theorists (e.g., Johnson et al., 2003; Trivers, 2006) have argued that it is possible to account for altruistic dispositions to punish third parties through individual-level processes such as kin selection, direct reciprocity, and indirect reciprocity. Other theorists (e.g., Fehr & Gächter, 2003; Gintis, Bowles, Boyd, & Fehr, 2005) have argued that we need to invoke group selection at a cultural level to account for dispositions to punish strangers in conditions of anonymity.

### Arms Races

Strategic interactions among members of groups create arms races in which better offensive tactics create a selective pressure for better defensive tactics, which create a selective pressure for better offensive tactics, and so on (Duntley, 2005). Tactics such as impression management and deception evolved as means of avoiding negative sanctions for selfishness, which in turn created conditions conducive to the selection of mechanisms equipped to detect selfishness and deception (Alexander, 1987). The complex and intricate social strategies displayed by modern humans are products of this process.

### THE ORIGIN OF CONSCIENCE

Darwin (1874) accounted for the origin of conscience by arguing the following:

[If a person] gratifies his own desires at the expense of other men . . . past and weaker impressions [of his transgressions will be] judged by the ever-enduring social instinct, and by his deep regard for the good opinion of his fellows . . . [As a consequence], retribution will surely come. He will feel remorse repentance, regret, or shame . . . [and] resolve to act differently.

(p.110)

This account can be reframed in terms of the acquisition of perspective-taking abilities and the dynamics of strategic interactions. When children misbehave, they consider imaginatively how their parents and others would react to their transgressions (Aronfreed, 1968; Freud, 1925; Higgins, 1987). The imagined approval and disapproval of others activates emotions such as pride, fear, shame, and guilt (Aronfreed, 1968), which children experience as emanating from a mental mechanism they label conscience. Fearful children have been found to develop particularly strong consciences (Kagan, 2003), and individuals with an impaired capacity for affective re-

sponsiveness have been found to develop weak consciences (Damasio, 1994; Lykken, 1992).

### THE ORIGIN OF MORAL JUDGMENT AND MORAL REASONING

The dynamics of strategic social interaction would have changed dramatically when early humans acquired the capacity to transmit ideas symbolically through language. Williams (1989) suggested the following:

the unparalleled human capability for symbolic communication has an incidental consequence of special importance for ethics. In biological usage, communication is nearly synonymous with attempted manipulation. It is a low-cost way of getting someone else to behave in a way favorable to oneself.

(p. 211)

This view implies that moral judgments originated in the human species as tools of social influence and tactics in strategic interaction. In particular, early humans used moral judgments to induce members of their groups to believe that they should behave in the ways prescribed by the moral judgments and that they should not behave in ways prohibited by them.

Symbolic language not only enables people to express their immediate approval and disapproval with words such as *good* and *bad*, it enables them to pass judgment on events that occurred in the past; to make judgments about events that will occur in the future; to share their observations and ideas with other members of their groups; and to buttress their judgments with reasons, explanations, and justifications designed to increase their persuasive power. Language enables people to hold others accountable for how they have behaved and to influence how they will behave. It enables them to enhance or diminish others' reputations through gossip and thereby uphold systems of indirect reciprocity (Alexander, 1987; Dunbar, 1996). Because moral judgments pertain to forms of social conduct that affect people's welfare and because they are associated with positive and negative social sanctions, they tend to evoke stronger affective reactions than other kinds of judgment do (Kagan, 1984).

### Making Commitments

Symbolic language also enables people to transform primitive threats and promises into long-term commitments. Commitments enable people to influence the immediate behavior of others by inducing them to believe that they will receive future outcomes. As pointed out by Nesse (2001) and others, it may be in people's long-term interest to honor their threats and promises, even when it is costly for them to do so, because the long-term benefits of upholding a reputation for keeping one's word may outweigh the short-term costs of honoring one's commitments.



## The Functions of Moral Reasoning

Reason is widely viewed as a process for deriving the truth, and moral reasoning is widely viewed as a process for deriving fair and valid moral decisions. Evolutionary theory casts reasoning and moral reasoning in a quite different light. In *The Economy of Nature and the Evolution of Sex*, Ghiselin (1974) made the following observation:

Man's brain, like the rest of him, may be looked upon as a bundle of adaptations. . . . We are anything but a mechanism set up to perceive the truth for its own sake. Rather, we have evolved a nervous system that acts in the interests of our gonads. . . . If fools are more prolific than wise men, then to that degree folly will be favored by selection.

(p. 1)

From an evolutionary perspective, adaptation trumps truth. We would expect people to be disposed to seek the truth when it contains (or, more exactly, contained) the greatest promise to solve the adaptive problems they face, but we would also expect people to be inclined to delude themselves when illusions better serve their needs (Alexander, 1987; Krebs, Denton, & Higgins, 1988; Trivers, 2000). What we need to determine is whether moral reasoning helped early humans adapt to their environments and propagate their genes or whether it is an incidental byproduct of nonmoral reasoning. And if, as I will argue, moral reasoning was adaptive, how did it foster the inclusive fitness of early humans?

Darwin's (1874) position on the adaptiveness of moral reasoning is unclear. In some places, Darwin implied that moral reasoning enabled early humans to enhance their individual fitness (e.g., "as the reasoning powers and foresight of the members became improved, each man would soon learn from experience that if he aided his fellow-men, he would commonly receive aid in return", p. 127). In other places, Darwin implied that moral reasoning enabled early humans to enhance the fitness of their groups (e.g., reason enabled early humans to "recognize the bearing of many virtues, especially of the self-regarding virtues [i.e., temperance, chastity], on the general welfare of the tribe", pp. 115–116). And in still other places, he seemed to view moral reasoning as a byproduct of nonmoral reasoning. The contemporary evolutionary biologist George C. Williams (1989) has argued that reasoning abilities evolved to help people advance their selfish interests—for example, by enabling them to distinguish between beneficial and costly forms of social influence. However, "in its boundless stupidity [the] evolutionary process incidentally designed machinery capable of answering other sorts of questions, such as, is this message one of help or harm for what I really want in the world?" (p. 212).

Darwin was less attentive than contemporary evolutionary theorists are to the adaptive value of moral reasoning in strategic interaction and moral argumentation and to the tendency of

people to use moral reasoning is a two-edged sword. As emphasized by cognitive-developmental theorists, people may use it to create and to uphold systems of cooperation, to preserve group harmony, and to derive fair solutions to conflicts of interest. However, as emphasized by Machiavellian intelligence theorists (Byrne & Whiten, 1988), people also may use moral reasoning to advance their interests in immoral ways. Indeed, when engaged in disputes, people may disagree about whether issues qualify as being moral as opposed, for example, to being matters of social exchange or convention, as well as whether the behaviors in question are right or wrong.

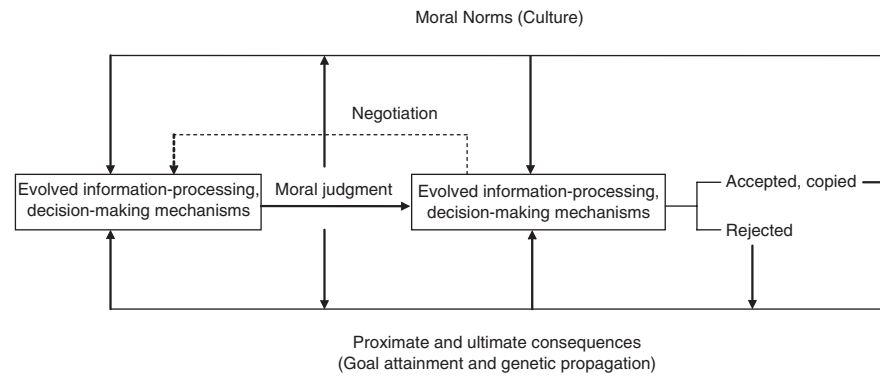
## THE SELECTION OF MORAL JUDGMENTS AND THE EVOLUTION OF MORAL NORMS

Members of groups may express all kinds of moral judgments. Of all the moral judgments that members of groups make to one another, what causes them to accept some and reject others? What causes some kinds of moral judgments to be copied and repeated sufficiently enough to constitute cultural moral norms? In large part, I believe the answer to this question lies in dynamical strategic interactions among individuals who possess evolved mechanisms designed for sending and receiving persuasive communications (see Figure 1).

As shown in Figure 1, the genetic, biological, and psychological costs and benefits of transmitting moral judgments are determined by recipients' reactions to them, which in turn are determined by the evolved information-processing and decision-making mechanisms they possess. Recipients accept and conform to some moral judgments and reject others. Senders (and recipients and observers) tend to repeat those that are accepted by others, and they refrain from repeating those that are not. Inasmuch as recipients are more receptive to some moral judgments than to others, they serve as agents of selection, determining which judgments succeed, get repeated, develop into moral norms, and contribute to culture.

Some theorists (e.g., Simon, 1990) have argued that, because it is impossible for people to learn on their own all the knowledge accumulated by members of their cultures and passed down through the generations, it is adaptive to believe the ideas expressed by others and comply with their injunctions. Think of how little we would know and how many mistakes we would make if we had to learn everything anew on our own. The benefits of invoking simple heuristics such as "believe what others say," "copy behaviors that have paid off for others," and "go along with the crowd" are obvious. According to Simon (1990), the fitness benefits of "docility" outweigh the fitness costs of being manipulated by others into behaving altruistically.

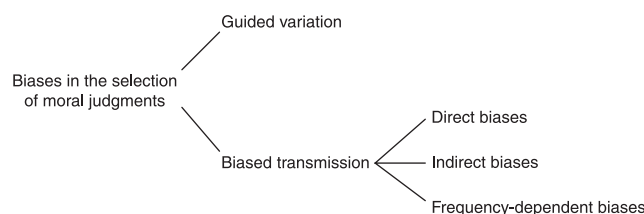
There are, however, two problems with this argument (in its extreme form). First, people are exposed to all kinds of contradictory ideas and could not possibly accept them all. Second, although evolved mechanisms may be far from perfect—they are simply the most adaptive of those available in the environments



**Fig. 1.** Relations between biological and cultural processes in the evolution of moral norms. Starting at the left, Person A sends a moral judgment to Person B that makes a small contribution to a moral norm. The judgment activates the evolved information-processing and decision-making mechanisms of Person B, who decides whether to negotiate (which creates a feedback loop), to accept and repeat the judgment, or to reject it. Copied judgments make additional contributions to moral norms. Person B's reaction to the moral judgment may have proximate consequences for Person A and B that affect their (and observers') tendency to repeat the judgment. In addition, Person B's reaction may have ultimate consequences on Person A's and B's survival and reproductive success and, therefore, the propagation of genes that helped design the evolved mechanisms in question. Note that repeating attractive moral judgments may lead to exponential increases in their frequency.

in which they were selected—individuals who possessed mechanisms that enabled them to discriminate between injunctions that enhanced and diminished their welfare would have fared better than those who did not.

The evidence suggests that although people may be gullible in some contexts, they are not evolved to accept the moral judgments of others indiscriminately. Richerson and Boyd (2005) and others have identified a set of selective “biases” in evolved social learning mechanisms. As illustrated in Figure 2, Richerson and Boyd have divided such biases into two basic categories. The first, entitled *guided variation*, is equivalent to individual learning. With respect to moral judgment, individuals repeat those that work for them, based on their personal experience. The second category, *biased transmission*, contains three types of bias. *Direct biases* induce individuals to accept moral judgments that they believe will best serve their interests. *Indirect biases* induce people to accept the moral judgments of those who have a vested interest in their welfare; those who have control over their fates; those who are powerful, successful, popular, and of high status (cf. Bandura, 1991); and those whom



**Fig. 2.** Biases in the selection of cultural variants identified by Richerson and Boyd (2005).

they like, admire, and respect (cf. Chaiken, 1987). Finally, *frequency-dependent biases* induce people to conform to the “tried and true” normative judgments that are most common in their groups (see also Flinn & Alexander, 1982).

### The Evolution of Culturally Relative Moral Norms

It is commonly assumed that evidence of cultural relativity in moral norms constitutes evidence that moral norms are products of social learning, not biological evolution. This assumption is misguided in two ways. First, it is rooted in a spurious nature-versus-nurture dichotomy. Nature and nurture are not opposing forces—they are complementary. Social learning mechanisms evolved because they helped our ancestors solve adaptive problems. They were (and continue to be) designed by genes interacting with ontogenetic experiences. Second, this assumption implies that evolved mechanisms are insensitive to variations in environments. As explained by many theorists (see Gangestad & Simpson, 2007), in the inherently unpredictable social environments in which early humans evolved, we would expect natural selection to have designed flexible, context-sensitive (“facultative”) mechanisms equipped to accommodate to, or fit with, variations in the form that adaptive problems assume in different environments:

Because evolved psychological mechanisms are functionally flexible, they are responsive to differences in local ecologies—to the unique opportunities, threats, and constraints afforded by the physical and social world around them—and so different ecologies afford superficially different cultural solutions to the same un-

derlying adaptive problems.

(Schaller, 2007, p. 369)

On this line of thought, many cross-cultural differences in the content of moral norms stem from cultural differences in the structure of adaptive problems. For example, cultures that have access to different foods may develop different dietary norms and food prohibitions (Durham, 1991), and in harsh environments such as those occupied by the Inuit, moral norms sanctioning infanticide and suicide in elders may have evolved to ensure the preservation of their younger relatives.

This is not, however, to say that all variations in cultural norms stem from variations in the form of adaptive problems (Henrich et al., 2001). Intelligent individuals from different cultures may develop different insights about how to solve the same problems, which may then be copied by members of their groups. Powerful individuals (or coalitions) may force others to conform to norms that foster their interests. Members of groups may agree to adopt arbitrary solutions to shared problems, such as agreeing to drive on the right (or left) side of the road, which work because they are adopted by all (Maynard Smith, 1982). The biases in social learning mechanisms discussed earlier also may give rise to arbitrary cultural norms. For example, people may copy the idiosyncratic ideas and practices of influential members of their groups (Richerson & Boyd, 2005). And as expressed by Darwin (1874), although “the judgment of the community will generally be guided by some rude experience of what is best in the long run for all the members; . . . this judgment will not rarely err from ignorance and weak powers of reasoning” (p. 118). Although maladaptive norms may evolve culturally, we would expect them to become extinct, due to their deleterious effects on individuals within groups, and (probably more so) because of their deleterious effects on competitions between groups.

### Dynamical Interactions

As modeled by Kenrick, Li, and Butner (2006), random aspects of the dynamic process of strategic social interaction also may play an important role in the generation of culturally relative moral norms. Consider, for example, a group that contains a random distribution of individuals with cooperative and aggressive traits. Kenrick et al. have shown that if members of such groups adopt the strategies of the majority of their neighbors, norms favoring either cooperation or aggression may emerge depending on the (random) assortment of individuals within the population. In addition, Kenrick et al. demonstrated that individual differences in thresholds for cooperation and aggression may exert profound effects on the evolution of norms. For example, a small number of uncooperative individuals may quickly corrupt cooperative groups.

### The Evolution of Universal Moral Norms

Following Aristotle, Darwin (1874) distinguished between two types of rules, akin to culturally universal and culturally relative moral norms. He suggested the following:

The higher [moral rules] are founded on the social instincts, and relate to the welfare of others. They are supported by the approbation of our fellowman and by reason. The lower rules . . . arise from public opinion, matured by experience and cultivation . . . [and may lead to] the strangest customs and superstitions, in complete opposition to the true welfare and happiness of mankind. (p. 118)

Attending to the process of strategic interaction and the adaptive value of resolving conflicts of interest in mutually beneficial ways helps refine Darwin’s analysis. Inasmuch as deferential, cooperative, and altruistic behavioral strategies evolved because they paid off genetically, we would expect people from all cultures to be receptive to moral judgments that exhort them to behave in these ways in appropriate conditions and, therefore, for all cultures to contain moral norms upholding these forms of conduct.

As illustrated in Figure 1, individuals who are equipped with symbolic language may attempt to resolve their differences by negotiating and engaging in moral argumentation. Although two self-serving individuals who have little to gain by resolving their differences may reach an impasse or end up in a fight, it is in the interest of those who have a lot to gain to resolve their conflicts in mutually acceptable ways and to be receptive to fair, balanced, and reversible solutions. When people engage in moral argumentation, they may attempt to activate each others’ moral intuitions and prosocial emotions (Haidt, 2001), or they may appeal to one another’s rational faculties (Saltzstein & Kasachkoff, 2004). As explained by the philosopher Singer (1981), rational arguments tend to generate universal and impartial standards such as the Golden Rule:

If I claim that what I do is right, while what you do is wrong, I must give some reason other than the fact that my action benefits me (or my kin, or my village) while your action benefits you (or your kin or your village).

(p. 118)

When people use reason and logical consistency as weapons in moral arguments, they often end up hoisted on their own petards.

### Gene–Culture Coevolution

Moral norms are the product of complex interactions between genetic and cultural evolution. As illustrated in Figure 1, genes guide the creation of the mental mechanisms that enable people to create, select, retain, transmit, and revise aspects of culture such as moral norms. Evolved mechanisms produce preferences for particular forms of culture. Outputs from these evolved

mental mechanisms not only contribute to culture, they also affect the probability that the genes that helped design the mechanisms (and therefore the types of mechanisms) will be propagated. For example, individuals who inherited genes that guided the construction of mechanisms that disposed them to reward and repeat fair moral judgments could (a) contribute to the creation of fair cultural norms and (b) propagate the genes that guided the construction of the mechanisms. Richerson and Boyd (2005) have emphasized the importance of cultural group selection in the evolution of moral norms.

We have come a long way toward an explanation of how a sense of morality originated and expanded in the human species, but the journey is not yet complete. People could make and comply with moral judgments and conform to moral norms without believing them, internalizing them, or incorporating them into their conceptions of morality. In addition, people could accept the validity of moral judgments without feeling any obligation to behave in accordance with them. To fully account for the acquisition of a moral sense, we must bridge the gaps between moral judgments, moral beliefs, and moral behaviors.

## SOURCES OF MORAL BELIEFS AND MORAL STANDARDS

Moral judgments are viewed by most theorists as products of moral beliefs. Dual-processing theorists such as Chaiken (1987) and Petty and Cacioppo (1986) have shown that people may derive beliefs in two ways: by processing information quickly, automatically, and mindlessly and by processing information in a more considered and controlled manner.<sup>12</sup> Evolutionary theory leads us to expect both types of decision-making mechanisms to be designed in ways that enabled early humans to solve adaptive problems and to be regulated by fitness-relevant cost-benefit considerations. It sensitizes us to the fact that the brain structures that enable people to derive beliefs and standards from reflective reasoning evolved later than, in addition to, and on top of structures that evolved earlier. Although the most recently evolved mechanisms may be equipped to modify more primitive mechanisms, they have not replaced them. Let us begin by considering the most primitive and automatic sources of moral beliefs and standards, then work progressively toward a consideration of those that are most complex and controlled.

### Moral Intuitions and Heuristics

People may derive moral beliefs from moral intuitions that stem from evolved dispositions shaped by culture (Haidt, 2001). For example, feelings of disgust evoked by physically impure phenomena may become both generalized to and embodied in feelings of moral impurity, giving rise to moral beliefs and judgments such as those that uphold dietary and sexual re-

strictions (Rozin, Haidt, & McCauley, 2000). People also may derive moral beliefs from a variety of heuristics such as those that induce them to feel that acts of commission are worse than acts of omission and that punishment should be proportionate to the outrageousness of an act (Hauser, 2006; Sunstein, 2005).

### Evolved Social Schemata

The anthropologist Alan Fiske (1992) has amassed evidence that people from all cultures are innately disposed to develop cognitive “schemata” that organize information about four types of social relations: (a) affectionate relations among people who share social bonds, (b) hierarchical relations among people who differ in social rank, (c) egalitarian exchanges among equals, and (d) economic relations aimed at maximizing cost-benefit ratios across different commodities. As documented by Cheney and Seyfarth (1990), de Waal (1996), and Haslam (1997), chimpanzees possess the first three schemata; the fourth appears to be unique to the human species. These schemata give rise to beliefs and judgments that uphold the altruistic, deferential, and reciprocating forms of social conduct discussed earlier in this article. Evidence that people possess domain-specific structures of moral reasoning designed to make decisions pertaining to helping others, to authority, and to distributive justice (Damon, 1980; Eisenberg, 1982) is consistent with Fiske’s (1992) model.

Decisions in each of the four domains vary in computational complexity (Fiske, 1992). Communal sharing decisions are based on a nominal (us vs. them) scale, authority ranking decisions are based on an ordinal scale, equality matching decisions are based on an interval scale, and market-pricing decisions are based on a ratio scale. We would expect the mental mechanisms that produce simple decisions to reside in primitive parts of the brain and to operate relatively automatically, with the mechanisms that produce more complex decisions residing in more recently evolved, uniquely human brain structures and requiring significantly more controlled processing.

### Social Learning and Self-Persuasion

Perhaps the most straightforward way in which people acquire moral beliefs is by internalizing the moral judgments made by others. As expressed by Gintis (2007), “in sharp contrast with other species, human preferences are *socially programmable*” (p. 4). In addition, the process of making moral judgments to others may induce people to form moral beliefs. A belief in the prescriptive judgments one makes may reap adaptive benefits by increasing their persuasive power (Trivers, 2000). People may persuade themselves in the process of persuading others (Festinger, 1964). People may be inclined to believe moral judgments and standards generated during moral negotiations because they actively participated in generating them, because they are supported by others, because they are backed up by reasons, and because they enable people to advance their interests in optimal ways. Several theorists have suggested that

<sup>12</sup> Automatic and controlled processes are most appropriately viewed along a continuum, but it is easier to discuss them in a dichotomous way.



moral dialogues are equipped to generate emergent preferences (Elster, 2000; Habermas, 1993).

### Self-Judgment

Researchers have found that as children's perspective-taking abilities develop, their cognitive representations of others (and consciences) become increasingly abstract, integrated, and general (Selman, 1980). As expressed by J.Q. Wilson (1993), "At first we judge others; we then begin to judge ourselves as we think others judge us; finally we judge ourselves as an impartial, disinterested third party might" (p. 33). Whether through this process or through a more autonomous mental agency, such as William James' (1890) "knower" or contemporary theorists' procedural knowledge (Linville, Fisher, & Salovey, 1989), children eventually acquire the ability to pass judgment on themselves. Discrepancies among internal moral standards and between moral standards and behaviors may evoke affective reactions such as guilt (Baumeister, et al., 1994; Higgins, 1987).

### Rational Deliberation

Humans are unique in their capacity to reflect on moral issues and derive moral beliefs and standards from moral reasoning. Like scientists, they may consider evidence systematically, evaluate it in terms of impartial standards, and deduce logical conclusions, and like judges they may consider the claims of interested parties, evaluate them in terms of principles of justice, and derive fair decisions (Kohlberg, 1984). In addition, they may anticipate "the more remote consequences of [their] actions" (Darwin, 1874, p. 129) and take measures to bring about the results they value. In contrast to the relatively direct link between activating stimuli and responses in primitive decision-making mechanisms, rational information processing involves significantly more intervening cortical mediation, interpretation, and cognitive construction. Recent studies have located the source of moral judgments in the most recently evolved and uniquely human parts of the brain (Moll, Zahn, di Oliveira-Souza, Kreuger, & Grafman, 2005).

Prosocial dispositions that evolved through kin selection and reciprocity tend to be activated by a relatively small circle of ingroup members. However, as Darwin (1874) observed, during the course of evolution, our "sympathies became more tender and widely diffused, so as to extend to men of all races . . . and [even] to the lower animals" (p. 121). Like Darwin, Singer (1981) suggested that reason was the root cause of the increase in people equipped to activate prosocial dispositions, especially as invoked in moral argumentation: "Ethical reasoning, once begun, pushes against our initially limited ethical horizons, leading us always toward a more universal point of view" (p. 119). In a similar vein, Bloom (2004) has argued that "once a creature is smart enough, impartiality—and an appreciation of moral codes such as the Golden Rule—will emerge as a consequence of this smartness" (p. 139). Reason is not, however, the

only process equipped to expand the circle of morality. For example, kin-selected altruistic dispositions may be activated by nonkin, people may identify with groups to which they have been assigned on an arbitrary basis, and small groups may form larger groups to pursue a common cause.

To summarize, the evidence suggests that people may derive moral beliefs and standards from a set of evolved mechanisms, some of which reside in the old brain and are shared by other animals and some of which reside in the new brain and are unique to humans. If this is the case, the challenge for researchers becomes to identify the factors that determine which mechanisms are activated and how they interact (if more than one is activated).

## THE ACTIVATION OF MORAL DECISION-MAKING MECHANISMS

Evolutionary accounts of moral decision making give rise to a number of expectations or predictions, most of which have been supported by research. With respect to the activation of moral decision-making mechanisms, evolutionary theory leads to the general prediction that people should be disposed to make moral decisions in the ways that best enhanced the inclusive fitness of early humans in the social environments in which they resided. Because complex forms of moral decision making are more costly than simpler forms, people should be inclined to invoke simple, automatic decision-making processes as their default (Gigerenzer, 2000; Gilovich, Griffin, & Kahneman, 2002). People should make moral judgments about recurring adaptive social problems in relatively automatic ways.

As suggested by Frank (1988), emotional reactions may be "strategically superior to those based on rational calculation" (Nesse, 2001, p. 17). Affective reactions such as sympathy, disgust, and righteous indignation should exert an immediate effect on moral decision-making processes (Frank, 1988; Haidt, 2001). People should have difficulty justifying decisions derived in these ways and, if called on to justify them, offer plausible post hoc rationalizations (Haidt, 2001).

People should be inclined to invoke simple forms of moral reasoning to solve simple, recurring moral problems (Fiske, 1992), to make quick decisions in contexts in which the costs of deliberation are high (Piliavin, Dovidio, Saertner, & Clark, 1981), and to generate simple judgments when they constitute the most effective forms of persuasion and impression management (such as, for example, when they are directed toward children; Krebs & Janicki, 2004). People should be inclined to adopt and preach the moral norms of their cultures without thinking much about them, as long as they work reasonably well, and to use mental shortcuts in contexts in which heuristics generate acceptable moral decisions (Chaiken, 1987; Gigerenzer, 2000; Kuhn, 1989, Sunstein, 2005).<sup>13</sup>

Moral reasoning should be customized to solve the different kinds of social problems faced by early humans, and therefore,

people should invoke different forms of moral reasoning in different domains, contexts, and conditions (Damon, 1980; Eisenberg, 1982; Krebs, Denton, et al., 1991). Moral reasoning should be susceptible to framing, directional, motivational, self-serving, nepotistic, and group-serving biases (Chaiken, Giner-Sorolla, & Chen, 1996; Krebs & Laird, 1998; Kunda, 2000; Lerner & Miller, 1978; Pyszczynski & Greenberg, 1987; Richerson & Boyd, 2001). And people should, in appropriate contexts, use moral reasoning for immoral purposes, such as avoiding responsibility and justifying immoral acts (Bandura, 1991; Haidt, 2001).

In contrast, people should engage in complex, controlled, and reflective forms of moral reasoning when other decision-making processes are ineffective, when the costs of deliberation are low, and when the benefits of making the right decision are high. People should be inclined to invoke sophisticated forms of moral reasoning to resolve conflicts among moral intuitions and moral norms (Haidt, 2001) and between the rights and duties of people participating in embedded systems of cooperation (Kohlberg, 1984), as well as to solve complex contemporary problems such as those that relate to the ethics of cloning and stem cell research (Pizzaro & Bloom, 2003). People should engage in reflective moral reasoning when they possess ample processing capacity, when they are challenged (e.g., in moral argumentation), when they have time to deliberate, when they are motivated to be accurate, when audiences are impressed by sophisticated moral judgments, and so on. Note that these conditions are characteristic of those in which cognitive-developmental theorists customarily assess moral reasoning.<sup>13</sup>

There is no question that people possess the capacity to derive moral decisions from moral reasoning. However, there is a big question about how often and in what contexts people invoke this tool rather than other tools in their moral decision-making tool boxes. As expected from an evolutionary analysis of morality, people rarely invoke their most sophisticated forms of moral reasoning in their everyday lives (Krebs & Denton, 2005). Indeed, Gibbs (2003), one of Kohlberg's colleagues, concluded that the impartial and reversible forms of moral reasoning that define the pinnacle of Kohlberg's hierarchy of moral development (Stages 5 and 6) are meta-cognitive and recommended dropping them from Kohlberg's sequence. Greene, Sommerville, Nystrom,

Darley, and Cohen (2001) have found that the brain areas activated when people respond to impersonal moral dilemmas, such as those on Kohlberg's test, are different from those activated when people make more personal moral decisions.

### Interactions Among Moral Decision-Making Mechanisms

Stimuli that emanate from the external world or from people's minds may activate different decision-making mechanisms simultaneously or sequentially. Jointly activated mechanisms may interact in complementary or antagonistic ways. Popular conceptions of human nature posit a conflict between evolved dispositions to behave in immoral ways and rational decisions to behave in moral ways—in Freud's (1925) terms, a conflict between the id and the ego (Campbell, 1978). However, people also may be naturally inclined to do the right thing and to invoke rational thinking to figure out ways of shirking their duties. Prosocial desires may conflict with antisocial desires (Darwin, 1874), egoistic motives may conflict with altruistic motives (Batson, 1991; Sober & Wilson, 1998), instincts may conflict with habits (Darwin, 1874), and one line of rational thought may conflict with another line of rational thought.<sup>15</sup>

As modeled by Haidt (2001), the process of moral decision making may involve complex, iterative, sequential interactions among automatic and controlled processes. On the one hand, people may impulsively intervene in emergencies, help those for whom they feel sorry, and obey powerful authorities, and rationalize their decisions afterward. On the other hand, the ways in which people interpret their primitive affective reactions may affect the kinds of emotions and intuitions they experience (Kohlberg, 1984; Pizzaro & Bloom, 2003; Schachter & Singer, 1962). Factual information, interpretive processes, perspective-taking, and moral reasoning may generate, structure, and change moral intuitions. In particular, perspective-taking may activate empathic reactions, engender altruistic motives, and expand the circle of those considered in moral decisions (Hoffman, 2000). Rational considerations may overpower intuitions. Encouraging those in intense affective states to reflect on the moral issues they face before making a decision, either in private or in interaction with others, may feed back on the form and intensity of their emotional states. Indeed, moral reasoning may even generate affective reactions and motivational states. For example, people may decide that "violations of reciprocity or justice, like violations of logic, 'shouldn't be' [and] the inference of unfairness [may] generate a motivation to restore the 'necessary' reciprocity or equality" (Gibbs, 2003, p. 36).

<sup>13</sup>In a world where information is costly and individual learning therefore error prone, imitating common behaviors is a simple heuristic that will get you the correct behavior most of the time at a smaller cost than subjecting alternative behaviors to more rigorous tests" (Richerson & Boyd, 2001, p. 193).

<sup>14</sup>Considered from this perspective, people's tendency to invoke sophisticated forms of moral reasoning on Kohlberg's moral judgment interview makes sense. In Kohlbergian interviews, people are presented with complex dilemmas that involve conflicts between moral norms, social obligations, and systems of cooperation. Interviewees customarily respond to Kohlbergian dilemmas in academic contexts in which reason and logic are valued, and the benefits of presenting oneself as a rational and intelligent person are high. Interviewers encourage interviewees to do their best thinking and give them as much time as they need to make decisions. Because Kohlbergian dilemmas are hypothetical, the costs of deriving impartial solutions are minimal (Krebs & Denton, 2005).

<sup>15</sup>Although theorists such as Saltzstein and Kasachkoff (2004) have implied that reason is the only agency qualified to resolve such conflicts, this is not necessarily the case. Animals that lack sophisticated reasoning abilities are able to resolve conflicts between prosocial and antisocial dispositions, and recent research on the neurological correlates of decision making suggests that the "brain's neural circuitry actually makes choices by internally representing the payoffs of various alternatives as neural firing rates, and choosing a maximal such rate (Glincher, 2003). . . ." (Gintis, 2007, p. 10).

Pizzaro and Bloom (2003) and Saltzstein and Kasachkoff (2004) have argued that moral reasoning may become automatic if repeated sufficiently often (Bargh, 1996).<sup>16</sup> In Darwin's terms, people may develop moral habits. This hypothesis needs to be tested. I have argued that it is implausible that the general forms of moral reasoning that define cognitive-developmental stages of moral development can become automatic (see Gibbs, 2006, and Krebs & Denton, 2006, for different views on this issue).

It has been argued that rational people may take measures in advance to control the kinds of affective reactions and intuitions they experience. For example, they may resolve to think before they act or to consider everyone's perspective before making a decision (Pizzaro & Bloom, 2003). Viewed from an evolutionary perspective, however, there is little reason to be optimistic about people's ability to keep such resolutions in strong situations (Haidt, 2003; Zimbardo, 2005). But all is not lost. When people anticipate encountering compelling temptations, they may take preventative measures to avoid them (Burnham & Phelan, 2000). For example, alcoholics may avoid bars, and sex addicts may avoid pornography. Even when encounters with temptations are inevitable, people may take measures to attend to their least activating aspects or bind themselves in advance, like Ulysses did when he tied himself to the mast of his ship to avoid the lure of the Sirens (Elster, 2000). The mental mechanisms necessary for such long-term planning are located in the newest parts of the brain, the frontal lobes.

### REFRAMING ONTOGENETIC ACCOUNTS OF THE ACQUISITION OF MORALITY

An evolutionary framework supplies a basis for reconceptualizing psychological models of the acquisition of morality in ways that integrate their insights and redress their limitations. Family contexts are microcosms of larger social groups. Members of families must resolve fundamental social dilemmas. Because parents and children need each other to propagate their genes, it is in their genetic interest to help one another and uphold familial systems of cooperation. For this reason, mechanisms have evolved that enable mothers and children to coordinate their behavior in mutually beneficial ways (Bugental & Goodnow, 1998). However, it often is in family members' interest to favor themselves and those with whom they share the largest complement of genes (Trivers, 1974). Conflicts of interest precipitate strategic interactions in which members of families attempt to induce one another to behave in ways that maximize their genetic benefits.

An evolutionary framework is equipped to organize the findings from research on socialization in a coherent manner. Because parents and other adults are more powerful than children,

they are more prone to invoke power-based tactics, such as threats and punishments, to activate deferential dispositions. Parents also may use love withdrawal (a form of social disapproval) and induction (i.e., persuasion) to induce their offspring to behave in prosocial ways (Baumrind, 1989). Because children are less powerful than their parents and more dependent on them, it is in their interest to comply with their parents' injunctions and cultivate their good favor. Because children are less knowledgeable than their parents, it is in their vested interest to learn from them (cf. Piaget, 1932). As recognized by Darwin (1874), children may acquire irrational beliefs and customs from their parents: "A belief constantly inculcated during the early years of life, whilst the brain is impressible, appears to acquire almost the nature of an instinct; and the very essence of an instinct is that it is followed independently of reason" (p. 118).

However, viewed from an evolutionary perspective, children are not expected to conform to their parents' injunctions indiscriminately or docilely; they are expected to resist conforming to those that run contrary to their adaptive interests and to actively attempt to manipulate and control other members of their families. Children are expected to negotiate and argue with their parents—and this may be a good thing. There is evidence that socialization practices that involve rational exchanges and discussions between parents and children foster moral development (Baumrind, 1989). Contemporary accounts of conscience that view the child "as an agent in moral socialization who actively processes parental moral messages," and engages in "discourse" with his or her parents (Kochanska & Aksan, 2004, p. 303) fit comfortably in a neo-Darwinian framework that emphasizes the role of strategic interaction in the selection of moral beliefs and standards.

An evolutionary analysis of morality implies a different interpretation from that offered by cognitive-developmental theorists of evidence that children acquire increasingly sophisticated structures of moral reasoning as they develop. Life history theory implies that children develop increasingly sophisticated moral reasoning abilities because these abilities are necessary for them to solve the increasingly complex and embedded social problems they encounter as they progress through the lifespan. At first, children's main social interactions occur within family contexts. The reason why young children are inclined to make predominantly "Stage 1" moral judgments that advocate obedience to authority (Kohlberg, 1984) is because it is adaptive for young children to adopt deferential strategies. As children grow older, they interact increasingly often with peers. They become increasingly prone to make "Stage 2" moral judgments that uphold reciprocal exchanges because reciprocity is a more adaptive strategy than deference in egalitarian relations among peers. With advancing age, young adults invest in cliques and coalitions. They are prone to make "Stage 3" and "Stage 4" moral judgments because such judgments uphold the harmonious types of ingroup relations, long-term commitments, and systems of indirect reciprocity that

<sup>16</sup> "Having learned (or decided) how to respond to certain moral situations and having responded subsequently to like situations in the same way over and over again, we now no longer have to deliberate before responding" (Saltzstein & Kasachkoff, 2004, p. 279).

foster their interests (see Krebs, 2005a, for elaborations of these ideas). The reason why people are prone to make different kinds of moral judgments at different ages on semiprojective tests of moral reasoning such as Kohlberg's *Moral Judgment Interview* is because different types of social relations and systems of cooperation are salient to them at different phases of their lives (Krebs & Denton, 2005).

Looked at in this way, we would not expect new structures of moral reasoning to transform and displace older structures, as Colby and Kohlberg (1987) have asserted. We would expect people to possess many domain-specific structures of moral reasoning, and we would expect people to acquire them in an additive-inclusive way (Eisenberg, 1982; Levine, 1979; Krebs, Vermeulen, et al., 1991), because people continue to experience throughout their lives the kinds of adaptive problems that early structures evolved to solve. Adults may, for example, find themselves in subordinate positions in which it is adaptive to believe that they should show deference to authority (Milgram, 1974). Military contexts and cults come to mind. Although the justifications that adults advance for obeying authority and engaging in tit-for-tat exchanges may be more sophisticated than those advanced by children—for example, because adults embed their justifications in principles that uphold more broadly based systems of cooperation—their decisions stem from essentially the same affective and cognitive processes.

Viewed from an evolutionary perspective, moral development consists more in the acquisition of the flexibility necessary to solve social problems in the most efficient, effective, and adaptive ways than in the ability to make highly sophisticated moral judgments in every context (Krebs & Denton, 2005). As Darwin (1874) implied, cardinal moral principles such as “foster the greatest good for the greatest number” may be translated as, “foster your ultimate adaptive interests by upholding the standards, social relations, and systems of cooperation that produce the greatest adaptive benefits for all.”

### WHAT A SENSE OF MORALITY IS

Understanding what morality is for—the adaptive functions it evolved to serve—helps explain what it is. The constellation of thoughts and feelings that constitute a sense of morality evolved to enable individuals to uphold cooperative social relations that maximized their biological benefits. To accomplish this, they had to acquire mechanisms that enabled them to resolve conflicts of interest in adaptive ways. The feature that all aspects of morality share is their purpose—to induce individuals to foster their interests in ways that also foster the interests of those with whom they interact. “In evolutionary theory, a moral person is simply one who pursues their ultimate genetic self-interest through psychological adaptations that embody a genuine, proximate concern for others” (Miller, 2007, p.103).

The reason why all people's sense of morality includes ideas about altruism, social contracts, rights and duties, justice, au-

thority, punishment, resisting temptation, reconciliation, purity, and so on is because these issues inevitably arise when people who can maximize their gains by giving less than their share and by taking more implicitly or explicitly agree to coordinate their efforts to foster their mutual interests. The domain of morality pertains to the formal and informal rules and sanctions that uphold the systems of cooperation that enable members of groups to survive, to reproduce, and to propagate their genes. Some ideas and affective experiences pertain to people's obligation to contribute to the welfare of those with whom they have formed cooperative relations (e.g., a sense of duty, responsibility, respect for authority, commitment, loyalty, and sympathy), whereas others pertain to people's sense of what they deserve by virtue of their contributions.

People are evolved to feel pride when they support their groups and contribute to the welfare of those with whom they have formed social bonds and to feel guilt and shame when fail to fulfill their social obligations. People are evolved to feel gratitude when others contribute to their welfare and to feel moral outrage when others betray them and violate norms of fairness. Conceptions of justice and the affective reactions that accompany them pertain to the balance between what members of groups owe and what they deserve.

Universal moral norms evolve because they constitute the most effective means of upholding mutually beneficial cooperative relations. Culturally relative norms evolve because people from different cultures experience different adaptive problems, because people solve the same adaptive problems in different ways, and because norms may emerge from social and cultural processes that bear only an indirect relation to biological evolution (Richerson & Boyd, 2005). As with language, the “grammar” of moral decision-making is universal, but the content may vary across cultures.

### CONCLUSION

A sense of morality is a complex and multifaceted phenomenon. To account for it, we must not only understand how it develops in children, but also how it originated in the human species. Revisiting Darwin's early insights in light of contemporary expansions and refinements of evolutionary theory gives rise to a framework equipped to accommodate relevant theory and research and to resolve many of the controversies in the field. The key to understanding what a sense of morality is and how people acquire it lies in explaining how the mechanisms that produce it evolved, which entails identifying the adaptive functions that they evolved to serve. The practical implications of an evolutionary approach to morality do not direct us to meddle with people's genes. They encourage us to map the design of the evolved mechanisms that give rise to moral decisions, discover how different mechanisms interact, and take measures to ensure that they are calibrated and activated in optimal ways (Krebs, 2004).



## REFERENCES

- Alcock, J. (2005). *Animal behavior: An evolutionary approach* (8th ed. Sunderland, MA: Sinauer Associates.
- Alcock, J. (in press). Evolutionary questions for evolutionary psychologists. In C.
- Crawford, & Krebs, D. (Eds.), *Foundations of evolutionary psychology* (pp. 25–46). Mahwah, NJ: Erlbaum.
- Alexander, R.D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Alexander, R.D. (1990). *How did humans evolve? Reflections on the uniquely unique species* (Museum of Zoology, Special Publication No. 1). Ann Arbor: University of Michigan Press.
- Aron, A., & Aron, E.N. (1986). *Love and the expansion of self: Understanding attraction and satisfaction*. New York: Hemisphere.
- Aronfreed, J. (1968). *Conduct and conscience*. New York: Academic Press.
- Atran, S. (2004). *In Gods we trust*. Oxford, United Kingdom: Oxford University Press.
- Axelrod, R. (1988). The further evolution of cooperation. *Science*, 242, 1385–1389.
- Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.
- Axelrod, R., & Hamilton, W.D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.
- 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.
- Bargh, M.A. (1996). Automaticity in social psychology. In E.T. Higgins & A.W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (Vol. 1, pp. 1–40). New York: Guilford Press.
- Barinaga, M. (1996). Social status sculpts activity of crayfish neurons. *Science*, 271, 290–291.
- Barkow, J.H. (2006). *Missing the revolution: Darwinism for social scientists*. Oxford, United Kingdom: Oxford University Press.
- Batson, C.D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Batson, C.D. (2000). Unto others: A service ... and a disservice. *Journal of Consciousness Studies*, 7, 207–210.
- Baumeister, R.F., Stillwell, A.M., & Heatherton, T.F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115, 243–267.
- Baumrind, D. (1989). Rearing competent children. In W. Damon (Ed.), *Child development today and tomorrow* (pp. 349–378). San Francisco: Jossey-Bass.
- Berkowitz, L. (1972). Social norms, feelings, and other factors affecting helping behavior and altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 62–106). New York: Academic Press.
- Bloom, P. (2004). *Descartes*. New York: Basic Books.
- Boehm, C. (2000). Conflict and the evolution of social control. *Journal of Consciousness Studies*, 7, 79–101.
- Boyd, R., & Richerson, P.J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171–195.
- Boyer, P. (2001). *Religion explained: The human instincts that fashion Gods, spirits, and ancestors*. London: William Heinemann.
- Brown, R. (1986). *Social psychology* (2nd ed.) New York: The Free Press.
- Brown, S.L., & Brown, M. (2006). Selective investment theory: Recasting the functional significance of close relationships. *Psychological Inquiry*, 17, 30–59.
- Bugental, D.B., & Goodnow, J.J. (1998). Socialization. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Volume 3. Social, emotional, and personality development* (5th. ed., pp. 389–462). Hoboken, NJ: Wiley.
- Burnham, T., & Phelan, J. (2000). *Mean genes*. New York: Perseus Publishing.
- Burnstein, E. (2005). Altruism and genetic relatedness. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 528–551). Hoboken, NJ: John Wiley & Sons.
- Buss, D. (2004). *Evolutionary psychology: The new science of the mind* (2nd Ed.). Boston: Allyn and Bacon.
- Byrne, R.W., & Whiten, A. (Eds.). (1988). *Machiavellian intelligence: Social experience and the evolution of intellect in monkeys, apes, and humans*. Oxford, United Kingdom: Oxford University Press.
- Campbell, D.T. (1978). On the genetics of altruism and the counterhedonic components in human nature. In L. Wispé (Ed.), *Altruism, sympathy, and helping: Psychological and sociological implications* (pp. 39–58). New York: Academic Press.
- Chaiken, S. (1987). The heuristic model of persuasion. In M.P. Zanna, J.M. Olson, & C.P. Herman (Eds.), *Social influence: The Ontario Symposium* (pp. 3–39). Hillsdale, NJ: Erlbaum.
- Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P.M. Gollwitzer & J.A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578). New York: Guilford Press.
- Cheney, D.L., & Seyfarth, R.M. (1990). *How monkeys see the world: Inside the mind of another species*. Chicago: University of Chicago Press.
- Cialdini, R.B., Borden, R.J., Thorne, A., Walker, M.R., Freeman, S., & Sloan, L.R. (1976). Three (football) field studies. *Journal of Personality and Social Psychology*, 34, 366–375.
- 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, United Kingdom: Cambridge University Press.
- Crespi, B.J. (2000). The evolution of maladaptation. *Heredity*, 84, 623–629.
- Cummins, D. (2005). Dominance, status, and social hierarchies. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 676–697). Hoboken, NJ: John Wiley & Sons.
- Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Grosset/Putnam.
- Damon, W. (1980). Patterns of change in children's social reasoning: A two-year longitudinal study. *Child Development*, 46, 1010–1017.
- Darwin, C. (1874). *The descent of man and selection in relation to sex*. New York: Rand, McNally & Company.
- Dawkins, R. (1989). *The selfish gene*. Oxford, United Kingdom: Oxford University Press.
- Decety, J. (2005). Perspective taking as the royal avenue to empathy. In B.F. Malle & S.D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 143–157). New York: Guilford Press.
- de Waal, F.B.M. (1982). *Chimpanzee politics*. London: Jonathan Cape.
- de Waal, F.B.M. (1991). The chimpanzee's sense of social regularity and its relation to the human sense of justice. *American Behavioral Scientist*, 34, 335–349.

- 1 de Waal, F.B.M. (1996). *Good natured: The origins of right and wrong*  
2 *in humans and other animals*. Cambridge, MA: Harvard Uni-  
3 versity Press.
- 4 Dugatkin, L.A. (1997). *Cooperation among animals: An evolutionary*  
5 *perspective*. New York: Oxford University Press.
- 6 Dunbar, R.I.M. (1996). Determinants of group size in primates: A  
7 general model. In G. Runciman, J. Maynard Smith, & R.I.M.  
8 Dunbar (Eds.), *Evolution of social behavior patterns in primates*  
9 *and man* (pp. 33–58). Oxford, United Kingdom: Oxford Uni-  
10 versity Press.
- 11 Duntley, J.D. (2005). Adaptations to dangers from humans. In D. Buss  
12 (Ed.), *The handbook of evolutionary psychology* (pp. 224–254).  
13 Hoboken, NJ: John Wiley & Sons.
- 14 Durham, W.H. (1991). *Coevolution: Genes, culture and human diversity*.  
15 Stanford, CA: Stanford University Press.
- 16 Eisenberg, N. (1982). *The development of prosocial behavior*. New York:  
17 Academic Press.
- 18 Eisenberger, N.I., Lieberman, M.D., & Williams, K.D. (2003). Does  
19 rejection hurt? An fMRI study of social exclusion. *Science*, 302,  
20 290–292.
- 21 Elster, J. (2000). *Ulysses unbound: Studies in rationality, precommit-*  
22 *ment, and constraints*. Cambridge, United Kingdom: Cambridge  
23 University Press.
- 24 Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans.  
25 *Nature*, 415, 137–140.
- 26 Fehr, E., & Gächter, S. (2003). Reply. *Nature*, 421, 912.
- 27 Festinger, L. (1964). *Conflict, decision, and dissonance*. Stanford, CA:  
28 Stanford University Press.
- 29 Fisher, H. (2004). *Why we love: The nature and chemistry of romantic*  
30 *love*. New York: Henry Holt.
- 31 Fiske, A.P. (1992). Four elementary forms of sociality: Framework for a  
32 unified theory of social relations. *Psychological Review*, 99, 689–  
33 723.
- 34 Flack, J.C., & de Waal, F.B.M. (2000). “Any animal whatever”: Dar-  
35 winian building blocks of morality in monkeys and apes. *Journal*  
36 *of Consciousness Studies*, 7, 1–29.
- 37 Flinn, M.V., & Alexander, R.D. (1982). Culture theory: The developing  
38 synthesis from biology. *Human Ecology*, 10, 383–400.
- 39 Frank, R.H. (1988). *Passions within reason: The strategic role of the*  
40 *emotions*. New York: Norton.
- 41 Frank, R.H. (2001). Cooperation through emotional commitment. In  
42 R. Nesse (Ed.), *Evolution and the capacity for commitment* (pp.  
43 57–76). New York: Russell Sage Foundation.
- 44 Freud, S. (1925). *Collected papers*. London: Hogarth Press.
- 45 Gangestad, S.W., & Simpson, J.A. (2007). *The evolution of mind:*  
46 *Fundamental questions and controversies*. New York: Guilford.
- 47 Gaulin, S.J., & McBurney, D.H. (2004). *Evolutionary psychology* (2nd  
48 ed.). Upper Saddle River, NJ: Pearson Education.
- 49 Ghiselin, M.T. (1974). *The economy of nature and the evolution of sex*.  
50 Berkeley: University of California Press.
- 51 Gibbs, J.C. (2003). *Moral development and reality: Beyond the theories*  
52 *of Kohlberg and Hoffman*. Thousand Oaks, CA: Sage.
- 53 Gibbs, J.C. (2006). Should Kohlberg’s cognitive developmental ap-  
54 proach to morality be replaced with a more pragmatic approach?  
55 Comment on Krebs and Denton (2005). *Psychological Review*,  
56 113, 666–671.
- 57 Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*.  
58 New York: Oxford University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases:*  
*The psychology of intuitive judgment*. Cambridge, United King-  
dom: Cambridge University Press.
- Gilovich, T., Keltner, D., & Nisbett, R.E. (2006). *Social psychology*.  
New York: Norton.
- Gintis, H. (2000). *Game theory evolving: A problem-centered intro-*  
*duction to modeling strategic interaction*. Princeton, NJ: Prince-  
ton University Press.
- Gintis, H. (2001). Forward. In R.M. Nesse (Ed.), *Evolution and the*  
*capacity for commitment* (pp. i–xviii). New York: Russell Sage  
Foundation.
- Gintis, H. (2007). A framework for the unification of the behavioral  
sciences. *Behavioral and Brain Sciences*, 30, 1–16.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining al-  
truistic behavior in humans. *Evolution and Human Behavior*, 24,  
153–172.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2005). *Moral sentiments*  
*and material interests: On the foundations of cooperation in eco-*  
*nomic life*. Cambridge, MA: MIT Press.
- Gintis, H., Smith, E.A., & Bowles, S. (2001). Costly signaling and  
cooperation. *Journal of Theoretical Biology*, 213, 103–119.
- Gouldner, A.W. (1960). The norm of reciprocity: A preliminary  
statement. *American Sociological Review*, 25, 161–178.
- 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.
- Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Co-  
hen, J.D. (2001). An fMRI investigation of emotional engagement  
in moral judgments. *Science*, 293, 2105–2108.
- Habermas, J. (1993). *Justification and application*. Cambridge, MA:  
MIT Press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social in-  
tuitionist approach to moral judgment. *Psychological Review*,  
108, 814–834.
- Haidt, J. (2003). The emotional dog does learn new tricks: A reply  
to Pizzaro and Bloom (2003). *Psychological Review*, 110, 197–  
198.
- Hamilton, W.D. (1964). The evolution of social behavior. *Journal of*  
*Theoretical Biology*, 7, 1–52.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162, 1243–  
1248.
- Hartung, J. (1987). Deceiving down: Conjectures on the management  
of subordinate status. In J. Lockart & D.L. Paulhus (Eds.), *Self-*  
*deception: An adaptive mechanism?* (pp. 170–185). Englewood  
Cliffs, NJ: Prentice-Hall.
- Haslam, N. (1997). Four grammars for primate social relations. In J.A.  
Simpson & D.T. Kenrick (Eds.), *Evolutionary social psychology*  
(pp. 297–316). Mahwah, NJ: Erlbaum.
- Hauser, M.D. (2006). *Moral minds: How nature designed our universal*  
*sense of right and wrong*. New York: Harper Collins.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., &  
McElreath, R. (2001). In search of homo economicus: Behavioral  
experiments in 15 small-scale societies. *American Economic*  
*Review*, 91, 73–78.
- Higgins, E.T. (1987). Self-discrepancy: A theory relating self and af-  
fect. *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.
- Hoffman, M. (2000). *Empathy and moral development: Implications for*  
*caring and justice*. Cambridge, United Kingdom: Cambridge  
University Press.
- James, W. (1890). *The principles of psychology* (Vols. 1–2). New York:  
Henry Holt.

- Janicki, M. (2004). Beyond sociobiology: A kinder and gentler evolutionary view of human nature. In C. Crawford & C. Salmon (Eds.), *Evolutionary psychology: Public policy and personal decisions* (pp. 51–72). Mahwah, NJ: Erlbaum.
- Johnson, D.P.P., Stopka, P., & Knights, S. (2003). The puzzle of human cooperation. *Nature*, 421, 911–912.
- Kagan, J. (1984). *The nature of the child*. New York: Basic Books.
- Kagan, J. (2003). Human morality and temperament. In G. Carlo & C. Pope-Edwards (Eds.), *Nebraska Symposium on Motivation: Vol. 51. Moral motivation through the life span*. Lincoln: Nebraska University Press.
- Kenrick, D.T., Li, N.P., & Butner, J. (2006). Dynamical evolutionary psychology: How social norms emerge from individual decisions rules. In P.A.M. Van Lange & A. Kruglanski (Eds.), *Bridging social psychology: The benefits of transdisciplinary approaches* (pp. 285–292). Hillsdale, NJ: Erlbaum.
- Kochanska, G., & Aksan, N. (2004). Conscience in childhood: Past, present, and future. *Merrill-Palmer Quarterly*, 50, 299–310.
- Kohlberg, L. (1984). *Essays in moral development: The psychology of moral development* (Vol. 2). New York: Harper & Row.
- Krebs, D.L. (2004). 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* (pp. 319–342). Hillsdale, NJ: Erlbaum.
- Krebs, D.L. (2005a). An evolutionary reconceptualization of Kohlberg's model of moral development. In R. Burgess & K. MacDonald (Eds.), *Evolutionary perspectives on human development* (pp. 243–274). Thousand Oaks, CA: Sage Publications.
- Krebs, D.L. (2005b). The evolution of morality. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 747–771). Hoboken, NJ: John Wiley & Sons.
- Krebs, D.L. (2006). The nature of altruism: Ultimate determinants and proximate mechanisms. *Psychological Inquiry*, 17, 48–50.
- Krebs, D.L. (2007). Understanding evolutionary approaches to human behavior. *Human Development*, 50, 286–291.
- Krebs, D.L. (in press). How selfish by nature? In C. Crawford & D.L. Krebs (Eds.), *Foundations of evolutionary psychology*. New York: Taylor Francis.
- 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* (pp. 220–263). Frankfurt, Germany: M. Suhrkamp.
- Krebs, D.L., & Denton, K. (2005). Toward a more pragmatic approach to morality: A critical evaluation of Kohlberg's model. *Psychological Review*, 112, 629–649.
- Krebs, D.L., & Denton, K. (2006). Explanatory limitations of cognitive-developmental approaches to morality. *Psychological Review*, 113, 672–675.
- Krebs, D., 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). New York: Springer-Verlag.
- 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. (1997). The forms and functions of real-life moral decision-making. *Journal of Moral Education*, 20, 131–145.
- Krebs, D.L., & Janicki, M. (2004). The biological foundations of moral norms. In M. Schaller & C. Crandall (Eds.), *Psychological Foundations of Culture* (pp. 25–148). Hillsdale, NJ: Erlbaum.
- 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., Vermeulen, S.C., Carpendale, J.I., & Denton, K. (1991). Structural and situational influences on moral judgment: The interaction between stage and dilemma. In W. Kurtines & J. Gewirtz (Eds.), *Handbook of moral behavior and development: Theory, research, and application* (pp. 139–169). Hillsdale, NJ: Erlbaum.
- Kruger, D.L. (2003). Evolution and altruism: Combining psychological mediators with naturally selected tendencies. *Evolution and Human Behavior*, 24, 118–125.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674–689.
- Kunda, Z. (2000). *Social cognition: Making sense of people*. Cambridge MA: MIT Press.
- Kurland, J.A., & Gaulin, S.J.C. (2005). Cooperation and conflict among kin. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 447–482). Hoboken, NJ: John Wiley & Sons.
- Laland, K.N., Odling-Smee, F.L., & Feldman, M. (2000). Group selection: A niche-construction perspective. *Journal of Consciousness Studies*, 7, 221–223.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make the work light: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 822–832.
- Leakey, R.E., & Lewin, R. (1977). *Origins*. New York: Dutton.
- Lerner, M.J., & Miller, D.T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, 85, 1030–1051.
- Levine, C.G. (1979). Stage acquisition and stage use: An appraisal of stage displacement explanations of variation in moral reasoning. *Human Development*, 22, 145–164.
- Lickliter, R., & Honeycutt, H. (2003). Developmental dynamics: Towards a biologically plausible evolutionary psychology. *Psychological Bulletin*, 129, 819–835.
- 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.
- Lorenz, K.Z. (1966). *On aggression*. London: Methuen.
- Lykken, D.T. (1992). *The antisocial personalities*. Hillsdale, NJ: Erlbaum.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, United Kingdom: Cambridge University Press.
- McCullough, M.E., Kilpatrick, S.D., Emmons, R.A., & Larson, D.B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127, 249–266.
- 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). Hillsdale, NJ: Erlbaum.
- Miller, G.F. (2007). The sexual selection of moral virtues. *The Quarterly Review of Biology*, 82, 97–125.
- Moll, J., Zahn, R., di Oliveira-Sourza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Neuroscience*, 6, 799–809.
- Nesse, R.M. (Ed.). (2001). *Evolution and the capacity for commitment*. New York: Russell Sage Foundation.
- Nowak, M.A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393, 573–577.



- 1 Petty, R.E., & Cacioppo, J.T. (1986). The elaboration likelihood model  
2 of persuasion. In L. Berkowitz (Ed.), *Advances in experimental*  
3 *social psychology* (pp. 123–205). New York: Academic Press.
- 4 Piaget, J. (1932). *The moral judgment of the child*, Routledge & Kegan  
5 Paul.
- 6 Piliavin, J.A., Dovidio, J.F., Saertner, S.L., & Clark, R.D. (1981).  
7 *Emergency intervention*. New York: Academic Press.
- 8 Pizzaro, D.A., & Bloom, P. (2003). The intelligence of the moral intu-  
9 titions: A reply to Haidt (2001). *Psychological Review*, 110,  
10 193–196.
- 11 Price, M.E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an  
12 anti-free rider psychological device. *Evolution and Human Be-*  
13 *havior*, 23, 203–231.
- 14 Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of  
15 cognitive and motivational perspectives on social inference: A  
16 biased hypothesis-testing model. *Advances in Experimental Social*  
17 *Psychology*, 20, 297–340.
- 18 Rawls, J. (1999). *A theory of justice* (Rev. ed.). Cambridge, MA: Har-  
19 vard University Press.
- 20 Rest, J.F. (1983). Morality. In J.H. Flavell & E.M. Markman (Eds.),  
21 *Handbook of child psychology: Vol. 3. Cognitive development*  
22 (4th ed., pp. 556–629). New York: Wiley.
- 23 Richards, R.J. (1987). *Darwin and the emergence of evolutionary the-*  
24 *ories of mind and behavior*. Chicago: University of Chicago Press.
- 25 Richerson, P.J., & Boyd, R. (2001). The evolution of subjective com-  
26 mitment: A tribal instincts hypothesis. In R. Nesse (Ed.), *Evo-*  
27 *lution and the capacity for commitment* (pp. 186–219). New York:  
28 Russell Sage Foundation.
- 29 Richerson, P.J., & Boyd, R. (2005). *Not by genes alone: How culture*  
30 *transformed human evolution*. Chicago: University of Chicago  
31 Press.
- 32 Ridley, M. (1996). *The origins of virtue: Human instincts and the*  
33 *evolution of cooperation*. New York: Viking.
- 34 Rozin, P., Haidt, J., & McCauley, C.R. (2000). Disgust. In M. Lewis &  
35 J.M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp.  
36 637–653). New York: Guilford Press.
- 37 Sachs, J.L., Mueller, U.G., Wilcox, T.P., & Bull, J.J. (2004). The evo-  
38 lution of cooperation. *Quarterly Review of Biology*, 79, 135–160.
- 39 Saltzstein, H.D., & Kasachkoff, T. (2004). Haidt's moral intuitionist  
40 theory: A psychological and philosophical critique. *Review of*  
41 *General Psychology*, 8, 273–282.
- 42 Schachter, S., & Singer, J.E. (1962). Cognitive, social and psycho-  
43 logical determinants of emotional state. *Psychological Review*, 69,  
44 379–399.
- 45 Schaller, M. (2007). Turning garbage into gold: Evolutionary univer-  
46 sals and cross-cultural differences. In S.W. Gangestad & J.A.  
47 Simpson (Eds.), *The evolution of mind: Fundamental questions*  
48 *and controversies* (pp. 363–371). New York: Guilford.
- 49 Segal, N.L., Hersherberger, S.L., & Arad, S. (2003). Meeting one's twin:  
50 Perceived social closeness and familiarity. *Evolutionary Psy-*  
51 *chology*, 1, 70–95.
- 52 Selman, R.L. (1980). *The growth of interpersonal understanding*. New  
53 York: Academic Press.
- 54 Shackelford, T.K., & Buss, D.M. (1996). Betrayal in mateships,  
55 friendships, and coalitions. *Personality and Social Psychology*  
56 *Bulletin*, 22, 1151–1164.
- 57 Shweder, R.A., Much, N.C., Mahapatra, M., & Park, L. (1997). The  
58 “big three” of morality (autonomy, community, divinity) and the  
“big three” explanations of suffering. In A. Brandt & P. Rozin  
(Eds.) *Morality and health* (pp. 119–169). New York: Routledge.
- Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. New  
York: Farrar, Straus and Giroux.
- Simon, H. (1990). A mechanism for social selection of successful  
altruism. *Science*, 250, 1665–1668.
- Skyrms, B. (2000). Game theory, rationality, and evolution of the social  
contract. *Journal of Consciousness Studies*, 7, 269–284.
- Sober, E., & Wilson, D.S. (1998). *Unto others: The evolution and*  
*psychology of unselfish behavior*. Cambridge, MA: Harvard Uni-  
versity Press.
- Sunstein, C.R. (2005). Moral heuristics. *Behavioral and Brain Sci-*  
*ences*, 28, 531–573.
- Tajfel, H., & Turner, J.C. (1985). The social identity theory of inter-  
group behavior. In S. Worchel & W.G. Austin (Eds.), *Psychology*  
*of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.
- 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. (1974). Parent-offspring conflict. *American Zoologist*, 14,  
249–264.
- Trivers, R. (1985). *Social evolution*, Menlo Park, CA, Benjamin  
Cummings.
- Trivers, R. (2000). The elements of a scientific theory of self-decep-  
tion. In D. LeCroy & P. Moller (Eds.), *Evolutionary perspectives on*  
*human reproductive behavior* (pp. 114–131). New York: New York  
Academy of Sciences.
- Trivers, R. (2006). Reciprocal altruism: 30 years later. In P.M. Kapeler  
& C.P. van Schaik (Eds.), *Cooperation in primates and humans*  
(pp. 67–84). New York: Springer-Verlag.
- Turiel, E. (2002). *The culture of morality*. Cambridge, England:  
Cambridge University Press.
- van den Berghe, P.L. (1983). Human inbreeding avoidance: Culture in  
nature. *Behavioral and Brain Sciences*, 6, 91–123.
- Vasquez, K., Keltner, D., Edenbach, D.H., & Banaszynski, T.L. (2001).  
Cultural variation and similarity in moral rhetorics: Voices from  
the Phillipines and United States. *Journal of Cross-Cultural*  
*Psychology*, 32, 93–120.
- Wedekind, C., & Milinski, M. (2000). Cooperation through image  
scoring in humans. *Science*, 288, 850–852.
- Williams, G.C. (1966). *Adaptation and natural selection: A critique of*  
*some current evolutionary thought*. Princeton, NJ: Princeton  
University Press.
- 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, NJ: Princeton University Press.
- Wilson, E.O. (1998). *Consilience: The unity of knowledge*. New York:  
Knopf.
- Wilson, J.Q. (1993). *The moral sense*. New York: Free Press.
- Wright, R. (1994). *The moral animal*. New York: Pantheon Books.
- Zahavi, A., & Zahavi, A. (1997). *The handicap principle*. New York:  
Oxford University Press.
- Zimbardo, P. (2005). A situationalist perspective on the psychology of  
good and evil: Understanding how good people are transformed  
into perpetrators. In A.G. Miller (Ed.), *The social psychology of*  
*good and evil* (pp. 21–50). New York: Guilford Press.



## Author Query Form

Journal **PPSC**

Article 72

*Dear Author,*

*During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers clearly on the query sheet if there is insufficient space on the page proofs. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.*

[illegible]