12

The Evolution of a Sense of Justice

DENNIS L. KREBS

Everyone possesses a sense of justice, however misguided it may be. How do people acquire this sense? Where does it come from? In this chapter, I argue that to account for the acquisition of a sense of justice, we must identify the mental mechanisms that produce it and explain how they originated and became refined in the course of human evolution. Explaining how a sense of justice originated in the human species helps us understand what it is, what it is for, how it is designed, what activates it, and why it sometimes fails to give rise to fair judgments and behaviors.

A Working Definition of a "Sense of Justice"

A sense of justice consists of thoughts and feelings about what is fair and unfair and what people deserve from and owe others (rights and duties). When we think of justice, we think of balanced scales. In *Nicomachean Ethics*, Aristotle distinguished three forms of justice. The first pertains to how resources should be distributed (*distributive justice*)—for example, in terms of principles of equality, equity, desert, and merit. The second pertains to agreements between people—promises, commitments, and other kinds of social contracts (*commutative justice*). The final type pertains to the righting of wrongs (*corrective justice*). It includes ideas such as forgiveness and a bunch of "r" words—revenge, reparation, restitution, and retribution ("getting even"). Overriding all of these forms of justice is *procedural justice*. To make fair decisions, people must use fair and impartial procedures such as the Golden Rule, balanced discussion, or democratic decision making.

Psychological Accounts of the Origin of a Sense of Justice

If you ask people how they acquired their sense of justice, most people, from the Western world at least, would advance a social learning account. They would say that they acquired a sense of justice from their parents and other mentors, who taught them to behave fairly, to share, to take turns, to keep their promises, and so on. Although it would be foolish to deny that social learning plays a role in the acquisition of a sense of justice, more is involved. If children internalized their parents' ideas about fairness, then children would possess the same ideas their parents do, but they do not. Children argue with their parents. They have minds of their own. They are able to think for themselves. And the ways in which they think about fairness changes as they develop. Cognitive-developmental theorists such as Kohlberg (1984) and Piaget (1932) have argued that children derive their conceptions of justice from structures of moral reasoning.

Like social learning, reasoning plays a role in determining people's sense of justice. However, like social learning, it does not account for all aspects of this sense. As demonstrated by Haidt (2001), people sometimes simply feel that a behavior is fair or unfair, right or wrong, without thinking about it or engaging in moral reasoning. If someone cheats you or breaks a promise to you, you may experience an immediate sense of righteous indignation without engaging in rational deliberation.

The goal of virtually all psychological research on a sense of justice is to decipher the design of the proximate mechanisms that produce it. Theoretical differences arise with respect to the types of mechanisms responsible for producing it (e.g., social learning versus reasoning versus affective mechanisms), the ways in which they are designed (e.g., whether people possess one overriding structure of moral reasoning or a bunch of different, domain-specific structures designed to deal with different aspects of justice), and the ways in which they interact (e.g., whether reason structures affective reactions, or whether affective reactions structure reason). Although adherents of different psychological approaches each tend to assume that their approach offers a full account of the acquisition of a sense of justice, it is clear that each approach accounts for only part of the process. A sense of justice stems from a system of mechanisms. Sometimes people derive conceptions of justice from one mechanism, sometimes from another. Sometimes more than one mechanism is activated, and when this occurs, the activated mechanisms may work in concert to support the same decision, or they may engender internal conflict. What is needed is an overarching framework that accounts for the origin of this system and integrates its components in meaningful ways. The thesis of this chapter is that evolutionary theory fills this bill.

An Overview

The mechanisms that produce a sense of justice did not emerge in the human species one sunny morning in full-blown glory. They emerged slowly over eons, through the modification of more primitive mechanisms. Although this was a continuous process, it is helpful heuristically to break it down into overlapping phases. I will sug-

gest that the first phase in the evolution of a sense of justice involved the evolution of cooperative behavioral dispositions and the affective reactions that support them. Precursors of this sense can be seen in chimpanzees and other primates. In the second phase, this primitive sense became refined and elaborated in the context of strategic interactions among members of groups motivated to induce one another to behave in cooperative ways. The acquisition of the capacity for symbolic language, perspective-taking, and sophisticated forms of intelligence played important roles in this process, which gave rise to moral judgments and moral norms. In the final phase, humans acquired the capacity to imagine ideal social systems; to reflect on moral issues; to figure out how, in principle, to solve complex moral problems; and to develop ideal conceptions of justice.

The Evolution of Cooperation

From an evolutionary perspective, the key to understanding the origin of a sense of justice lies in identifying the adaptive functions it evolved to serve. I will argue that the overarching function of a sense of justice is to induce members of groups to uphold fitness-enhancing forms of cooperation. To understand the emergence of the mechanisms that produce a sense of justice, we must first understand the emergence of the mechanisms that induce animals to cooperate.

There is tremendous adaptive potential in cooperation. In conducive contexts, two or more animals that work together and exchange goods and services can enhance their fitness much more effectively than they could by going it on their own. This does not, however, guarantee the evolution of cooperative dispositions. All kinds of traits and behaviors could enhance animals' fitness better than those they already possess. For cooperative dispositions to evolve, individuals must inherit genes that guide the creation of mechanisms that dispose them to behave in cooperative ways, and these mechanisms must pay off better genetically than competing mechanisms such as those that dispose them to behave in selfish ways.

The Fundamental Social Dilemma

Assume that, originally, animals were disposed to help only themselves. It is relatively easy to account for the evolution of mutualistic behaviors such as group hunting and group defense, because the animals that engage in such behaviors could be coordinating their efforts to maximize their own biological gains, helping others only incidentally. In contrast, it is considerably more difficult to account for the selection of mechanisms that dispose animals to engage in equitable exchanges. As expressed by the philosopher Rawls (1999) in the opening pages of *A Theory of Justice*,

> Although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since

> social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. (Rawls, 1999, p. 4)

Modeled in evolutionary terms, assume that members of a group inherit genes that dispose them to adopt one of two strategies—either to behave fairly (i.e., to cooperate) or to behave selfishly (i.e., to cheat). If all members of a group inherited genes that disposed them to cooperate, everyone could obtain more for himself or herself though gains in trade than he or she could by failing to cooperate, and the group could prevail in competitions against less cooperative groups. Cooperation could produce a utopia for all. The problem is, if other members of one's group behave cooperatively, each individual can come out ahead by doing less than his or her share and taking more. If those who are disposed to behave selfishly contribute more replicas of their genes to future generations than those who are disposed to behave fairly, selfish dispositions will be selected and evolve. Ironically, however, as the number of selfish members of a group increases, there are fewer and fewer cooperative individuals to exploit, jeopardizing the system of cooperation and forcing selfish individuals to interact with one another, to their mutual detriment.

Theoretical Resolutions of the Fundamental Social Dilemma: The Selection of Cooperative Strategies

Game theorists have created computerized simulations of evolution in which they have pitted cooperative strategies against selfish strategies. These theorists have found that certain conditionally cooperative strategies, such as various forms of tit-for-tat (that is to say, strategies based on the decision rule, "make an initial cooperative overture, then copy the response of your partner,") and variations such as tit-for-two-tats, are equipped to defeat unconditionally selfish strategies and evolve in favorable conditions. The cooperative strategies gain their power either by reducing the costs and increasing the benefits of behaving fairly or by increasing the costs and reducing the benefits of behaving unfairly. The genetic costs of contributing one's share can be reduced by engaging in cooperative exchanges with those who share one's genes, by selectively engaging in exchanges with other cooperators, and by reaping indirect benefits from acquiring a reputation as a cooperator. The net genetic costs of failing to contribute one's share can be increased by diminishing the probability that recipients or others will interact with those who behave selfishly and by increasing the probability that selfish individuals will be punished—either by their interaction partners or by other members of their group. Accounting for the evolution of dispositions to punish third parties is tricky, because if we assume that it is costly to inflict punishments, those who refused to accept responsibility for administering punishments would fare better than those who accepted responsibility.

Reciprocity in Nonhuman Animals

The most appropriate place to look for precursors of the forms of cooperation practiced by humans is in other primates. Studies by de Waal and others (see Kappeler & Schaik, 2006, for a review) have established that chimpanzees engage in calculated forms of delayed reciprocity in which they remember who has helped them, track credits and debts to particular partners, and repay them either in kind or in some other currency. For example, chimpanzees are more likely to assist those who have assisted them in agonistic exchanges with others ("one good turn deserves another") and to aggress against those who have sided with others against them ("an eye for an eye") (De Waal & Luttrell, 1988). In addition, chimpanzees are more likely to share food with those who have groomed them earlier in the day. If we accept the idea that chimpanzees inherit mechanisms that dispose them to reciprocate and engage in other forms of cooperation, we can conclude that they possess mental mechanisms that enable them to solve fundamental social dilemmas and induce them to engage in primitive forms of fairness.

Games that Primates Play

The social lives of chimpanzees and members of other social species are dynamic. Members of primate groups engage in ongoing contests in which they attempt to induce one another to behave in ways that benefit them by invoking tactics such as begging, offering, enticing, screaming, threatening, attacking, and shunning. De Waal (1991) has suggested that the "active reinforcement of others" (p. 338) is responsible for the emergence of prescriptive rules in groups of chimpanzees.

In their essence, the games that humans play when they are in small groups are the same as the games that other primates play. Like other primates, humans engage in strategic interactions and attempt to press one another's prosocial buttons. They use physical, material, and social rewards and punishments to induce others to treat them right. In Darwin's (1874) words, "man [is] influenced in the highest degree by the wishes, approbation, and blame of his fellow-men, as expressed by their gestures and language" (p. 106).

The Origin of a Sense of Justice

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). According to Trivers (2006), "such cheating is expected to generate strong emotional reactions, because unfair arrangements, repeated often, may exact a very strong cost in inclusive fitness" (p. 77). In a similar vein, de Waal and Brosnan (2006) have suggested that "the squaring of accounts in the negative domain . . . may represent a precursor to

human justice, since justice can be viewed as a transformation of the urge for revenge, euphemized as retribution, in order to control and regulate behavior" (p. 88). On the positive side, "the memory of a received service, such as grooming, induces a positive attitude toward the same individual, a psychological mechanism described as 'gratitude' by Trivers (1971)" (p. 93).

The affective precursors to a sense of justice discussed by Trivers and de Waal stem primarily from the reactions of animals to the ways in which *they* are *treated* by members of their groups. There is, however, more to humans' sense of justice than these reactions. Humans also experience emotional reactions to the ways in which they and others treat third parties.

Affective Reactions to Third-Party Injustice

Although other primates display negative reactions to members of their troupes who violate prosocial norms and take measures to punish them (Boehm, 2000), humans may be the only species that is disposed to punish free riders and those who behave unfairly toward third parties. Summarizing the findings from several studies, Gachter and Herrmann (2006) conclude:

> Overall, the results suggest that free riding causes negative emotions . . . [that are] consistent with the hypothesis that emotions trigger punishment. . . . [T]he majority of punishments are executed by above-average contributors and imposed on below-average contributors. . . . [P]unishment increases with the deviation of the free rider from other members' average contribution. . . . [E]vidence from neuroscientific experiments supports the interpretation that emotions trigger punishment. (p. 297)

Although evolutionary theorists agree that humans are disposed to punish third-party cheaters, they do not agree about how the mechanisms that give rise to these dispositions evolved. On one side, mainstream evolutionary theorists argue that the disposition to punish free riders evolved through standard forms of selection (kin selection, reciprocal altruism, indirect reciprocity, and costly signaling). For example, Trivers (2006) has suggested that because the groups formed by early humans consisted mainly of kin, we would expect the mechanisms that dispose contemporary humans to punish third parties to "misfire" by being activated by members of groups who are not kin. Trivers's account implies that "the human brain applies ancient cooperative heuristics even in modern environments" (Gachter & Herrmann, 2005). Other mainstream evolutionary theorists such as Alexander (1987) and Nowak and Sigmund (1998) have argued that the disposition to punish third parties could have been reinforced by the fitness-enhancing gains of an enhanced social image or a reputation for cooperation. On the other side, theorists such as Fehr and Gächter (2002) and Gintis, Bowles, Boyd, and Fehr (2003) have argued that biological evolution is not, by itself, equipped to account for the disposition to punish free riders in one-shot games among anonymous players and that this disposition could have evolved only through gene-culture coevolution. The theoretical differences between theorists who

have advanced exclusively individual-level selection models and theorists who have advanced coevolutionary models are significant psychologically mainly with respect to their potential to produce hypotheses about how the mechanisms in question are designed.

Affective Reactions to Treating Others Fairly and Unfairly

When we attribute a sense of justice to people, we imply that they possess standards of fairness that they apply to themselves as well as to others. If the function of negative reactions to unfair behaviors committed by others is to motivate people to uphold systems of cooperation by punishing cheaters, we might also expect people to feel bad when they cheat and to be inclined to punish themselves. In fact, people often do feel bad when they cheat others, but it is unclear whether such negative reactions stem from the same mechanisms as their reactions to the transgressions of others.

There is an important difference between inducing oneself to cooperate and inducing others to cooperate. As discussed, in most contexts people are able to maximize their benefits by inducing others to do their share, or more than their share, while doing less than their share themselves. From an adaptive perspective, we would not expect people to be unconditionally motivated to behave fairly or to be naturally inclined to pass judgment on themselves in an impartial way. Rather, we would expect the mechanisms that guide decisions about fairness to be calibrated in ways that maximized the genetic benefits to early humans, inducing individuals to feel inclined to behave only as fairly as they needed to maximize their benefits from social exchanges. In support of these expectations, there is a great deal of evidence that people are inclined to react more strongly to being treated unfairly by others than to treating others unfairly, to hold others to higher standard of fairness than they hold themselves, and to reckon costs and benefits for themselves and others in different ways (Greenberg & Cohen, 1982). As expressed by Trivers (2006), "[A]n attachment to fairness or justice is self-interested and we repeatedly see in life . . . that victims of injustice feel the pain more strongly than do disinterested bystanders and far more than do the perpetrators" (p. 77).

People's negative reactions to others' injustices usually involve anger, which seems to emerge automatically. In contrast, people's negative reactions to their own injustices may be acquired more indirectly, through social learning. As emphasized by Darwin (1874), humans are highly motivated to seek the approval and avoid the disapproval of members of their groups. Contemporary learning theorists such as Aronfreed (1968) have adduced evidence that children acquire negative reactions to their own transgressions through classical and instrumental conditioning. Children come to feel good about behaving fairly and bad about cheating because they are rewarded when they behave fairly and punished when they behave unfairly. Although such inputs structure children's early conceptions of right and wrong, they do not account for a fully developed sense of justice, as I will explain.

The Expansion and Refinement of Cooperative Systems in the Human Species

Humans engage in concrete forms of reciprocity and feel angry when others cheat them in much the same way as chimpanzees do. However, *in addition*, humans engage in more complex forms of social exchange. They give to others over long periods of time before receiving any returns; they invest in long-term relationships; they trade across widely diverse domains (often using money as a common medium); they reckon equity in highly refined ways; they engage in indirect forms of reciprocity; they create rules and formalize systems of sanctions that uphold cooperative systems; they coordinate their efforts on a massive scale to accomplish such tasks such as constructing skyscrapers and building bridges.

The unique forms of cooperation practiced by modern humans became possible when early humans acquired the intellectual and linguistic abilities necessary to create them and uphold them. As expressed by Williams (1989), "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). Coupled with intelligence, symbolic language would have enabled early humans to translate their affective reactions to the behavior of members of their groups into words and communicate such reactions to those who performed the behaviors and to third parties. Not only would it have enabled them to express their immediate approval and disapproval with words such as "good" and "bad," but it would also have enabled them to pass judgment on events that occurred in the past and to make judgments about events that could occur in the future. It would have enabled them to transform primitive threats and promises into long-term social contracts and commitments (Nesse, 2001) and to verbalize disapproval when others violated implicit social contracts such as those that govern monogamous marriages. It would have enabled them to enhance or diminish others' reputations through gossip (Alexander, 1987; Dunbar, 1996) and to buttress their judgments with reasons, explanations, and justifications designed to increase their persuasive power.

The Expansion and Refinement of a Sense of Justice

Intelligence and language are two-edged swords. On the one hand, they enable humans to create and uphold significantly more complex forms of cooperation than those practiced by any other species. On the other hand, they enable humans to engage in significantly more complex forms of cheating. Although we would expect people to be naturally inclined to make self-serving moral judgments, the process of strategic interaction is equipped to counteract such biases. Because recipients are unreceptive to judgments that exhort them to behave in ways that do not advance

their interests, blatantly self-serving judgments do not work, and because they do not work, people are disinclined to make them.

Moral Argumentation

Language and intelligence endow humans with the capacity to resolve their conflicts of interest through negotiation and discussion. Many theorists have focused on the significance of moral argumentation in the production of standards of justice (e.g., Damon & Hart, 1992; Habermas, 1993; Piaget, 1932). When people engage in moral argumentation, they may attempt to push one another's emotional buttons (Haidt, 2001), or they may appeal to one another's rational faculties (Saltzstein & Kasachkoff, 2004). As explained by the philosopher Singer (1981), publicly expressed rational arguments tend to generate universal and impartial standards: "[I]f 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 hoist on their own petards.

The Evolution of Rules and Justice Norms

The process of strategic interaction and the adaptive value of resolving conflicts of interest through moral argumentation have implications for the evolution of rules of conduct and universal norms of justice. Members of groups make rules to formalize their agreements about how they should be treated by others, and they invoke sanctions to induce others to uphold the rules. What goes around comes around (Alexander, 1987), such that the rules that members of groups invent to control the behavior of others end up controlling their behavior. Inasmuch as recipients are more receptive to some moral prescriptions than to others, recipients serve as agents of selection, determining which prescriptions succeed, get repeated, and develop into rules and moral norms. We would expect people to be particularly receptive to moral judgments and rules that prescribe fitness-enhancing forms of cooperation and to judgments that enable them to resolve conflicts of interest in mutually beneficial ways. Consistent with these expectations, there is evidence that judgments and rules that uphold fair, balanced, and reversible solutions to social conflicts—such as those prescribed by the norm of reciprocity and the Golden Rule—constitute universal moral norms (Brown, 1991; Gouldner, 1960; Sober & Wilson, 1998; Wright, 1994).

Clearly, however, not all moral rules and norms are fair or rational. 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)

Earlier I discussed differences between negative affective reactions to injustices committed by others and negative reactions to injustices committed by oneself. In the same vein, there is a significant difference between believing that others should uphold standards of justice and believing that one is obliged to uphold them. People could espouse norms of justice in order to manipulate others into behaving in cooperative ways without believing in the norms or incorporating them into their own conceptions of justice. We would, however, expect the process of strategic interaction to reduce the gap between conceptions of one's own and others' rights and duties.

To begin with, as emphasized by socialization theorists, people may be persuaded to accept as valid the standards preached by others. Evolutionary theory offers a basis for predicting which, of the many ideas to which people are exposed, they will be disposed to accept. It leads us to expect people to be most receptive to norms and standards that have enhanced their fitness in the past and that they believe will enhance their fitness in the future. Thus, for example, we would expect people to be receptive to standards preached by those with a vested interest in their welfare and to standards that are widely accepted by other members of their groups (see Richerson & Boyd, 2005, for a more extended discussion of this issue).

In addition, preaching standards of justice to others may induce those who preach them to accept them as their own. Believing in the validity of the prescriptive judgments one makes may reap adaptive benefits by increasing their persuasive power (Trivers, 1985). 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 them to advance their own interests in optimal ways.

The Origin of Conscience

Most people locate their sense of justice in a mental mechanism they call their conscience. The conditioned reactions to one's transgressions discussed earlier may form the core of conscience. Animals such as dogs appear to display affective reactions akin to guilt when they anticipate punishment for their transgressions (Aronfreed, 1968). Humans differ from other animals, however, in their ability to construct portable cognitive representations of others and store them in their minds, to view events from others' perspectives, and to imagine how others will respond to their behavior (Selman, 1980). In their imagination, people experience others as observing them when they are in private and passing judgment on their behavior (Aronfreed, 1968; Higgins, 1987).

Ironically, perhaps, the mechanisms that enable people to take the perspective of others may have evolved as tools designed to improve early humans' ability to manipulate others in the context of strategic interactions. There is tremendous adaptive potential in the ability to anticipate the moves of others in social games—what they are thinking; what they intend to do; whether they will cooperate, pay one back, detect one's deception; and so on. To accomplish this, people internalize mental representations of others and view events, including those which they themselves are directly involved in, from their perspectives. After people internalize mental images of others, they may experience these images as approving and disapproving of the things they do in private, and this may be experienced as a "voice of conscience."

As children's perspective-taking abilities develop, their cognitive representations of others become increasingly abstract, integrated, and general (Selman, 1980). As expressed by 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). We would expect highly developed perspective-taking processes to give rise to fairer decisions than more primitive perspective-taking processes.

To summarize, an evolutionary analysis suggests that conscience is a mental mechanism that originated as a tool in strategic interaction. Conscience consists of internalized images of others that enable people to predict how others will react to their behaviors. In imagining the negative reactions of others, people experience an anticipatory fear or embarrassment, which they experience as a sense of guilt or shame. As people internalize an increasingly large number of cognitive representations and as they integrate them in their minds, the perspective from which they judge themselves becomes increasingly abstract and impartial.

Reframing Traditional Psychological Accounts of the Acquisition of a Sense of Justice

An evolutionary framework supplies a basis for reconceptualizing psychological models of the acquisition of a sense of justice in ways that integrate their insights and redress their limitations. The family contexts in which parents teach children to behave fairly are microcosms of larger social groups. Members of families face 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. However, it may be in each member's interest to favor himself or herself and those with whom he or she shares 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. The ways in which members of families resolve their conflicts of interest affect the ways in which their conceptions of justice are structured and calibrated.

Evolutionary theory leads us to expect the mechanisms that regulate strategic interactions between parents and children to be designed in fitness-enhancing ways. It follows that we would not expect children to conform to their parents' injunctions indiscriminately or docilely. We would expect children to resist injunctions that run contrary to their interests and actively attempt to manipulate and control other members of their families. 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 an evolutionary framework that emphasizes the role of strategic interaction in the development of a sense of justice. From this perspective, the key to instilling a balanced sense of justice in children lies in structuring their early interactions in fair ways and inducing them to discover by their experience that it pays to cooperate and treat others fairly.

An evolutionary analysis implies a different interpretation from that offered by cognitive-developmental theorists of evidence that children acquire increasingly sophisticated structures of justice reasoning as they develop. The anthropologist 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—(1) affectionate relations among people who share social bonds, (2) hierarchical relations among people who differ in social rank, (3) egalitarian exchanges among equals, and (4) economic relations aimed at maximizing cost/benefit ratios across different commodities. Chimpanzees possess the first three schemata; the fourth appears to be unique to the human species (de Waal, 1996; Haslam, 1997).

Life history theory implies that the reason people are prone to invoke increasingly sophisticated schemata and structures of moral reasoning as they develop is that they need increasingly sophisticated schemata and standards of justice to solve the increasingly complex and embedded social problems they encounter as they progress through the life span. The reason young children view justice primarily in terms of obedience to authority (Kohlberg, 1984) is that it is adaptive for young children to subordinate themselves to older, wiser, and more powerful members of their groups. The reason older children view justice primarily in terms of concrete reciprocity is that reciprocity is a more adaptive strategy than obedience in egalitarian relations among peers (Piaget, 1932). The reason young adults view justice primarily in terms of principles that uphold long-term commitments, harmonious in-group relations, and systems of indirect reciprocity is that these forms of cooperation are best equipped to foster their interests (see Krebs, 2005a, 2005b, for elaborations of these ideas). The sophisticated forms of justice reasoning that define Kohlberg's highest stages of moral development constitute creative ideas about how to resolve conflicts of interest and reap the benefits of cooperation in optimal ways. From an evolutionary perspective, cardinal moral principles such as "foster the greatest good for the greatest number" equate to injunctions to foster one's ultimate adaptive interests by upholding the standards, forms of conduct, and systems of cooperation that, if adopted by everyone, would produce the greatest gains.

From a life history perspective, we would not expect new structures of justice reasoning to "transform and displace" older structures, as Colby and Kohlberg (1987) have hypothesized. We would expect people to acquire structures of justice reasoning in an "additive-inclusive" way (Eisenberg, 1982; Levine, 1979), because adults continue to experience the kinds of adaptive problems that early structures evolved to solve. Adults may, for example, find themselves in subordinate positions in which it would be adaptive for them to believe that they should show deference to authority (Milgram, 1974). Viewed in this manner, the acquisition of a sense of justice 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). 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 may stem from essentially the same affective and cognitive processes.

The Activation of Mechanisms that Produce a Sense of Justice

Given a suite of evolved mechanisms equipped to contribute to people's sense of justice, the main task for those who seek to account for this phenomenon is to explain how these mechanisms are activated and, if more than one is activated, how they interact. Because complex forms of moral cognition are more costly than simpler forms, we would expect people to be inclined to use simple, automatic forms as their default (Gigerenzer, 2000; Gilovich, Griffin, & Kahneman, 2002). We would expect affective reactions such as gratitude and righteous indignation to exert an immediate effect on people's sense of justice (Haidt, 2001; Sunstein, 2005), and we would not be surprised that people have difficulty justifying decisions derived in these ways or, if called upon to justify them, that they offer plausible but invalid post hoc rationalizations (Haidt, 2001).

We also would expect people to invoke simple forms of justice reasoning to solve simple, recurring social problems (Fiske, 1992), to make quick decisions in contexts in which the costs of deliberation are high, and to generate simple judgments when such judgments constitute the most effective forms of persuasion and impression management (such as, for example, when they are directed toward children) (Krebs & Janicki, 2004). We would expect people to adopt and to preach the moral norms of their cultures without thinking much about them, as long as they worked reasonably well, and to use mental shortcuts in contexts in which heuristics generate acceptable moral decisions (Chaiken, 1987; Gigerenzer, 2000; Sunstein, 2005).

We would expect conceptions of justice to be customized to solve different kinds of social problems and, therefore, for people to invoke different conceptions of justice in different domains, contexts, and conditions (Damon, 1980; Eisenberg, 1982;

Krebs & Denton, 2005; Krebs, Vermeulen, Carpendale, & Denton, 1991). We would expect the cognitive apparatus that gives rise to conceptions of justice to be susceptible to framing, directional, motivational, self-serving, nepotistic, and group-serving biases (Chaiken, Giner-Sorolla, & Chen, 1996; Krebs & Laird, 1998; Kunda, 2000; Pyszczynski & Greenberg, 1987; Richerson & Boyd, 2005). And we would not be surprised to find that people sometimes use justice reasoning for immoral purposes, such as avoiding responsibility and justifying immoral acts (Bandura, 1991; Haidt, 2001).

There is nothing in this evolutionary analysis of the acquisition of a sense of justice that is inconsistent with the idea that people have the capacity to derive conceptions of justice from sophisticated forms of moral reasoning. As demonstrated by cognitive-developmental theorists, most people do possess this capacity. However, an evolutionary framework induces us to ask how often, and in what contexts, people invoke this tool rather than other tools in their moral-decision-making tool boxes. We would expect people to invoke sophisticated forms of moral reasoning to derive decisions about justice when they work better than alternative methods and when the biological benefits from invoking them outweigh the costs. For example, we would expect people to invoke sophisticated forms of moral reasoning to resolve conflicts among moral intuitions and moral norms (Haidt, 2001) and the rights and duties of people participating in embedded systems of cooperation (Kohlberg, 1984). We would expect people to 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 the costs of deliberation are low, when the benefits of deliberation are high, 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 assess moral reasoning.

Conclusion

To understand how people acquire a sense of justice, we must understand why people need one and what goals it helps them to achieve. The mechanisms that give rise to a sense of justice evolved to help early humans maximize their gains from cooperative social interactions. A sense of justice induces members of groups to distribute resources in fair ways (distributive justice), to honor the commitments they make to others (commutative justice), to punish cheaters (corrective justice), and to develop effective ways of resolving conflicts of interest and making fair decisions (procedural justice).

Contemporary humans inherit primitive predispositions to react positively to being treated fairly and negatively to being treated unfairly, to pass judgment on those who treat others fairly or unfairly, and to feel obliged to pay others back. This core is refined and expanded during the process of strategic interaction in every generation

as people reward and punish one another for behaving in cooperative and uncooperative ways, preach norms of fairness, negotiate mutually beneficial solutions to their conflicts of interest, and attempt to create ever more effective systems of cooperation. To achieve these goals, people use the tools with which they have been endowed by natural selection, especially language, perspective-taking abilities, and social intelligence. Although it is naïve to expect people to possess a universal sense of justice that consistently disposes them to make fair and impartial decisions that jeopardize their adaptive interests, it is realistic to expect people to be able to counteract one another's biases in ways that enable them to make fair decisions in contexts in which such decisions advance everyone's interests in optimal ways.

References

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

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

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

Boehm, C. (2000). Conflict and the evolution of social control. *Journal of Consciousness Studies, 7,* 79–101.

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

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.

Colby, A., & Kohlberg, L. (Eds.). (1987). *The measurement of moral judgment* (Vols. 1–2). Cambridge: Cambridge University Press.

Gachter, S., & Herrmann, B. (2006). Human cooperation from an economic standpoint. In P. M. Kappeler & C. P. van Schaik (Eds.), *Cooperation in primates and humans: Mechanisms and evolution* (pp. 275–302). Berlin: Springer-Verlag.

Damon, W. (1980). Patterns of change in children's social reasoning: A two-year longitudinal study. *Child Development, 46,* 1010–1017.

Damon, W., & Hart, D. (1992). Self-understanding and its role in social and moral development. In M. H. Bornstein & E. M. Lamb (Eds.), *Developmental psychology: An advanced textbook* (2nd ed., pp. 421–465). Hillsdale, NJ: Erlbaum.

Darwin, C. (1874). *The descent of man and selection in relation to sex*. New York: Rand McNally.

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.

de Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

de Waal, F. B. M., & Brosnan, S. F. (2006). Simple and complex reciprocity in primates. In P. M. Kappeler & C. P. van Schaik (Eds.), *Cooperation in primates and humans: Mechanisms and evolution* (pp. 85–106). Berlin: Springer-Verlag.

De Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of reciprocity in three primate species: Symmetrical relationship characteristics or reciprocity? *Ethology and Sociobiology, 9,* 101–118.

Dunbar, R. I. M. (1996). Determinants of group size in primates: A general model. In G. Runciman, J. Maynard Smith, & R. I. M. Dunbar (Eds.), *Evolution of social behavior patterns in primates and man* (pp. 33–58). Oxford: Oxford University Press.

Eisenberg, N. (1982). *The development of prosocial behavior*. New York: Academic Press.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415,* 137–140.

Festinger, L. (1964). *Conflict, decision, and dissonance*. Stanford, CA: Stanford University Press.

Fiske, A. P. (1992). Four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review, 99,* 689–723.

Gachter, S., & Herrmann, B. (2006). Human cooperation from an economic perspective. In P. M. Kappeler & C. P. van Schaik (Eds.), *Cooperation in primates and humans: Mechanisms and evolution* (pp. 275–301). Berlin: Springer-Verlag.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world.* New York: Oxford University Press.

Gilovich, T., Griffen, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

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

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

Greenberg, J., & Cohen, R. L. (1982). *Equity and justice in social behavior*. New York: Academic Press.

Habermas, J. (1993). *Justification and application*. Cambridge, MA: MIT Press.

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

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.

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

Kappeler, P. M., & van Schaik, C. P. (2006). *Cooperation in primates and humans: Mechanisms and evolution*. Berlin: Springer-Verlag.

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

Kochanska, G., & Aksan, N. (2004). Conscience in childhood: Past, present, and future. *Merrill-Palmer Quarterly, 50,* 299–310.

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.

Krebs, D. L. (2005b). The evolution of morality. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 747–771). Hoboken, NJ: John Wiley.

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., & 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 and J. Gewirtz (Eds.), *Handbook of moral behavior and development: Theory, research, and application (pp. 139–169)*. Hillsdale, NJ: Erlbaum.

Kunda, Z. (2000). *Social cognition: Making sense of people*. Cambridge MA: MIT Press.

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.

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

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.

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

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. *Advances in Experimental Social Psychology, 20,* 297–340.

Rawls, J. (1999). *A theory of justice* (rev. ed.). Cambridge, MA: Harvard University Press.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Saltzstein, H. D., & Kasachkoff, T. (2004). Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology, 8,* 273–282.

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

Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. New York: Farrar, Straus and Giroux.

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

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences, 28,* 531–573.

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-deception. 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. Kappeler & C. P. van Schaik (Eds.), *Cooperation in primates and humans: Mechanisms and evolution* (pp. 67–84). Berlin: Springer-Verlag.

Williams, G. C. (1989). A sociobiological expansion of *Evolution and Ethics*. In J. Paradis & G. Williams (Eds.), *Evolution and Ethics* (pp. 179–214). Princeton, NJ: Princeton University Press.

Wilson, J. Q. (1993). *The moral sense*. New York: Free Press.

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