Psychological Bulletin
2013, Vol. 139, No. 2, 477–496

0033-2909/13/$12.00 DOI: 10.1037/a0029065

A Solution to the Mysteries of Morality

Peter DeScioli
Brandeis University

Robert Kurzban
University of Pennsylvania and Chapman University

We propose that moral condemnation functions to guide bystanders to choose the same side as other bystanders in disputes. Humans interact in dense social networks, and this poses a problem for bystanders when conflicts arise: which side, if any, to support. Choosing sides is a difficult strategic problem because the outcome of a conflict critically depends on which side other bystanders support. One strategy is siding with the higher status disputant, which can allow bystanders to coordinate with one another to take the same side, reducing fighting costs. However, this strategy carries the cost of empowering high-status individuals to exploit others. A second possible strategy is choosing sides based on preexisting relationships. This strategy balances power but carries another cost: Bystanders choose different sides, and this discoordination causes escalated conflicts and high fighting costs. We propose that moral cognition is designed to manage both of these problems by implementing a dynamic coordination strategy in which bystanders coordinate side-taking based on a public signal derived from disputants' actions rather than their identities. By focusing on disputants' actions, bystanders can dynamically change which individuals they support across different disputes, simultaneously solving the problems of coordination and exploitation. We apply these ideas to explain a variety of otherwise mysterious moral phenomena.

Keywords: moral psychology, evolution of morality, condemnation, choosing sides, side-taking

People can use morality to choose sides in conflicts. When two people get into a dispute, their family, friends, and other bystanders often get involved and take a side. Sometimes bystanders choose sides based on relationships, such as supporting a sibling over a stranger or supporting a friend over an acquaintance. Sometimes, however, people choose sides based on moral considerations, who is "right" and who is "wrong." Bystanders might side against a sibling who wrongfully stole money from a stranger, or they might side against a friend who lied to an acquaintance. In these cases, the moral status of particular actions, such as theft or deception, influences how people choose sides. We propose that moral condemnation functions to guide bystanders to choose the same side as other bystanders in conflicts, and this function uniquely explains a wide range of empirical observations about the information-processing structure of moral cognition.

Our proposal builds on previous work outlining the challenges posed by several core mysteries of morality (DeScioli & Kurzban, 2009b). This framework distinguishes two basic issues to be addressed. One issue is *moral conscience*, or why people conform to moral rules. This question requires explaining why people avoid certain behaviors such as incest or assault. A second, distinct issue is *moral condemnation*, or why people judge other people's actions to be "wrong." This question requires explaining the empirical pattern of moral judgments including why people want violators to be punished for specific behaviors such as incest or assault (Chudek & Henrich, 2011; Darley & Schultz, 1990; Gray & Wegner, 2009; Gutierrez & Giner-Sorolla, 2007; Haidt, 2001, 2007; Haidt & Joseph, 2004, 2008; Kadri, 2005; Killen, 2007; Knobe, 2005; Levy, 1993; Lieberman, Tooby, & Cosmides, 2003, 2007; Mikhail, 2007; Shweder, Much, Mahapatra, & Park, 1997; Simoons, 1994; Smetana & Braeges, 1990; Tetlock, 2000, 2002; Turiel, 1998).

Previous evolutionary approaches to morality have focused on the first issue, moral conscience, especially the evolution of altruistic behavior (de Waal, 1996; Joyce, 2006; Ridley, 1996; Wright, 1994). In contrast, we focus on the second issue, moral condemnation. Beginning with condemnation is potentially productive because explaining moral condemnation naturally and simultaneously provides a framework for explaining features of moral conscience. In a social world in which other people condemn and punish actions of particular types—those specified by moral rules—avoiding these actions can be understood as a strategy for avoiding condemnation (e.g., DeScioli, Christner, & Kurzban, 2011).

Specifically, we argue that the cognitive mechanisms that underlie moral condemnation are designed around the problem of choosing sides in conflicts. Bystanders can coordinate to choose the same side and reduce fighting costs if they choose sides based on a public signal, following the game theoretic logic of a correlated equilibrium (Aumann, 1974). Furthermore, bystanders can dynamically change who they support by using as a source of public signals the disputants' actions (what actors have done)

This article was published Online First July 2, 2012.

Peter DeScioli, Department of Psychology, Brandeis University; Robert Kurzban, Department of Psychology, University of Pennsylvania, and Economic Science Institute, Chapman University.

We would like to thank Jon Haidt, Debra Lieberman, Michael Bang Petersen, David Pinsof, Alex Shaw, and Chris Taylor for thorough and thoughtful comments on drafts of this article. We thank Ilya Altshteyn, John Christner, Sarah Hailey, Omar Sultan Haque, Maxim Massenkoff, Steven Pinker, Sophie Scolnik-Brower, Kyle Thomas, and Leslie Zebrowitz for insightful discussion about ideas in this article.

Correspondence concerning this article should be addressed to Peter DeScioli, Department of Psychology, Brandeis University, Waltham, MA 02453. E-mail: pdescioli@gmail.com

rather than their identities (who the actors are). This strategy, *dynamic coordination*, can allow bystanders to coordinate with each other without concentrating power in particular individuals. We propose that moral cognition is designed to implement this dynamic coordination strategy.

Empirical Challenges for Understanding Moral Condemnation

Evolutionary theories of morality have focused on explanations for altruism (Alexander, 1987; de Waal, 1996; Gigerenzer, 2010; Greene, 2008; Haidt, 2007; Hauser, 2006; Wright, 1994). These models can potentially explain why people have cognitive systems for benefiting other people. However, they do not readily explain moral condemnation—why people think other people should be punished for violating moral rules. A theory of moral condemnation needs to explain its empirically observed properties, including three central features: moral judgment, moralistic punishment, and moral impartiality. We summarize these issues here (for more detailed discussion, see DeScioli & Kurzban, 2009b).

Historically, morality and altruism have been very closely linked in the scientific literature. Darwin (1871) approached the human "moral faculties" with the idea that "to do good unto others . . . is the foundation-stone of morality" (p. 159). He claimed that groups with individuals who were willing to "sacrifice themselves for the common good" gained an advantage in between-group competition. Darwin, then, took the task of explaining morality to be the same as explaining altruism, why people sacrifice (enduring costs) to do good for others (delivering benefits). Modern theorists have largely adopted Darwin's approach, while exploring other evolutionary pathways to altruism such as kin selection and reciprocal altruism. Richard Alexander (1987), in *The Biology of Moral Systems*, wrote: "The problem, in developing a theory of moral systems that is consistent with evolutionary theory from biology, is in accounting for the altruism of moral behavior in genetically selfish terms" (p. 93). More recently, Haidt (2007) wrote that "people are selfish, yet morally motivated" (p. 998), implying that morality is the opposite of selfishness. Many other researchers similarly take the perspective that explaining morality is the same as explaining altruism (de Waal, 1996; Gigerenzer, 2010; Greene, 2008; Hauser, 2006; Wright, 1994).

Altruism theories can potentially explain why people choose actions that benefit other people, but they do not straightforwardly explain why people morally judge other people's behavior. People show intense interest in moral wrongdoing (Dunbar, 2004; Wiessner, 2005) beginning in childhood (Darley & Schultz, 1990; Ross & Den Bak-Lammers, 1998; Turiel, 1998). Moral judgment involves complex unconscious inferences (Mikhail, 2007), and does not reflect only computations about other people's welfare (Gutierrez & Giner-Sorolla, 2007; Haidt, 2001; Haidt & Hersh, 2001; Haidt, Koller, & Dias, 1993; Tetlock, 2000). In fact, moral judgment can be insensitive to expected consequences, showing patterns of "nonconsequentialism," due to its particular focus on wrongful actions rather than outcomes (see Categorical Imperatives). In identifying violating actions, moral judgment is sensitive to a number of factors such as harm, intent, knowledge, causality, and force, but the reverse can also occur, with beliefs about these factors being generated post hoc to support moral conclusions (Alicke, 1992; Haidt, 2001; Knobe, 2005; Young & Phillips, 2011).

Darwin (1871) himself discussed a formidable problem for altruism theories: the incredible variety of moral rules, especially destructive prohibitions. Moral judgment is applied to a variety of content domains, including violence, altruism, property, authority, sex, food, communication, and many others (Haidt, 2007; Haidt & Graham, 2007; Haidt & Joseph, 2004, 2008; Levy, 1993; Shweder et al., 1997; Simoons, 1994). Nonetheless, people are able to compare severity across domains (Robinson & Kurzban, 2007). There is striking cross-cultural variation in moral rules such that behaviors viewed as highly immoral—and punishable—in one group are accepted, even promoted, in other groups (e.g., contraception, Riddle, 1997; interest-bearing loans, Bentham, 1787/1952). At the same time, people show intolerance of moral variation (Haidt, Rosenberg, & Hom, 2003), holding that their own moral judgments are universal rather than contingent on factors such as personal taste, group membership, or the pronouncements of authorities (Goodwin & Darley, 2008; Killen, 2007; Posada & Wainryb, 2008; Smetana & Braeges, 1990; Smetana, Schlagman, & Adams, 1993; Tooby & Cosmides, 2010; Turiel, 1998), leading to vigorous moral debates aimed at achieving consensus. Finally, moral judgment can be highly destructive (Darwin, 1871), particularly when moral rules prohibit harmless or beneficial behavior. Modern examples include "honor killings" of women (Appiah, 2010; United Nations, 2000) and organized militias that execute homosexual people (Sarhan & Burke, 2009).

Next, altruism theories do not easily explain people's moralistic punishment. Moralistic punishment is puzzling because it is costly for the punisher (Boyd, Gintis, Bowles, & Richerson, 2003; Boyd & Richerson, 1992; Dreber, Rand, Fudenberg, & Nowak, 2008; Gardner & West, 2004; Henrich & Boyd, 2001; Gintis, Smith, & Bowles, 2001; Sigmund, 2007). When people are punished, they might seek revenge (McCullough, Kurzban, & Tabak, 2010), and punishers risk being the target of retaliation. People are sensitive to these costs: When the ability to retaliate is included in economic games, punishment drops substantially (Cinyabuguma, Page, & Putterman, 2006; Denant-Boemont, Masclet, & Noussair, 2007; Herrmann, Thoni, & Gachter, 2008). Nonetheless, whether people are willing to perform costly punishment themselves (Kurzban & DeScioli, 2009), it is clear that people want violators to be punished (Robinson, Kurzban, & Jones, 2007). Some researchers have argued that costly punishment promotes group welfare (Boyd, Gintis, Bowles, & Richerson, 2003), but other researchers have found that punishment reduces group welfare (Dreber et al., 2008; Herrmann et al., 2008; Sigmund, 2007) and that punishment behavior responds to variables relevant to individual reputation rather than group welfare (Barclay, 2006; Kurzban, DeScioli, & O'Brien, 2007).

Finally, altruism theories do not explain why people often claim that their moral judgments are impartial, independent of their loyalties to the people being judged. Humans try to appear impartial in moral judgments, and at times they actually show some degree of impartiality (DeScioli & Kurzban, 2009b; Lieberman & Linke, 2007; Tetlock, 2002). Of course, people are also frequently partial (e.g., Bernhard, Fischbacher, & Fehr, 2006; van Prooijen, 2006), but it is not clear why the ideal of moral impartiality exists at all. Evolutionary processes that cause altruism such as kin selection (Hamilton, 1964) or reciprocity (Trivers, 1971)—often

invoked to explain morality—are expected to produce partial mechanisms that discriminate in favor of family, friends, or groups. Motivations to appear impartial need to oppose these evolved altruism mechanisms in order to maintain a reputation for impartiality.

These empirical observations are not well explained by existing theoretical perspectives based on altruism. One particularly challenging issue is the focus of moral judgment on actions, which we now turn to in greater detail.

Categorical Imperatives

Kant (1785/1993) famously argued that lying is always morally wrong, no matter what benefits result, even saving lives. More generally, he argued that morality consists of "categorical imperatives," a set of actions that are morally wrong regardless of the goals these actions are intended to achieve. Laboratory research has found that people often show Kantian moral thinking, focusing on specific actions rather than expected consequences (reviewed in DeScioli & Kurzban, 2009b). For instance, in the footbridge trolley dilemma, 90% of people judged that it is impermissible to push one person off of a footbridge to save five people (Mikhail, 2007).

These observations pose a problem for altruism theories that hold that morality is for improving welfare. The biologist Mivart (1871) recognized this problem and argued that Darwin, like John Stuart Mill, did not appreciate the "difference between the ideas 'useful' and 'right'" (p. 203), where the latter refers to avoiding wrongful actions rather than maximizing welfare. Mivart rejected Darwin's (1871) theory that human morality was an adaptation for altruism because a mechanism for altruism would be expected to be welfare-maximizing, attuned to expected consequences, not actions.

The problem posed by categorical imperatives can be expressed formally with decision theory. When making a choice, a decision maker can use utility maximization:

$$\max_{a \in A} u(\mathbf{y}), \tag{1}$$

where the decision maker chooses the action a^* from the choice set A to maximize $u(\mathbf{y})$, which depends only on the consequences, the vector $\mathbf{y}$ of payoffs to the decision maker and other people in the situation. By applying different weights to one's own and other people's payoffs, this decision procedure can encompass dispositions ranging from extreme selfishness to extreme altruism. Importantly, choices are based solely on the payoffs resulting from the actions rather than on the actions themselves.

In contrast, Kantian decisions can be expressed as choosing subject to moral constraints on the actions:

$$\max_{a \in A} u(\mathbf{y}), \text{ subject to the constraint, } a \notin W, \tag{2}$$

where W refers to a set of actions labeled morally wrong. In Kantian decisions, morally wrong actions are excluded regardless of the payoffs they generate. In philosophical terms, these decisions are *nonconsequentialist* in that judgments of behavior are not based only on the expected consequences.

Research shows that moral decisions are influenced by both action constraints and consequences. As mentioned above, in the footbridge trolley problem, most people said it was impermissible to kill one person to save five people, but in the switch version of the problem, most people judged that it was permissible to redirect the trolley onto a new path that will kill one person in order to save five people (Mikhail, 2007). This observation is consistent with other research indicating that when violations occur as a by-product rather than as a means, the moral action constraints are less binding, perhaps due to how byproducts are encoded in action representations (Mikhail, 2007; Royzman & Baron, 2002; Waldmann & Dieterich, 2007). Hence, both action constraints and consequences influence moral judgments, but it is not clear precisely how these factors are integrated in moral cognition.

The presence of nonconsequentialism in moral judgment—specifically, the use of action constraints—strongly suggests that the function of the system giving rise to these judgments is not to deliver benefits (DeScioli & Kurzban, 2009b), contradicting the common view that morality is designed for altruism (de Waal, 1996; Joyce, 2006; Ridley, 1996; Wright, 1994). The action constraint, W, pertains to the means by which benefits are achieved, whereas a benefit-delivery system would be focused on the goal rather than the means of attaining the goal.

Indeed, cognitive adaptations in general are expected to be consequentialist because evolution favors mechanisms that achieve better fitness consequences. Of course, consequentialist computations do not require conscious awareness of fitness goals, or (unattainable) complete information about all possible consequences. Instead, consequentialist reasoning requires using a form of Decision Rule 1 above in which the choice among alternative actions depends on estimated values for the consequences of each action. For example, burying beetles kill some of their offspring to feed the bodies to other offspring (Mock, 2004): These (presumably nonconscious) consequentialist computations maximize inclusive fitness while violating the human action constraint against infanticide. Indeed, many animal species frequently face trade-offs between the welfare of some individuals and the welfare of other individuals, and evolved psychological mechanisms usually base these decisions on fitness costs and benefits (Mock, 2004)—consequentialism—rather than using action constraints.

Some researchers have argued that moral action constraints function as altruism heuristics: The rules typically maximize welfare even if they fail in some cases (Gigerenzer, 2010). This idea resembles *rule consequentialism* in moral philosophy, which holds that simple rules are the best way to maximize welfare. There are several problems with the altruism-heuristic theory. First, the need for simplifying heuristics does not explain why moral rules focus on behavior per se. Rather than use action constraints, altruism heuristics could use simple cues to estimate welfare outcomes, for instance, using physical injury as a cue for utility, such as the heuristic rule "minimize total physical injuries" rather than the action constraint "do not kill." Indeed, classic heuristics such as representativeness use "attribute substitution," where a cue is substituted for an outcome, rather than action constraints (Kahneman & Frederick, 2002). Second, when participants are asked to assess both wrongness and welfare, their welfare judgments accurately track welfare outcomes (Haidt, 2001; Haidt & Hersh, 2001; Haidt et al., 1993; Tetlock, 2000), indicating that welfare information was easy to compute but ignored in wrongness judgments. Third, moral judgments are not simple but are intricate and complex (Mikhail, 2007), and further, their nuances track dimensions of perpetrators' actions rather than welfare outcomes. Finally, the

altruism-heuristic model predicts that increasing people's altruistic dispositions toward other people will lead to greater use of action constraints such as "do not kill," but instead the reverse occurs. Kurzban, DeScioli, and Fein (2012) found that participants reported greater willingness to kill one brother to save five brothers than to kill one stranger to save five strangers. Altruism causes people to be less likely, not more likely, to use Kantian action constraints.

Mysteries of Morality

The empirical observations reviewed above—third-party judgment, moralistic punishment, impartiality, moral variety, action constraints—pose difficult challenges for traditional evolutionary theories based on altruism. Below, we develop a functional theory of moral cognition based on side-taking rather than altruism. We first describe the adaptive problems people face when choosing sides in disputes and a potential solution to these problems that we term dynamic coordination. Then we examine whether moral cognition might perform a dynamic coordination function. We return to the challenging empirical observations from this section to see how these observations can be explained as part of a side-taking strategy. Next, we consider how the content of moral rules is shaped by people's agreements and disagreements about which actions will be prohibited. Finally, we offer a few examples of how the side-taking theory can be applied to illuminate otherwise mysterious moral phenomena.

The Problem of Choosing Sides

Like many other animal species, humans have disputes over resources, and these conflicts vary in intensity, ranging from minor disagreements to heated arguments to lethal violence (e.g., Daly & Wilson, 1988). Unlike most other animals, however, humans frequently recruit bystanders for support in disputes, expanding dyadic conflicts to larger ones. This ability creates a new set of adaptive problems unique to multiparty conflicts (Harcourt, 1992). In this section, we focus on two key adaptive problems faced by bystanders who choose sides: (a) discoordination with other bystanders and (b) exploitation resulting from excessive support and empowerment of a few individuals. Subsequently, we propose that moral cognition is designed to implement a particular strategy for solving these two problems.

In most animal species, individuals do not intervene in other individuals' conflicts (Harcourt, 1992). Disputes are dyadic, and group dominance hierarchies reflect individuals' relative fighting abilities (Krebs & Davies, 1993). In some species, individuals fight in groups such as armies of ants (Whitehouse & Jaffe, 1996), but group alignment is usually fixed (i.e., ants cannot switch sides). Even in primate species, coalitional alignment is usually determined by kinship rather than being flexible and unpredictable. In baboons, for instance, individuals always side with those in their matriline and do not switch sides (Seyfarth & Cheney, 2012).

There are several exceptions including chimpanzees, macaques, and dolphins. In chimpanzees, males switch alliance partners, sometimes over short periods of time. Newton-Fisher (2002) wrote that chimpanzees "show little long-term loyalty to one another and can be extremely fickle in their allegiances" (p. 125). This gives rise to complex social dynamics, so much so that chimpanzee interactions have been likened to "politics" (de Waal, 1982). This social environment has shaped adaptations designed to manipulate social relationships to gain power (Whiten & Byrne, 1997). Similar findings suggest that Assamese macaques have adaptations for managing shifting coalitions (Schülke, Bhagavatula, Vigilant, & Ostner, 2010). In bottlenose dolphins, males form small alliances that further combine into superalliances in disputes with rival coalitions over access to females (Connor, 2007). In species in which alliances can form, individual power is no longer sufficient to guarantee the apex of the dominance hierarchy because two smaller individuals can jointly depose a single, more powerful individual (Newton-Fisher, 2002).

Humans frequently intervene in other individuals' conflicts to provide coalitional support (Black, 1998; Cooney, 1998, 2003; Harcourt, 1992; Harcourt & de Waal, 1992; Phillips & Cooney, 2005). When bystander intervention is possible, individuals must defend themselves not only against other individuals but also against collections of individuals. Moreover, human coalitions are shifting (Kurzban, Tooby, & Cosmides, 2001), and bystanders' side-taking decisions are difficult to predict. This feature creates a complex strategic environment in which potential allies and enemies are difficult to discern.

Crucially, the possibility of intervention also creates a new strategic position—the bystander role—with the challenging problem of choosing sides. Now, individuals must try to reduce not only the costs of their own fights but also the costs of joining in other individuals' fights. Choosing sides is a critical decision because being on the losing side can have high fitness costs, even death in some cases (Tiger, 1969).

In this section, we examine the strategic predicament of bystanders who choose sides in conflicts. We note that by "conflicts" we do not mean only zero-sum games or violent conflicts, but rather we refer broadly to conflicts of interest, in a technical sense, including mixed-motive games characterized by both shared and opposed interests (Schelling, 1960). Many disputes occur in cooperative relationships in which interests are largely shared, but there is also room for disagreement. Further, many disputes are resolved with considerable subtlety and do not escalate to overt hostility. Even in these subdued cases, conflicts of interest have important consequences, and individuals can recruit other people to help promote their interests. In general, the subtlety of human conflict resolution is indicative of the potentially high costs of fighting and the cognitive abilities that humans have for avoiding escalation. The broad problems faced by bystanders discussed here apply across a range of conflicts varying in escalation from subtle to violent.

Third-Party Coordination

When a bystander chooses sides in a conflict, an important consideration is which side other bystanders will support. We refer to the bystanders or outsiders to a conflict as "third parties," distinguishing them from the two initial parties between whom the dispute began. We assume that third parties incur greater costs from being on the losing side than the winning side and, further, that numerical superiority provides an advantage. These two factors together give rise to an important adaptive problem: avoiding being on the minority side. Third parties need to anticipate which side the majority will take to avoid being outnumbered and suf-

fering a costly defeat. When all third parties seek to side with a majority, they collectively face a coordination problem that requires synchronizing their side-taking decisions.

Third parties also need to coordinate to avoid the high costs of fights between evenly matched sides. In the nonhuman animal literature, research shows that animals assess relative fighting ability and retreat if they determine that they are outmatched (Arnott & Elwood, 2009; Parker, 1974). The costliest fights occur when individuals are closely matched, requiring further escalation to decide the contest (Arnott & Elwood, 2009; Krebs & Davies, 1993). In multi-individual conflicts, by synchronizing side-taking, third parties can avoid long, escalated battles between evenly matched groups. When all third parties choose the same side, the fight is heavily lopsided and decided quickly with low costs to third parties. If, on the other hand, third parties split into closely matched groups, then they suffer costly protracted conflicts. The costs of discoordination—splitting between disputants—means that each player stands to gain by supporting the disputant whom other bystanders support. If third parties can synchronize their side-taking choices, then they can minimize the costs of becoming entangled in other people's fights.

Bandwagoning

One strategy third parties can use is *bandwagoning*, siding with the more powerful individual based on the relative power of the disputants and their respective supporters. Siding with the more powerful disputant prevents discoordination when used by all third parties. Hyenas show this pattern: Third parties choose sides based on a publicly known status hierarchy, siding with the higher ranked fighter (Engh, Siebert, Greenberg, & Holekamp, 2005). This strategy of supporting of the higher ranked individual is also found in nonhuman primates (Chapais, Girard, & Primi, 1991; Cheney & Seyfarth, 1990; Silk, Alberts, & Altmann, 2004): "Female vervets, macaques, and baboons typically support the higher ranking of the two opponents when forming alliances with lower ranking individuals" (Cheney & Seyfarth, 2007, p. 92). For the bandwagon strategy to work, like other coordination strategies, individuals need to establish "common knowledge" (Aumann, 1976; Schelling, 1960) of the status hierarchy, meaning that everyone knows the information, and further, everyone knows that everyone else knows it (and that everyone knows this, and so on).

When third parties favor the powerful, the strong get stronger as more individuals take their side, creating a feedback loop of increasing power (Snyder, 1997). The bandwagon decision rule solves the problem of discoordination because all third parties end up on the majority side of a heavily lopsided fight. When third parties use this decision rule, the resulting social structure is a steeply stratified dominance hierarchy in which the highest status individual wins all of their disputes, the second-ranked wins against everyone except the first-ranked, and so on until the last-ranked who always loses. The distribution of wins and losses will resemble a dominance hierarchy without side-taking in which all fights are one-on-one. A key difference, however, is that the fighting costs for higher status winners are dramatically reduced because they fight not alone but with the support of all third parties. When choosing sides is possible, bandwagoning exacerbates power differences.

The bandwagon strategy accomplishes third-party coordination but creates a new problem that we call *despotism*. When bystanders reliably side with the higher status fighter, the highest status individuals can initiate conflicts to advance their interests at little cost. Their aggression is essentially subsidized by third parties. High-status individuals can use this power to monopolize food, shelter, mates, etc. In chimpanzees, for example, high-status individuals use their power to control key resources (Boehm, 1999). In general, low-ranking individuals would benefit from mounting an attack to depose oppressive leaders. However, this strategy provides no long-term benefit unless third-party coordination in conflicts can be accomplished by an alternative method other than siding with the powerful.

Alliance Building

Another approach to choosing sides, which counters despotism, is an *alliance-building* strategy. Here individuals choose sides based on preexisting alliances and commit to support their allies against others when conflicts emerge. Lower status individuals can form alliances and commit to side with one another, rather than bandwagon, to defend against despotic individuals (Boehm, 1999). When individuals credibly signal that they will take an ally's side in future conflicts, that individual becomes especially valuable to the ally, making the ally more likely to side with the individual in future conflicts, which makes the ally more valuable, and so on. This feedback process causes increasing affinity among allies, a phenomenon referred to as "integrative spirals," whereas the opposite occurs among enemies (Snyder, 1984). The result of alliance building is that individuals tend to side with other individuals to the extent that those individuals side with them.

Whereas a bandwagon strategy requires knowing a single status hierarchy, alliance building is cognitively more complex (DeScioli & Kurzban, 2009a; DeScioli, Kurzban, Koch, & Liben-Nowell, 2011). Each individual has their own distinct set of loyalties or rankings that determines whom they will side with for all possible pairwise disputes. The matrix of all individuals' rankings of everyone else defines a loyalty landscape that determines alliance support for all possible disputes. To estimate the value of a particular ally (or an adversary's alliance support), individuals need to know everyone's loyalty rankings. An alliance-building rule would set the individual's ranking of others according to their loyalties to the individual, thus choosing sides to protect their most valuable allies. Studies of friendship have provided evidence that humans use this decision rule in their close relationships (DeScioli & Kurzban, 2009a; DeScioli, Kurzban, et al., 2011). When all individuals choose sides based on alliances, power tends to be balanced in the group such that no individuals have much more support than anyone else (DeScioli & Kimbrough, 2012; Snyder, 1984, 1997).

Alliance building eliminates despotism. In alliance building, individuals solicit support by offering support, and an individual cannot support everyone because to side with one individual is to side against another individual. Each person must side against any given individual, on average, 50% of the time, and these spurned individuals will, if building alliances, tend to side against that person with the same frequency. Hence, individuals cannot amass unanimous group support through alliance building.

Alliance formation sets in motion an escalating arms race that has been called the "alliance security dilemma" (Snyder, 1984). Individuals' alliances cause other individuals to form counteralliances in defense. When alliances form, those outside of the alliance are threatened by the possibility of joint action by the allied individuals. Those who remain unallied and independent are vulnerable to exploitation. The formation of defensive alliances offers protection, but at a cost—potential entanglement in other people's disputes. The result of an alliance security dilemma, similar to other arms races, is that individuals are no more secure than before any alliances were formed because everyone else has alliances too. Further, the total costs of fighting can be dramatically increased because individuals are now entangled in other people's disputes. Nonetheless, bystanders are often better off getting involved and choosing sides, despite the fighting costs, in order to preserve their alliances so they can avoid being exploited by other people's alliances.

Importantly, alliance building revives the problem of third-party discoordination. If everyone sides with their personal allies in disputes, then conflicts will tend to split the group into closely matched coalitions (DeScioli & Kimbrough, 2012). Thus, third parties suffer the costs of protracted conflict for every dispute. Consistent with this idea, ethnographic research shows that societies are more violent when individuals have a stronger sense of community and loyalty because disputes escalate as individuals' allies get involved (Black, 1998; Chaux, 2005; Cooney, 1998, 2003; Phillips & Cooney, 2005). So, although the formation of alliances might be an important strategy in a social world where third parties choose sides, commitments to side with preexisting allies carry the costs of third-party discoordination and escalated conflict. How can third parties choose sides without suffering from either discoordination or despotism?

Discoordination and Despotism: A Four-Player Example

To clarify the problems of discoordination and despotism, consider the following example. The simplest possible case in which these problems could arise is a game with four players $\{P_1, P_2, P_3, P_4\}$ in which two players are randomly chosen to be fighters who dispute over a resource with a value V (Maynard Smith, 1982), and the other two players are bystanders who each choose sides with one of the disputants. We assume that spurning both individuals is worse than supporting one disputant (i.e., a negative payoff in this example), motivating bystanders to choose sides. We further assume that the players are equally powerful so that the disputant with more supporters wins the resource. If there is a clear power asymmetry (three against one), then the fight is settled with a threatening display at no cost. If a tie occurs (two against two), then the resource is randomly allocated to one disputant, but all players incur a fighting cost C.

We focus on the side-taking decisions of the third parties or bystanders. Figure 1 shows the bystanders' decisions in a 2×2 game matrix. Bystander 1 and Bystander 2 each choose between two strategies, *Fighter 1* or *Fighter 2*, indicating which side they will take. The payoffs are shown for all four players in the order: Fighter 1, Fighter 2, Bystander 1, Bystander 2. If the bystanders choose the same side, then the chosen fighter outnumbers the opponent (three vs. one) and so wins the resource value V at negligible cost, giving the loser and both bystanders a payoff of 0. However, if the bystanders choose different sides, then the two sides are evenly matched, two versus two. Each fighter has a 50% chance of winning the resource V, giving an expected value of $\frac{1}{2}V$. Additionally, all players suffer the cost C from an escalated dispute. This game is a coordination game with two equilibria (*Fighter 1, Fighter 1*) and (*Fighter 2, Fighter 2*). The bystanders need to choose sides with the same fighter in order to avoid the costs of an escalated dispute.

		Bystander 2	
		Fighter 1	*Fighter 2*
Bystander 1	*Fighter 1*	$V, 0, \mathbf{0}, \mathbf{0}$	$\frac{1}{2}V - C, \frac{1}{2}V - C, \mathbf{-C}, \mathbf{-C}$
	Fighter 2	$\frac{1}{2}V - C, \frac{1}{2}V - C, \mathbf{-C}, \mathbf{-C}$	$0, V, \mathbf{0}, \mathbf{0}$

Figure 1. Side-taking game. A 2×2 matrix game in which two players, Bystander 1 and Bystander 2, each choose to side with Fighter 1 or Fighter 2. The matrix shows the payoffs to Fighter 1, Fighter 2, Bystander 1, and Bystander 2, respectively, in terms of the resource value, V, and the costs of escalated fighting, C. The bystanders' payoffs are in bold.

How can third parties coordinate their decisions? To use a bandwagon strategy, all players $\{P_1, P_2, P_3, P_4\}$ observe a public status hierarchy. For instance, suppose the status ranks are $P_1 > P_2 > P_3 > P_4$, such that P_1 has the highest status and P_4 has the lowest status. For all possible disputes, third parties side with the higher status disputant. For example, in a fight between P_1 and P_3, the bystanders (P_2 and P_4) would both side with P_1 against P_3. Third parties can use this strategy to coordinate their decisions. Even if the particular hierarchy is disadvantageous for an individual as a fighter, such as P_4 in the present example, it is still beneficial, when acting as a third party, for P_4 to choose sides based on status.

When all players use a bandwagon strategy, power and resources are unequally distributed. To quantify this inequality, suppose that all possible disputes occur once, such that each player has three disputes as a fighter (rather than bystander) that end in a win, loss, or tie. For a bandwagon strategy based on status, the win–loss–tie records for the four players would be 3–0–0, 2–1–0, 1–2–0, and 0–3–0, for the highest to lowest ranked players, respectively. Clearly, the low status individuals do not fare well in this hierarchy, but they might nonetheless choose sides based on status to avoid discoordination costs, C.

The inequality in power created by bandwagoning creates a threat of despotism: Third-party support can be exploited by high-status individuals to monopolize resources. To emphasize this problem, imagine an addition to the game: After the dispute is resolved, each third party in turn has the opportunity to contest the resource following the same rules. In this extended version of the game, the highest status player would despotically contest and monopolize all resources.

To counter despotism, players can use an alliance-building strategy. In the present example, suppose that P_3 and P_4 form an alliance that specifies that they will always side with each other in disputes, ranking each other first, and in others' disputes they default to the bandwagon strategy. (We make the simplifying assumption that they can credibly commit to an alliance.) Now, the win–loss–tie records are 1–0–2, 0–1–2, 1–0–2, and 0–1–2. Both P_3 and P_4 have improved records so they could both favor their alliance, depending on the magnitudes of C and V. Specifically,

both players improved two losses to two ties. Hence, they will favor the alliance when the additional payoff from two ties, $2 \times (\frac{1}{2}V - C)$, is greater than the additional fighting costs they incur when acting as bystanders in two of their ally's disputes against higher status fighters, $2 \times C$, giving the condition $V > 4C$, which describes when both players would favor the alliance. In response to this alliance, P_1 and P_2 should cement a counteralliance, ranking each other first, to avoid a further shift to a status hierarchy controlled by P_3 and P_4 (see Snyder, 1984, 1997).

Next, P_2 and P_4 are at a disadvantage, so they could benefit by forming a secondary alliance, ranking each other second. That is, they will agree to side with each other, except against their primary allies (e.g., P_2 will side with P_4, except against P_1). Similarly, P_1 and P_3 should form a counter–secondary alliance. Now, all fights end in ties, the win–loss–tie records are 0–0–2, 0–0–2, 0–0–2, and 0–0–2, respectively, and no one has more power or resources than anyone else, solving the problem of despotism. This outcome would be stable because any advantage by one individual creates an opportunity for a beneficial alliance among those who are disadvantaged. Hence, the construction of alliances and counteralliances promotes equal distributions of power and resources (DeScioli & Kimbrough, 2012; Snyder, 1984, 1997). However, the problem of discoordination is now extreme: All fights are evenly matched with high fighting costs.

In sum, in this simplified model, when third parties bandwagon and choose sides based on status, they successfully coordinate but also enable despotic exploitation. When third parties choose sides based on alliances, they eliminate despotism but fail to coordinate, leading to costly escalated fights. We offer this example as the simplest possible case showing the problems of bystander coordination and despotism in order to highlight their core features. The model can be elaborated in many ways including, importantly, adding more agents, which creates a crowd of bystanders and an *n*-player coordination problem among them. Human disputes frequently involve more than four players (Black, 1998; Cooney, 1998), and these additional players heighten the challenge of coordination as well as the potential for despotic exploitation (Snyder, 1997). How can third parties choose sides in a way that avoids both of these problems?

Dynamic Coordination for Choosing Sides

An alternative strategy for third-party coordination, and the core principle of the present theory, is using a *correlated equilibrium* (Aumann, 1974), a solution concept in which players coordinate by making decisions based on a public signal. A traffic light is an example of a device for creating a correlated equilibrium. Drivers want to pass through an intersection while coordinating with other drivers to avoid a crash. A signal everyone can observe—and, crucially, everyone knows that everyone else can observe—creates a correlated equilibrium. Given "common knowledge" (Aumann, 1976; Schelling, 1960) of the traffic light, it is in each player's interest to make decisions based on the light, proceeding when the light facing them is green and stopping when it is red. To be useful for coordination, the signal system needs to be clearly observable, like ringing bells on ships that signal the change of shift. However, the signals themselves can be arbitrary. There need be no relationship, at all, between the nature of the signal and the behavior that is being coordinated, just as there is no intrinsic relationship between lights and driving through intersections.

How can third parties use the correlated equilibrium concept to solve their coordination problem? As an example, consider a simple strategy: Third parties could flip a coin and choose sides based on the result. If the coin flip is in view of all third parties and everyone knows which disputant is heads and which is tails, then third parties can use the coin flip to ensure that they all take the same side. By coordinating on a coin flip, third parties can create a heavily lopsided dispute that will be resolved quickly at low cost to third parties. Anyone who ignored the coin flip, for instance, prioritizing their loyalty to one disputant, would endure the cost of discoordination with the majority.

The key advantage of a coin flip (or an analogous signal) is that it is impartial, not tied to individual identity. The coin flip is an example of a dynamic coordination strategy because third parties can use the device to synchronize side-taking while dynamically changing which individual they support across different conflicts. Because coordination is achieved without reference to individuals' identities, the problem of despotism is avoided: There are no extremely powerful individuals because every person has an equal chance of winning each conflict. In the four-player example above (see Alliance Building), the players' win–loss–tie records would all be 1.5–1.5–0, on average, with no ties occurring, and with no individual with a greater chance of victory than any other individual. Once a correlated equilibrium mechanism is implemented, individuals who choose sides based on status or alliance would be at a disadvantage because they would frequently be in the losing minority. Thus, a correlated equilibrium mechanism for choosing sides (such as a public coin flip) would be an evolutionarily stable strategy (Maynard Smith, 1982) against alternative decision rules based on status or alliance.

Coordinating on a public signal is not without costs and is not always advantageous. The primary cost is that this strategy will often require individuals to side against family, friends, and ingroups. The strategy requires impartiality and the sacrifice of personal ties in order to detach decisions from disputants' identities. This cost is inevitable for successful coordination because two disputants will each have their own family and friends who prioritize them above their adversary. If everyone acts on their loyalties, then discoordination will result. In general, third parties are forced to choose between the costs of betraying their loyalties and the costs of discoordination with other third parties. These costs will depend on many factors such as the value of the particular relationship, the relative formidability of potential condemners, the decisions of other bystanders, etc. Successful performance will require third parties to closely tailor their use of loyalty versus impartiality to the details of particular conflicts.

Importantly, the payoffs of the dynamic coordination strategy also depend on the strategies used by other players. This is a familiar feature of all coordination games, which are characterized by multiple equilibria (Schelling, 1960). Depending on the precise resource values, fighting costs, and probabilities of repetition, there can be multiple evolutionarily stable strategies, which can include bandwagoning and alliance building. This possibility raises important questions about how social groups might transition among social orders based on bandwagoning, alliance building, and dynamic coordination (see also Moral Impartiality and Moral Disagreement: Strategic Morality). It also suggests the

psychological hypotheses that people will seek to align their strategies with their perceptions of other people's strategies and, further, that people will use signaling and communication to influence group behavior toward the strategy that is most individually advantageous.

In sum, the correlated equilibrium concept allows dynamic coordination, in which third parties are able to coordinate while dynamically changing which individuals they support. The nature of the signal is not critical: Casting lots, cracks in a turtle shell, configurations of oracle bones, the outcome of a poison ordeal, the conventions of a legal system—all of these signals can allow third parties to coordinate. What is critical is that the signal must be public knowledge, must stand out among other possible signals, and crucially, must not be tied to individual identity.

Choosing Sides Based on Actions

What features of conflict events could be used as signals for coordination? In addition to individuals, there are also the actions taken by those individuals in the dispute, which could provide a signal for coordination. Just as humans naturally recognize other people's identities (e.g., Pascalis, de Haan, & Nelson, 2002), humans also naturally and automatically parse other people's behavior streams into identifiable component actions (Baldwin & Baird, 2001; Kurby & Zacks, 2008; Zacks & Swallow, 2007). This ability further allows particular linguistic labels, verbs, to refer to particular types of actions (Pinker, 2007). Action parsing can allow humans to parse conflict events into component actions that are publicly observable and identifiable. People share the same action parsing mechanisms, and the outputs of these systems are consistent across individuals (Baldwin & Baird, 2001; Kurby & Zacks, 2008; Zacks & Swallow, 2007). Hence, people's spontaneously generated action representations are available to be used as a basis for dynamic coordination among third parties.

Suppose that group members construct, in advance, a set W of n distinct actions $\{w_1, \ldots, w_n\}$ that third parties can use to choose sides. Specifically, they will side against an individual who chooses an action in the set W. To be useful, the set W would have to include actions that are likely to occur in conflicts, and it should be as comprehensive as possible such that any given conflict will tend to include at least one action in the set W. To facilitate this goal, the set should be open: When new types of conflict arise—perhaps due to new discoveries, technologies, or cultural forms—new actions can be added to the set W.

Further, the action-based coordination mechanism would have to be able to handle disputes in which both disputants have chosen an action in the set W. To accomplish this goal, the actions $\{w_1, \ldots, w_n\}$ could be ranked and assigned relative magnitudes. Third parties can side against the individual who chose the action in the set W with the greatest magnitude. To facilitate third-party coordination, the assignments of magnitudes to actions would need to be "common knowledge" (Aumann, 1976; Schelling, 1960), would need to be established in advance of conflicts, and would require consensus among third parties on the ranking of actions used to choose sides. Any disagreement about the magnitudes for different actions could interrupt coordination, threatening third parties with the high costs of escalated disputes.

In sum, dynamic coordination based on actions could be an advantageous strategy for choosing sides. This strategy tends to produce lopsided disputes with decisive victories rather than closely matched disputes with escalated fighting. By using actions rather than identities, coordination is achieved without empowering particular individuals to despotically monopolize resources.

Moral Cognition as a Dynamic Coordination Device

The dynamic coordination theory of morality holds that evolution favored individuals equipped with moral intuitions who chose sides in conflicts based, in part, on "morality" rather than relationships or status (see Figure 2). To perform this function, humans parse the behavior of disputants (as observed or verbally described) into identifiable actions and compare these actions against the set of moral wrongs, W. Group members construct this set of moral rules in advance of conflict, sometimes with negotiation and debate, and these rules can persist over long time periods. To achieve dynamic coordination, third parties choose sides against the individual who has chosen the action with the greatest wrongness magnitude.

Of course, the output of moral cognition is only one of the critical factors that a bystander will consider when choosing sides. Importantly, bystanders should consider their loyalties to family, friends, and ingroups which will often conflict with their moral judgment, producing countervailing motives. They should also consider the welfare consequences to the people involved, even if they are all strangers (e.g., trolley problems), in order to anticipate other people's reactions, and again, these welfare computations might yield outputs that conflict with the outputs of moral computations (Kurzban et al., 2012). The final decision will need to account for these and a variety of other factors. The moral strategy will be most useful when there are high costs of discoordination and despotism that outweigh other considerations.

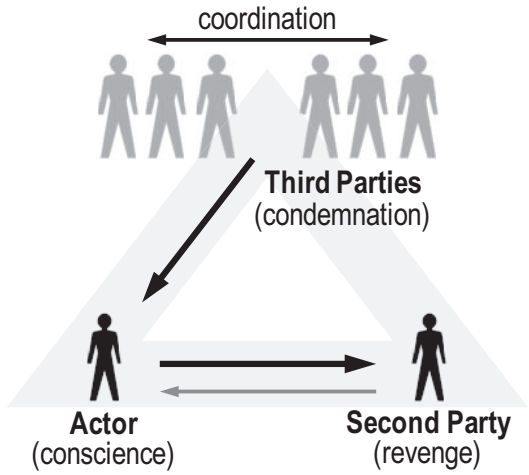

Figure 2. Diagram of a strategic interaction among two disputants and a set of third parties who choose sides. Arrows show the activity of one role toward another as described in parentheses. Third parties face a problem of coordination, particularly because some individuals are more closely affiliated with the disputant on the left and other individuals are closer to the disputant on the right. Moral cognition allows third parties to coordinate by taking sides against the actor who has chosen the action with the greatest wrongness magnitude.

In this way, moral cognition allows third parties to coordinate moralistic attacks, avoiding costly fights among themselves, without concentrating power in potential despots. We turn now to discuss how these ideas can explain the three principal components of moral condemnation: moral judgment, moralistic punishment, and moral impartiality. We conclude the section by discussing the relationship between condemnation and conscience.

Moral Judgment

We propose that the moral dimension of human experience functions to map the contours of other people's side-taking decisions for potential disputes. Humans naturally and effortlessly assign moral values to people's actions so that, if necessary, they can predict which side other bystanders will take in disputes. Moral representations include a *perpetrator* and a *victim* (DeScioli, Gilbert, & Kurzban, 2012; DeScioli & Kurzban, 2009b; Gray & Wegner, 2009; Gray, Young, & Waytz, 2012) because these designations indicate which disputant to side against and which disputant to support, respectively, whereas the *condemner* role identifies other third parties for coordination. People's deep interest in other people's wrongdoing and other bystanders' moral opinions allows them to anticipate and respond to conflicts. In preparation for these perilous situations, individuals can use moral gossip to probe and negotiate which side they will take. In this sense, moral cognition is comparable to other computational systems designed to detect coalitional alignment using cues such as race, accent, or dress (Kinzler, Dupoux, & Spelke, 2007; Kurzban et al., 2001).

The dynamic coordination theory predicts that moral cognition will focus on locating coordination points that are independent of disputants' identities. Moral cognition focuses on actions because they provide a source of public signals for coordination. This explains why moral cognition uses Kantian "categorical imperatives": Lying is always wrong, as Kant (1785/1993) argued, because moral imperatives are not designed to promote beneficial behavior, but rather to allow third parties to synchronize side-taking.

The particular contents of moral rules are not critical to their dynamic coordination function (but see The Content of Morality), just as the specific colors of traffic lights are not critical to their function ("go" could be signaled by blue instead of green). Action-based coordination requires a set of moral wrongs that will pick out at least one action for a variety of conflicts about violence, sex, food, communication, property, etc. However, coordination by itself places few constraints on which specific actions will be moralized. For example, it is useful to have rules for fights over resources, but whether the rule is "do not steal" or the opposite "do not refuse to share" is not critical for third-party coordination, even if these different rules lead to opposite "wrong" individuals for a given conflict. As long as third parties agree on which rule applies, dynamic coordination is accomplished regardless of which side is found guilty and which innocent.

A third-party coordination function predicts that there will be a diversity of moral rules to cover the diversity of human conflicts. The various content domains of morality—violence, property, communication, sex, food, beliefs—can be understood as domains in which conflicts can occur, giving rise to the need to coordinate side-taking. Similarities among rules within each content domain result from their common source in a type of dispute, and these similarities allow moral rules to be grouped together like any other set of objects (Haidt, 2007; Rosch & Mervis, 1975). Each type of dispute will tend to have its own domain-specific mechanisms (e.g., mate-guarding mechanisms differ from lie detection mechanisms), but what is distinctively moral about different content domains, such as infidelity or deception, is what they share in common—they cause disputes and side-taking. Hence, moral rules share a common set of properties across content domains (DeScioli, Asao, & Kurzban, 2012)—features associated with a dynamic coordination function.

Because novel types of conflicts can arise, moral cognition should be capable of generating new moral rules for new disputes. The cognitive process of "moralization" (Rozin, 1999) can be understood as a mechanism for extending moral strategies to novel types of conflict. Still, different moral rules need to be comparable, despite their diversity, to be used for coordination when both sides have chosen morally wrong actions. Indeed, people are able to compare the severity of different moral violations across content domains, such as violence and property offenses (Robinson & Kurzban, 2007). The output of moral judgment is not simply "right" or "wrong," but rather "wrong" actions are highly differentiated in wrongness magnitudes (unlike, e.g., grammar judgments).

Because third-party coordination places few constraints on rule content, this function is consistent with high variation in moral rules across cultures (Shweder et al., 1997). The particular rules in a cultural group might be determined by any number of factors, giving those rules greater signaling power than alternatives (see The Content of Morality). Emotions such as disgust, public conventions, welfare concerns, specific precedents, and other factors can converge and diverge to favor one rule over another. Furthermore, different human groups exploit different ecologies, encounter novel conflicts, and invent new moral rules to handle these situations. Because sources of conflict differ across cultures, moral rules should also be expected to differ. Convergence in moral rules is expected when similar problems arise across ecologies. To anticipate one argument below, to the extent that "do not steal" is a better rule than "do not refuse to share" from the standpoint of modern economic growth, these rules might have an advantage in cultural group selection (A. Smith, 1776).

The prevalence of destructive moral rules is consistent with a coordination function. The present theory holds that moral cognition functions to coordinate condemnation, not to promote beneficial behavior or to deter harmful behavior (but see The Content of Morality). This strategic goal is served by locating identifiable actions for condemnation regardless of whether these actions are harmful or beneficial. When beneficial or harmless actions provide useful signals for coordinated condemnation, moral cognition will have destructive effects by coordinating aggression toward individuals who have done no harm (McWilliams, 1996; Nozick, 1974). The dynamic coordination function, then, can explain why moral rules frequently have destructive consequences—a key feature left unexplained by previous altruism-based theories of morality.

A challenging issue for the dynamic coordination hypothesis is the occurrence of wrongs that appear to be solitary such as masturbation or suicide. If moral judgment is for choosing sides, then it is expected to focus on situations with two disputants. One possibility is that these wrongs are a consequence of the open-

ended nature of moral cognition, designed for accommodating new disputes. Once implemented, this feature might allow for a wide variety of actions to be moralized, even perhaps actions without victims. Moreover, in order to debate and agree on moral rules, humans might have evolved the ability to contemplate the wrongness of actions independent of current disputes or current victims. A related possibility is that people do in fact represent victims for apparently solitary moral violations. Indeed, previous research found that people who view "victimless" violations as wrong also tend to perceive victims of these actions (DeScioli, 2008; DeScioli et al., 2012; see also Gray & Wegner, 2009). For instance, among participants who thought suicide was wrong, 88% thought there was a victim, compared to 39% for those who thought it was not wrong. Similarly, for drug use these values were 89% and 40%, respectively, and for consensual incest between siblings, these values were 77% and 8%, respectively. Hence, it seems that people do represent victims even for victimless crimes which might reflect an underlying cognitive template that includes perpetrator and victim roles.

A dynamic coordination function imposes some important constraints on moral cognition. First, moral rules cannot be conceptualized as mere opinions or personal taste because this would undermine consensus building. If every person had their own moral rules, then moral judgment would be useless for third-party coordination. Indeed, many people view moral rules as more like objective facts than subjective tastes (Goodwin & Darley, 2008). This could also explain why people are averse to differences of opinion in moral matters (Haidt et al., 2003). Second, moral rules cannot be tied to the identities of individuals, particularly high-status individuals, because this would allow for despotism. These are precisely those properties of moral cognition—universalism and independence from authority—that appear at the onset of children's moral reasoning (Posada & Wainryb, 2008; Smetana & Braeges, 1990; Smetana et al., 1993; Turiel, 1998). In short, a coordination function can explain why moral rules are highly variable and also why people vehemently reject variation, asserting the universality of their own morals.

Moralistic Punishment

The dynamic coordination theory offers a novel perspective on why people want costs to be imposed on moral violators: Moralistic punishment functions to signal which side the punisher is on for potential disputes. To choose sides effectively and convincingly, an individual must be prepared to support the favored party in taking any hostile measures that are necessary to secure their interests against the adversary. Minimally, taking sides would require verbal encouragement of those who prosecute the opponent. More boldly, individuals can engage in costly punishment themselves to clearly signal which side they are on, which could otherwise be in doubt.

Previous research provides evidence that moralistic punishment performs a signaling function (Kurzban et al., 2007; Piazza & Bering, 2008). In these experiments, third parties were given the opportunity to punish, at a cost, individuals who behaved selfishly toward someone else in an economic game. The experimental manipulation varied the anonymity of participants' punishment decisions. The results showed that there was greater moralistic punishment when decisions were public. This audience effect suggests that moralistic punishment sends a signal to other people, but what exactly is being signaled remains a matter of debate (Barclay, 2006). We suggest that punishment signals which side the punisher is on for potential conflicts arising from the situation.

Timing is important for costly moral signals. An early gesture of moralistic punishment can serve as a rallying cry to support one side. An early signal can foster dynamic coordination by broadcasting the moral righteousness of the favored party to other third parties trying to coordinate condemnation. Later, if coordination is successful, the costs of punishment are reduced via the growing numerical advantage of the condemners against the condemned. At this point, punishment is cheap and third parties can be more brazen, hurling insults and throwing stones to signal their alignment with the majority.

A useful design feature in a moral mechanism for taking sides is the ability to block concern for the wrongdoer's suffering. Sympathy for a wrongdoer's pain could motivate helping behavior, and such a display might be interpreted as siding with the perpetrator, potentially drawing attacks from condemners. Mechanisms that induce indifference or malevolence toward wrongdoers can preclude this outcome. This idea could explain historic phenomena such as public executions in which crowds of people enjoyed watching brutalities against people labeled as wrongdoers (e.g., Kadri, 2005). Similarly, in modern laboratory studies, participants showed insensitivity to pain and suffering when the person was viewed as a wrongdoer (T. Singer et al., 2006).

This account of moralistic punishment draws attention beyond the punisher and the perpetrator to the larger social world in which they are embedded. Humans live in a world of alliances in which disputes originating between two individuals can quickly escalate to include family, friends, and groups (Black, 1998; Chaux, 2005; Cooney, 1998, 2003; Phillips & Cooney, 2005). To minimize the costs of entanglement, third parties need to synchronize their side-taking decisions, and further, they need to unambiguously communicate their alignments to everyone else. In this strategic environment, moralistic punishment functions to signal which side the punisher is on, instead of being aimed at deterring behaviors such as lying, stealing, or incest. Moreover, what is signaled is not an enduring relationship but rather a momentary allegiance struck among third parties to avert an escalating conflict between rival alliances.

Moral Impartiality

The dynamic coordination hypothesis predicts that moral judgment will be impartial—independent of the actors' identities. This feature distinguishes dynamic coordination from alternative strategies for choosing sides. To avoid the perils of despotism and discoordination, third parties must coordinate based on some observable signal that is independent of the identities of the disputants.

Impartiality is a fundamental and universal feature of moral cognition (DeScioli & Kurzban, 2009b; Tetlock, 2002; Tooby & Cosmides, 2010). Importantly, this claim does not imply that people will always judge impartially. On the contrary, impartiality is costly to people's relationships, and they will need to balance the benefits of partiality against the costs of discoordination with the adversary's supporters. People can use different strategies to

choose sides, but the distinctively moral strategy is one that ignores individuals' identities.

It is the appearance of impartiality that is most critical. If the appearance can be maintained, despite actual partiality, then the benefits of coordination and supporting allies can be reaped simultaneously. The problem for third parties is negotiating side-taking with other third parties who have stronger loyalties to the adversary. Partiality creates the problem of alliance-based escalation in the first place. Individuals cannot use their own personal ties as a basis to persuade the adversary's allies to abandon their personal ties. At the same time, there is a strong incentive to fake impartiality while actually supporting allies or opposing enemies. For example, this strategy could be facilitated by cognitive systems that start with moral conclusions and generate post hoc beliefs about harm, intentions, causality, and force (Alicke, 1992; Haidt, 2001; Knobe, 2005; Young & Phillips, 2011) in order to use these beliefs to provide a seemingly impartial basis for judgment, aimed at persuading other people (Mercier & Sperber, 2011). In turn, humans should be vigilant against other people's deceptive use of impartiality, which might explain strong reactions against moral hypocrisy (Kurzban, 2010).

The strategic importance of impartiality makes it a point of vulnerability against alternative social orders based on bandwagoning or alliances. Individuals with high status or strong alliances might prefer a social world in which bystanders choose sides based on power or relationships, rather than morality. These individuals need to undermine impartiality in order to establish alternative rules for bystander side-taking. One way to accomplish this goal is to take advantage of the openness of moral cognition. This openness allows people to moralize defiance of authority or disloyalty to allies (Haidt, 2007). These paradoxical moral rules can promote partiality, thereby disabling moral coordination while favoring a social order in which people choose sides based on power or relationships. On the other hand, the moral force of impartiality can be leveraged to counter and delegitimize rules against defiance and disloyalty, thereby favoring a social order based on moral rules. Hence, impartiality is a major battleground on which individuals compete to influence group side-taking strategies—authority versus relationships versus morality (see Moral Disagreement: Strategic Morality).

Impartiality sets moral cognition in stark relief against altruistic behavior such as kin altruism. Human parental care, for instance, is not impartial. People care for their own children more than others' children. Indeed, relatedness is a very strong predictor of altruistic behavior (Burnstein, Crandall, & Kitayama, 1994; Gaulin, McBurney, & Brademan-Wartell, 1997; M. S. Smith, Kish, & Crawford, 1987). For moral decisions, however, altruistic mechanisms are undermined by moral judgment, leading humans to fail to maximize benefits to kin when doing so requires violating a moral rule. Kurzban et al. (2012) found that many participants reported that they would not kill one sibling to save five siblings, which is the opposite of what kin selection predicts and what has been observed in similar decisions in nonhuman species such as burying beetles, which kill some offspring to feed the bodies to other offspring (Mock, 2004).

In sum, the puzzle of moral impartiality can be explained as a critical element of a strategy for choosing sides. If third parties were to choose sides based on relationships, then they would risk costly discoordination. At a minimum, individuals must publicly declare impartial judgment—even if this damages their alliances—to try to reach consensus with other third parties.

Moral Conscience

We have applied the dynamic coordination model to three features of moral condemnation: judgment, punishment, and impartiality. In this section, we turn our attention to moral conscience. By explaining condemnation, the dynamic coordination model can simultaneously explain why people have conscience mechanisms for avoiding morally wrong behavior—to prevent moralistic punishment (e.g., DeScioli, Christner, & Kurzban, 2011). Punishment is a powerful force that can potentially explain the evolution of any behavioral adaptation selected for by the costs imposed by punishers (Boyd & Richerson, 1992). Moral conscience might function in part as a defense system for avoiding actions that, if detected, could provoke a coordinated attack by third parties.

This theory casts a new light on Kantian categorical imperatives and nonconsequentialism in moral judgment. Recall that Kantian decisions involve making choices subject to action constraints:

$$\max_{a \in A} u(\mathbf{y}), \text{ subject to the constraint, } a \notin W.$$

This decision procedure can be understood as a mechanism for anticipating and avoiding condemnation by third parties, who use the action set W as a source of public signals for coordinating punishment. That is, actors' moral action constraints are a counteradaptation to a different cognitive mechanism used by third parties, which specifies choosing sides against the player who has chosen an action in the set W.

Nonconsequentialism occurs because moral cognition is not primarily designed for promoting altruistic behavior or beneficial consequences, but rather for third-party coordination. The moral rule "do not kill" works just as well for third-party coordination when killing one person can save five people. Hence, the coordination theory is consistent with empirical results from moral dilemmas such as the footbridge trolley problem, showing that moral judgments do not track welfare outcomes (Mikhail, 2007). Instead of welfare, moral judgments are attuned to details of the actor's behavior such as whether the violation occurred as a means or a byproduct (Cushman, Young, & Hauser, 2006; Mikhail, 2007). This observation can be understood as resulting from the role of action parsing in third-party coordination. Any structural features of behavior that influence action parsing can in principle influence wrongness judgments by changing how observable and identifiable the wrongful actions are for a given event. If so, then moral judgment tracks the quality of the coordination signal available to third parties, rather than the benefits or harm caused by actors.

Nonetheless, because people can moralize a range of identifiable actions, they can in principle moralize nonconsequentialist decisions themselves, as, for example, advocated by utilitarian philosophers. That is, it is possible to use the expected consequences of actions to make moral judgments and coordinate side-taking. There are several reasons why this approach is not more prevalent. First, consequentialist behavior might not be a category that is sufficiently identifiable to be useful for coordination, perhaps being too high level compared with more basic categories such as

lying, killing, and stealing. Second, welfare consequences might be particularly difficult to use for coordination given that they tend to be the basis of the dispute in the first place. That is, different sides will tend to disagree on the weight to put on each disputant's welfare, potentially making welfare judgments ill-suited for coming to a consensus. In sum, nonconsequentialism in moral conscience might be explained as a defensive strategy, which in turn can be explained by the details of the coordination problem confronting bystanders who choose sides.

The Content of Morality

The openness of moral cognition to new and revised moral rules creates the potential for moral variability and secondary processes that shape this variability. In this section, we explore how the evolution of moral mechanisms gives rise to a secondary strategic game in which individuals seek to establish moral rules that serve their personal interests (Kurzban, Dukes, & Weeden, 2010; Tooby & Cosmides, 2010; Weeden, 2003). We suggest that this secondary and derivative game shapes the observed content of moral rules, both the themes and variations across cultures. Once morality was possible, new adaptations evolved, designed to steer the particular rules used by a group to the individual's advantage.

The threat of moralistic punishment creates opportunities for those who can influence the set of moral wrongs, *W*, in their group. By advocating particular moral rules, the weapon of collective punishment can be directed toward behaviors against one's personal interests. This goal could be accomplished through adaptations for proposing, debating, negotiating, and revising moral rules.

Moral Consensus: Rawlsian Morality

People will favor moral rules that are in their personal interest, and some self-serving rules also happen to be in everyone else's interest. A rule against killing, for instance, can benefit anyone who can be killed: Everyone can agree that other people should not kill, even if they themselves prefer to be able to kill, particularly because they do not know in advance whether they will be the killer or the victim.

We refer to this type of rule as *Rawlsian morality* (Kurzban et al., 2010) after the philosopher John Rawls (1971), who argued for adopting the set of moral rules that people would choose if they did not know their own identity in society (Binmore, 2007). For example, it is in most people's interest to favor a rule against deception because they do not want to be deceived. Similarly, a rule against breaking contracts is in everyone's interest because they do not want others to break contracts with them, and moreover, their own contracts are more credible when enforced. Rawlsian moral rules are not the only moral rules, but we would expect them to be the most stable and universal (Binmore, 2005; Binmore & Samuelson, 1994).

However, we emphasize that Rawlsian rules are self-serving only when other people are using moral rules to choose sides. When, for instance, individuals choose sides based on status, high-status individuals have no incentive to favor a moral rule against violence. If a low-status individual assaults a powerful person, then the bandwagoning group will side with the powerful individual, and a moral rule would offer no additional protection. This idea can explain why people ignore basic moral rules when they are in extremely hierarchical groups (e.g., the Nazi regime; Zimbardo, 2007). When morality-based coordination is interrupted in extreme environments such as warfare or rioting, people can be expected to discount or ignore moral evaluations, instead making decisions based on other factors such as self-interest, altruism, or group loyalty. Rather than being blindly "internalized," people's observance of moral rules depends on circumstances, especially whether other people are currently using moral rules to choose sides. Rawlsian morality holds general appeal only in a particular social order: When bystanders choose sides based on impartial rules (rather than status), then rules against actions such as killing, assaulting, and lying are favored by everyone, including both high- and low-status individuals.

In sum, Rawlsian morality is the prototypical moral category, including rules against killing, stealing, and lying. These rules serve most people's interests, and they likely contribute to the common public sentiment that morality is an intrinsically benign and positive force in society, as well as the tendency among researchers to view morality as a form of altruism (DeScioli & Kurzban, 2009b). This uncritical positivity ignores an abundance of destructive and antagonistic moral rules, possibly as part of an intuitive strategy to deny the validity of moral rules against one's interests. Humans are, understandably, a morally anxious species because of their dependence on moral rules for security in a complex world of rival alliances. The stability and universality of Rawlsian morality offers comfort while so many other moral rules are under intense negotiation, the issue we turn to next.

Moral Disagreement: Strategic Morality

There is not, of course, always agreement on moral rules. People disagree about abortion, animal rights, drug use, homosexuality, promiscuity, digital property rights, and many other issues. Individuals stand to gain by proposing and defending moral rules that benefit themselves. When different people have different interests, moral disagreement is likely to result.

For example, research shows that humans pursue a range of mating strategies including short-term and long-term mating (Buss, 2006). If individuals differ in their mating strategies, then they will tend to disagree about rules that impede or facilitate particular strategies. Weeden (2003) applied this idea to abortion, and his data suggest that those people who benefit from promiscuous sexual practices are more likely to be pro-choice, whereas those people who benefit from monogamy are more likely to be pro-life. Similar differences would be expected to be observed for rules about fornication, adultery, and contraception. People's life history strategies determine the moral regime that most benefits them, though these are not the reasons they present when defending their moral positions (Kurzban, 2010; Kurzban et al., 2010). Instead, people claim their positions derive from general principles of potentially universal appeal, which could be aimed at persuading other people who are otherwise indifferent or opposed to their position (Tooby & Cosmides, 2010).

Digital property rights offer another example. People who produce information—music, books, scientific theories—benefit from being able to sell this information. Consumers of information benefit from being able to acquire these products at no cost. Digital

consumers benefit from a "share everything" rule, whereas producers benefit from a "take nothing" rule.

Moral rules about authority and loyalty might be expected to be particularly contentious because they can potentially undermine moral impartiality. Particularly powerful individuals can benefit by moralizing disobedience, whereas powerless individuals moralize the opposite, oppression and coercion. If the powerful can establish rules against insubordination, then bystanders will choose sides based on authority, effectively converting a moral social order into an authoritarian order where the bandwagon strategy prevails. Similarly, well-connected individuals with particularly strong alliances can benefit by moralizing disloyalty, whereas less connected individuals moralize the opposite, favoritism. Again, what is at stake is whether bystanders will choose sides based on relationships or based on impartial action constraints. Consistent with these ideas, there is particularly strong disagreement for moral rules about authority and loyalty (Haidt & Graham, 2007). This observation can be explained by the fundamental tension between authority, loyalty, and impartial morality as alternative strategies for choosing sides.

These struggles to control the rules have been discussed in the legal literature. The "conflict model" holds that criminal laws are created through an "on-going struggle between vested interest groups which seek to have their particular values legitimated and supported by the coercive power of the state" (Thomas, Cage, & Foster, 1976, p. 110). An analogous but informal process occurs in people's interpersonal lives as individuals argue about the moral rules that will govern their side-taking decisions in disputes.

In short, moral rules restrict an actor's scope of behavior, and some rules impose greater costs on some people than others. This leads to conflicts about the content of moral rules, particularly when strategic interests are in play (Robinson & Kurzban, 2007), though we emphasize that the strategic elements are not always obvious, even to actors themselves (Kurzban, 2010; Weeden, 2003).

Moral Epidemiology

A dynamic coordination function requires flexible moral systems that can learn the local moral rules of particular groups and can create new rules for new conflicts through the moralization of actions (Rozin, 1999). Moral learning mechanisms allow humans to mint, acquire, and transmit moral rules. These processes create the potential for an epidemiological approach that can explain moral content in terms of the factors that influence the social transmission of moral rules. As we argue above, Rawlsian rules are likely to be adopted because they are self-serving, and the transmission of strategic rules depends on struggles among opposing interest groups to control the rules. We now address two other factors that influence rule transmission: cultural group selection and emotions.

In cultural group selection, a practice spreads by providing a competitive advantage to groups that have the practice. That is, socially acquired representations can produce feedback loops on their frequency by enhancing group success (Boyd & Richerson, 1985; see Binmore, 2005, for extended discussion of cultural group selection for social rules). Consider, for instance, two types of property rules. One rule is "share everything," in which a person who withholds resources from another person is condemned. The other rule is "take nothing," in which a person who takes an object from another person is condemned. Depending on the distribution of resources, one of these rules might be more advantageous than the other for group success. The Nuer conquest of the Dinka provides an example of one group gaining an advantage over another group due to cultural property rules (Boyd & Richerson, 2005). Another potential example is the spread of "take nothing" rules in the modern world through colonization by cultural groups that developed these rules (Hayek, 1944).

Moral rules might also spread due to their interaction with human emotional programs. Consider disgust, an emotion closely linked with moral condemnation. Given strategic debate over moral rules, the moral learning system might be designed to accept moral rules about behaviors that the person would not want to do anyway. Disgust can be thought of as a motivational system that steers people away from particular fitness costs associated with pathogens (Kurzban & Leary, 2001; Oaten, Stevenson, & Case, 2009), inbreeding (Lieberman et al., 2003, 2007), and toxins. We suggest that disgust acts as an input to moral cognition, increasing the perceived wrongness of an action. By condemning disgusting actions, people can favor those moral rules that do not impinge on their own interests. This idea can explain why rules are more likely to persist over time if they are about disgusting actions (Nichols, 2002) as well as why experiencing disgust increases condemnation of moral offenses (Eskine, Kacinik, & Prinz, 2011; Schnall, Haidt, Clore, & Jordan, 2008; Wheatley & Haidt, 2005). This model implies that different disgust reactions in different people can lead to moral disagreement. If, for example, some people view eating meat as disgusting, they will favor rules against it, but other people who do not experience disgust will tend to disagree with this moral stance (Fessler, Arguello, Mekdara, & Macias, 2003; Fessler & Navarrete, 2003; Rozin, Markwith, & Stoess, 1997). Similarly, individual differences in people's sexual aversion to incest influence their moral judgments of incest (Lieberman & Lobel, 2012).

Dynamic Coordination Explanations for Moral Phenomena

This section examines several important moral phenomena using the dynamic coordination theory. Because the present model posits a very different function for moral cognition than previous models of morality, it makes a number of predictions that diverge substantially from other models and explains a number of phenomena that are puzzling from the point of view of other accounts. Here we provide examples of how the theory can be applied rather than a comprehensive account of phenomena it can explain, which can be developed in future research.

Trials by Ordeal and Combat

Throughout history, a variety of cultural groups have used trials by ordeal and combat to resolve disputes (Kadri, 2005). In these practices, the moral wrongness of a disputant is determined by subjecting them to fire, boiling oil, freezing water, dueling, and other painful rituals. Emerging unscathed from the ordeal is interpreted as a sign of innocence, whereas injury signifies guilt. Evidence of the use of these trials can be found in the Old and New Testaments, the Code of Hammurabi, and in places ranging from India to Burma to England and continental Europe (Eidelberg,

1979). The practices continue in modern times, such as a case in Liberia in 2007 when four people accused of theft had red hot metal applied to their legs and were judged innocent because they did not flinch.

Why have so many human groups independently developed the practice of determining who is right and who is wrong by using arbitrary contests? Ordeals do not seem well designed to target punishment toward cheaters and other harmful individuals. Trials by ordeal and combat can be understood as dynamic coordination devices that allow bystanders to coordinate side-taking. By choosing sides based on the outcome of an ordeal, rather than status or personal alliances, bystanders can avoid despotism and discoordination. In this sense, trial by ordeal is functionally similar to a public coin flip used for coordination. Especially the more graphic ordeals, such as using fire or boiling oil, could provide a clear public signal. Viewing the outcome of the ordeal could allow everyone in the local community to come to the same conclusion about the person's guilt, knowing that others will reach the same conclusion as well. Trials by ordeal and combat might share this deep structural similarity with moral condemnation, which could explain why people interpret these trials in moral terms rather than as amoral techniques for dispute resolution.

Criminal Law

Theories of criminal law have traditionally focused on the functions of retribution and deterrence. We suggest that an additional approach is to view criminal law as a cultural extension of moral cognition that performs a dynamic coordination function. In conflicts, bystanders not only need to coordinate taking sides but also need to coordinate on the magnitude of punishment that fits the crime. The adversaries of the guilty party will tend to seek harsher punishments, whereas the perpetrator's family and friends will favor more lenient penalties. To preserve the possibility of coordination and the benefits of impartiality, these opposing groups need to coordinate their decisions about the magnitude of punishment. To accomplish this aim, individuals need to be able to rank moral violations and to assign punishments accordingly. Indeed, research shows that judgments of offense severity are highly consistent across individuals in domains such as harm and theft (Robinson & Kurzban, 2007). We suggest that moral intuitions for coordinating punishment decisions explain the emphasis on proportionality in criminal law.

The dynamic coordination function differs from traditional notions of retribution because the function is bystander coordination rather than satisfying a need for revenge or for perpetrators to suffer. We do not deny that people have revenge systems (McCullough et al., 2010, in press), but these mechanisms focus on violations against oneself or one's allies and often aim for maximum rather than proportional retaliation. In contrast, third-party punishment and the emphasis on proportionality seem better explained by a coordination function.

A coordination function also sharply contrasts with the utilitarian perspective, which holds that punishment should aim to deter harmful behavior. Deterrence theorists focus on two parameters as inputs: the damage done and the probability of detection. From a coordination perspective, the probability of detection, which is often unknown, does not provide a clear public signal for coordination, and hence, this is one variable that distinguishes these models. Indeed, contrary to the deterrence model, research suggests that punishment decisions are insensitive to the probability of detection (Baron & Ritov, 2009; Carlsmith, Darley, & Robinson, 2002).

A number of more specific features of criminal law might be explained by a coordination function. For instance, many legal scholars have debated the rationale for treating attempted crimes as less severe than completed crimes. This difference seems peculiar given that a failed attempt can result from seemingly irrelevant factors, such as skill with a gun or vagaries of chance. If, however, the law functions as an action-based coordination device, then attempted crimes do not generate a violating action—in legal terms the *actus reus*, or "guilty act"—and therefore do not produce a strong signal for action-based coordination.

Moralistic Religion

Why are the major organized religions so moralistic (Kurzban & DeScioli, 2009)? At the core of Judaism, the Ten Commandments specify actions that will be punished (ranked in order of severity). The symbol of Christianity is a crucifix for executing wrongdoers. Religious organizations moralize countless behaviors including art, science, usury, sexuality, and reproduction. Condemnation is directed not only at members but also at nonmembers, as illustrated by the long history of brutal persecutions for heresy and blasphemy (Levy, 1993). Today religious organizations continue to mint moral rules for new medical technologies such as contraceptives, stem cells, gene therapy, cloning, embryo cryopreservation, and artificial fertilization.

Gellner (1988) argued that one of the original functions of organized religion was to coordinate side-taking in disputes. With the development of agriculture, human groups were able to generate greater surpluses of resources. The opportunity to take and the need to defend these surpluses supported groups of warriors who specialized in fighting. These groups of fighters formed escalating alliances and counteralliances, which created the problem of choosing sides in disputes. Gellner argued that organized religion functioned to help local warlords choose sides in disputes by declaring which side was legitimate, or "right," and which side was "wrong," writing, "The sword may dominate, but the priests help crystallize cohesion among swordsmen. They arbitrate among them, and enable them to gang up successfully" (p. 276).

If organized religion functions to coordinate side-taking, then this could explain why morality is a central part of these organizations. Religious leaders might use the same strategic logic as moral cognition to solve a similar problem arising at the level of neighboring political groups. Many of the accoutrements of organized religion can be interpreted as instruments designed to amplify their coordination signal. Enormous cathedrals act as physical coordination points designed to broadcast uniquely powerful moral judgments. Devoted religious congregations that hold regular meetings can be understood as pervasive communication networks ever ready to receive condemnation signals from holy leaders. Religious rituals with public, repetitive, and coordinated movements generate common knowledge (Chwe, 2001). Elaborate supernatural stories written in holy books augment the uniqueness and attraction of sacred condemnation signals. Interestingly, for coordination purposes, it is not critical for people to believe in the divinity of the religion's authority. As Gellner wrote,

> This does not, once again, imply that individual thugs are so overawed in their hearts by the organization's claim to exclusive moral authority. . . . It merely requires that all of them publicly go through the appropriate motions of respect, and hence that each single one knows that the others will respect the doctrine, so that, in following it, he will join the larger battalions. (p. 96)

By claiming "moral authority," religious leaders commit what appears to be a conceptual contradiction. The moral strategy is distinct from an authority strategy in which people bandwagon and choose sides based on status. Morality is intrinsically antiauthority because it focuses on actions rather than a person's identity or status. Of course, this idea implies that if an individual or organization seeks unrivaled power, it will need to preempt or disable people's antidespotic moral countermoves. Perhaps one solution is to attempt to construe morality as derived from the particular status hierarchy of the organization. This is likely to be very difficult and might require special conceptual tricks such as identifying the highest ranking leader as an impartial nonhuman supernatural agent. This assertion of a divinity itself creates many suspicions and discrepancies that each need to be addressed with fictional elaborations of ever-increasing complexity. Still, if creative entrepreneurs can resolve these issues, then they can transition from a moral social order to an authoritarian social order in which the purveyors hold disproportionate power and influence.

The Action–Omission Distinction

A key characteristic of moral judgment is that morally violating omissions are judged as less wrong than commissions (Cushman et al., 2006; DeScioli, Bruening, & Kurzban, 2011; DeScioli, Christner, & Kurzban, 2011; P. Singer, 2009). The omission effect occurs across a variety of moral domains. For instance, it is judged less wrong to keep extra change from a cashier than to take extra change, and it is judged less wrong to allow illicit sex to occur than to initiate it (DeScioli et al., 2012). This distinctive information-processing pattern offers an important test case for theories about moral cognition as well as potentially providing insight into a number of other effects in moral cognition.

P. Singer (2009) has emphasized the global harms that result from the omission effect. People in modern affluent societies enjoy many luxuries while billions of other people suffer from preventable diseases, water shortage, and malnutrition. People's omissions of helping cause tremendous harm, yet moral judgment is largely silent on the issue, focused instead on sins of commission with consequences that are trivial in comparison. Leniency and indifference toward omissions is difficult to reconcile with the idea that morality functions to improve welfare. From the perspective of coordination, however, the signal generated for coordination is critical, not the harmfulness of a behavior. Ongoing, vague, and manifold omissions do not reliably produce discrete, distinct, and unique action representations that can be used to synchronize third-party condemnation. This is not to say that inaction cannot be moralized—this can be facilitated by redescribing, for instance, an omission of payment as (the "action" of) *breaking* a contract.

To understand omissions, it is important to distinguish decisions to omit (conscience) from judgments about others' omissions (condemnation). DeScioli, Christner, and Kurzban (2011) tested whether the omission effect in conscience is explained by reduced condemnation for omissions. Participants were given the opportunity to take money from someone by either commission or omission. The experimental manipulation varied whether a third party had the opportunity to punish the decision maker. Participants were more likely to choose omissions when they could be punished, showing that the choice of omissions is not an error but rather a strategy for avoiding condemnation.

This finding by itself leaves unexplained why condemnation is reduced for omissions. DeScioli, Bruening, and Kurzban (2011) proposed that omissions are condemned less harshly because it is more difficult to coordinate condemnation for omissions than commissions. They found that the omission effect is eliminated when the perpetrator presses a button that does nothing but leaves transparent and public physical evidence of their decision. This result suggests that it is not an absence of causality that explains the omission effect but the lack of material evidence available for coordination.

In sum, the omission effect in condemnation can be understood as a strategy for third-party coordination of punishment, whereas the omission effect in conscience can be understood as a counterstrategy used by actors to avoid coordinated punishment by third parties. Similar explanations might apply to other structural properties of moral cognition; that is, the omission effect might provide critical insight into the information-processing structure of many different effects in moral cognition. For example, people judge offenses as less wrong when they occur as a side-effect rather than as a means to an end—the well-known principle of double effect (Mikhail, 2007). This effect could be explained in terms of a lower quality signal for coordination, due to the way that action parsing mechanisms encode side-effects in action representations.

Conclusion

Humans are anxious about morality and for good reason. The moral rules that people create and negotiate perform critical functions in their fast-paced social lives. Disputes are constantly arising, and bystanders need to choose sides. Deciding based on relationships leads to discoordination and escalation, whereas deciding based on individual status concentrates power and enables exploitation. The moral strategy offers an alternative approach—action-based coordination—that can both synchronize side-taking and balance power. However, for the moral strategy to work, people need to agree, first, on whether to use morality at all and, second, on which moral rules to use. To maintain this consensus, humans need cognitive adaptations for constant vigilance against moral disagreement, including the construction of supporting belief systems that broadcast and reinforce shared moral commitments.

We suggest that this adaptive anxiety helps explain why scientific progress in understanding morality has been so difficult and unwelcome. In his treatise on human evolution, Darwin (1871) devoted two chapters to morality because he thought, with great foresight, that this trait would be the most difficult for people to accept as a product of evolution. Even the codiscoverer of natural selection, Alfred Russel Wallace, thought morality was the only human trait that did not evolve, instead favoring a supernatural explanation. Scientific investigation can change people's views about morality, specific moral rules, and supporting belief systems, and new ideas threaten disagreement and discoordination in people's interpersonal lives.

However, in the modern world scientific knowledge of morality is crucial. One reason is that moral condemnation can be extremely destructive. Moral judgment is designed for bystander coordination rather than promoting welfare and can therefore cause great harm both interpersonally and on a global scale. Prominent examples include honor killings of women (Appiah, 2010; United Nations, 2000), violence against homosexual people (Sarhan & Burke, 2009), mass imprisonment of drug users (Global Commission on Drug Policy, 2011; Marlatt, 1996), and failure to reduce HIV transmission by denying health services to sex workers (Rekart, 2005). A second reason is what morality obscures from view—people's omissions that result in tremendous harm. As P. Singer (2009) has emphasized, people's omissions leave billions of people without food, water, and medicine.

We can better understand moral cognition by applying the theory of evolution by natural selection (Darwin, 1871; Dawkins, 1976; Williams, 1966), computational theory of mind (Chomsky, 1957; Marr, 1982; Minsky, 1985; Pinker, 1997), and the evolutionary biology of strategy (Maynard Smith, 1982). Together, these foundations allow us to approach morality as an evolved, computational system that performs strategic functions.

We propose that moral condemnation is caused by an evolved suite of computational devices that are designed to implement a dynamic coordination strategy for choosing sides in other people's conflicts. Moral cognition takes action representations derived from conflict events as a primary input, using these actions as a source of public signals for bystander coordination. Observed actions are compared to a set of moral wrongs in which moralized actions are stored, revised, retrieved, and assigned magnitudes. The coordination rule is to choose sides against the disputant who has chosen the action with the greatest wrongness magnitude. This rule allows bystanders to synchronize side-taking while dynamically changing which individuals they support, thereby avoiding concentrating power in a few individuals.

It is difficult to overstate the power of the selective forces created by adaptations for condemnation. In this newly moral world, an otherwise powerful and well-connected individual could at any moment be seized upon and stoned to death by a crowd of angry moralists (see also Boehm, 1999). Few natural predators could be more dangerous and deadly than a moral mob composed of not only enemies but also family and friends. This lethal threat would select for adaptations designed to keep individuals on the right side of Kantian coordination rules. Through natural selection, humans became equipped with an increasingly sophisticated moral conscience for steering clear of moral mobs. These cognitive mechanisms would prospectively compare the individual's potential actions against the set of moral wrongs in order to avoid actions that could trigger coordinated condemnation by third parties.

The existence of action-based condemnation and conscience sets up a secondary strategic game in which individuals try to influence the set of moral rules to serve their interests. Rawlsian moral rules are favored by most people and hence are the most stable and universal. Strategic moral rules are favored by special interest groups, and they fluctuate within and between cultures as opposed groups struggle for control. Other moral rules can spread and persist due to the competitive advantages they provide to groups or because their social transmission is enhanced by the arousal of emotions such as disgust. The potency of all of these epidemiological processes derives from the fact that a dynamic coordination function does not critically depend on the content of moral rules (unlike other functions such as altruism), creating the potential for moral variability, secondary processes that shape this variability, and strong motivations to suppress other people's moral dissensions to establish a self-serving consensus.

The dynamic coordination theory can account for many features of moral cognition that are left unexplained by previous theories. The theory depicts moral cognition as highly sophisticated in the computations that it performs and the strategies that it enacts, more so than is generally assumed by previous theories. The human mind is the most advanced computational control system in the universe, and moral cognition is one of its most critical processes, running in nearly all of our social interactions and many of our private reflections. We suggest that moral computations are as impressively engineered as cognitive mechanisms for language or vision. By continuing to apply the logic of reverse engineering, we can uncover the structure of moral thought, unlock its strategic design, and solve the mysteries of morality.

References

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

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63,* 368–378. doi:10.1037/0022-3514.63.3.368

Appiah, K. A. (2010). *The honor code: How moral revolutions happen.* New York, NY: Norton.

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour, 77,* 991–1004. doi:10.1016/j.anbehav.2009.02.010

Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics, 1,* 67–96. doi:10.1016/0304-4068(74)90037-8

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics, 4,* 1236–1239. doi:10.1214/aos/1176343654

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5,* 171–178. doi:10.1016/S1364-6613(00)01615-6

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior, 27,* 325–344. doi:10.1016/j.evolhumbehav.2006.01.003

Baron, J., & Ritov, I. (2009). The role of probability of detection in judgments of punishment. *Journal of Legal Analysis, 1,* 553–590. doi:10.1093/jla/1.2.553

Bentham, J. (1952). A defence of usury. In W. Stark (Ed.), *Jeremy Bentham's economic writings.* London, England: Allen & Unwin. (Original work published 1787)

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442,* 912–915. doi:10.1038/nature04981

Binmore, K. (2005). *Natural justice.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195178111.001.0001

Binmore, K. (2007). The origins of fair play. In P. J. Marshall (Ed.), *Proceedings of the British Academy* (Vol. 151, pp. 151–193). Oxford, England: Oxford University Press. doi:10.5871/bacad/9780197264249.003.0006

Binmore, K., & Samuelson, L. (1994). An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics, 150,* 45–63.

Black, D. (1998). *The social structure of right and wrong.* New York, NY: Academic Press.

Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior.* Cambridge, MA: Harvard University Press.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United of America, 100,* 3531–3535. doi:10.1073/pnas.0630443100

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

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13,* 171–195. doi:10.1016/0162-3095(92)90032-Y

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures.* New York, NY: Oxford University Press.

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology, 67,* 773–789. doi:10.1037/0022-3514.67.5.773

Buss, D. M. (2006). Strategies of human mating. *Psychological Topics, 15,* 239–260.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83,* 284–299. doi:10.1037/0022-3514.83.2.284

Chapais, B., Girard, M., & Primi, G. (1991). Non-kin alliances, and the stability of matrilineal dominance relations in Japanese macaques. *Animal Behaviour, 41,* 481–491. doi:10.1016/S0003-3472(05)80851-6

Chaux, E. (2005). Role of third parties in conflicts among Colombian children and early adolescents. *Aggressive Behavior, 31,* 40–55. doi:10.1002/ab.20031

Cheney, D., & Seyfarth, R. (1990). *How monkeys see the world.* Chicago, IL: University of Chicago Press.

Cheney, D., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind.* Chicago, IL: University of Chicago Press.

Chomsky, N. (1957). *Syntactic structures.* The Hague, the Netherlands: Mouton.

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm–psychology and the emergence of human prosociality. *Trends in Cognitive Sciences, 15,* 218–226. doi:10.1016/j.tics.2011.03.003

Chwe, M. S.-Y. (2001). *Rational ritual: Culture, coordination, and common knowledge.* Princeton, NJ: Princeton University Press.

Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics, 9,* 265–279. doi:10.1007/s10683-006-9127-z

Connor, R. C. (2007). Dolphin social intelligence: Complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362,* 587–602. doi:10.1098/rstb.2006.1997

Cooney, M. (1998). *Warriors and peacemakers: How third parties shape violence.* New York: New York University Press.

Cooney, M. (2003). The privatization of violence. *Criminology, 41,* 1377–1406. doi:10.1111/j.1745-9125.2003.tb01023.x

Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science, 17,* 1082–1089. doi:10.1111/j.1467-9280.2006.01834.x

Daly, M., & Wilson, M. (1988). *Homicide.* New York, NY: Aldine de Gruyter.

Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology, 41,* 525–556. doi:10.1146/annurev.ps.41.020190.002521

Darwin, C. (1871). *Descent of man, and selection in relation to sex.* New York, NY: Appleton.

Dawkins, R. (1976). *The selfish gene.* Oxford, England: Oxford University Press.

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory, 33,* 145–167. doi:10.1007/s00199-007-0212-0

DeScioli, P. (2008). *Investigations into the problems of moral cognition* (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia, PA.

DeScioli, P., Asao, K., & Kurzban, R. (2012). *Omissions and byproducts across moral domains.* Manuscript in preparation.

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior, 32,* 204–215. doi:10.1016/j.evolhumbehav.2011.01.003

DeScioli, P., Christner, J., & Kurzban, R. (2011). The omission strategy. *Psychological Science, 22,* 442–446. doi:10.1177/0956797611400616

DeScioli, P., Gilbert, S., & Kurzban, R. (2012). Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry, 23,* 143–149.

DeScioli, P., & Kimbrough, E. (2012). *Alliance formation and the structure of conflict.* Manuscript in preparation.

DeScioli, P., & Kurzban, R. (2009a). The alliance hypothesis for human friendship. *PLoS ONE, 4,* e5802. doi:10.1371/journal.pone.0005802

DeScioli, P., & Kurzban, R. (2009b). Mysteries of morality. *Cognition, 112,* 281–299. doi:10.1016/j.cognition.2009.05.008

DeScioli, P., Kurzban, R., Koch, E. N., & Liben-Nowell, D. (2011). Best friends: Alliances, friend ranking, and the MySpace social network. *Perspectives on Psychological Science, 6,* 6–8. doi:10.1177/1745691610393979

de Waal, F. B. M. (1982). *Chimpanzee politics: Power and sex among apes.* Baltimore, MD: Johns Hopkins University Press.

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

Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don't punish. *Nature, 452,* 348–351. doi:10.1038/nature06723

Dunbar, R. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8,* 100–110. doi:10.1037/1089-2680.8.2.100

Eidelberg, S. (1979). Trial by ordeal in medieval Jewish history: Laws, customs and attitudes. *Proceedings of the American Academy for Jewish Research, 46/47,* 105–120. doi:10.2307/3622459

Engh, A. L., Siebert, E. R., Greenberg, D. A., & Holekamp, K. E. (2005). Patterns of alliance formation and post-conflict aggression indicate spotted hyenas recognize third party relationships. *Animal Behaviour, 69,* 209–217. doi:10.1016/j.anbehav.2004.04.013

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological Science, 22,* 295–299. doi:10.1177/0956797611398497

Fessler, D. M. T., Arguello, A. P., Mekdara, J. M., & Macias, R. (2003). Disgust sensitivity and meat consumption: A test of an emotivist account of moral vegetarianism. *Appetite, 41,* 31–41. doi:10.1016/S0195-6663(03)00037-0

Fessler, D. M. T., & Navarrete, C. D. (2003). Meat is good to taboo: Dietary proscriptions as a product of the interaction of psychological mechanisms and social processes. *Journal of Cognition and Culture, 3,* 1–40. doi:10.1163/156853703321598563

Gardner, A., & West, S. A. (2004). Cooperation and punishment, especially in humans. *American Naturalist, 164,* 753–764. doi:10.1086/425623

Gaulin, S. J. C., McBurney, D. H., & Brakeman-Wartell, S. L. (1997). Matrilateral biases in the investment of aunts and uncles: A consequence and measure of paternity uncertainty. *Human Nature, 8,* 139–151. doi:10.1007/s12110-997-1008-4

Gellner, E. (1988). *Plough, sword and book: The structure of human history.* Chicago, IL: University of Chicago Press.

Gigerenzer, G. (2010). Moral satisficing: Rethinking moral behavior as

bounded rationality. *Topics in Cognitive Science, 2,* 528–554. doi: 10.1111/j.1756-8765.2010.01094.x

Gintis, H., Smith, A. E., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology, 213,* 103–119. doi:10.1006/jtbi.2001.2406

Global Commission on Drug Policy. (2011). *War on drugs: Report of the Global Commission on Drug Policy.* New York, NY: Open Society Institute.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition, 106,* 1339–1366. doi:10.1016/j.cognition.2007.06.007

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96,* 505–520. doi:10.1037/a0013748

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23,* 101–124.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 3, pp. 35–79). Cambridge, MA: MIT Press.

Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion, 7,* 853–868. doi:10.1037/1528-3542.7.4.853

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814–834. doi:10.1037/0033-295X.108.4.814

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316,* 998–1002. doi:10.1126/science.1137651

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20,* 98–116. doi:10.1007/s11211-007-0034-z

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology, 31,* 191–221. doi:10.1111/j.1559-1816.2001.tb02489.x

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133,* 55–66. doi:10.1162/0011526042365555

Haidt, J., & Joseph, C. (2008). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Vol. 3. Foundations and the future* (pp. 367–392). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195332834.003.0019

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65,* 613–628. doi:10.1037/0022-3514.65.4.613

Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology, 33,* 1–36. doi:10.1111/j.1559-1816.2003.tb02071.x

Hamilton, W. D. (1964). The genetic evolution of social behaviour. I. *Journal of Theoretical Biology, 7,* 1–16. doi:10.1016/0022-5193(64)90038-4

Harcourt, A. H. (1992). Coalitions and alliances: Are primates more complex than non-primates? In A. H. Harcourt & F. B. M. de Waal (Eds.), *Coalitions and alliances in humans and other animals* (pp. 445–471). New York, NY: Oxford University Press.

Harcourt, A. H., & de Waal, F. B. M. (Eds.). (1992). *Coalitions and alliances in humans and other animals.* New York, NY: Oxford University Press.

Hauser, M. D. (2006). *Moral minds.* New York, NY: HarperCollins.

Hayek, F. (1944). *The road to serfdom.* Chicago, IL: University of Chicago Press.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208,* 79–89. doi:10.1006/jtbi.2000.2202

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science, 319,* 1362–1367. doi:10.1126/science.1153808

Joyce, R. (2006). *The evolution of morality.* Cambridge, MA: MIT Press.

Kadri, S. (2005). *The trial: A history, from Socrates to O. J. Simpson.* New York, NY: Random House.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of intuitive judgment: Extensions and applications* (pp. 49–81). New York, NY: Cambridge University Press.

Kant, I. (1993). *Grounding for the metaphysics of morals* (J. W. Ellington, Trans.). Indianapolis, IN: Hackett. (Original work published 1785)

Killen, M. (2007). Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science, 16,* 32–36. doi:10.1111/j.1467-8721.2007.00470.x

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 12577–12580. doi:10.1073/pnas.0705345104

Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences, 9,* 357–359. doi:10.1016/j.tics.2005.06.011

Krebs, J. R., & Davies, N. B. (1993). *An introduction to behavioral ecology* (3rd ed.). Oxford, England: Blackwell.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12,* 72–79. doi: 10.1016/j.tics.2007.11.004

Kurzban, R. (2010). *Why everyone (else) is a hypocrite: Evolution and the modular mind.* Princeton, NJ: Princeton University Press.

Kurzban, R., & DeScioli, P. (2009). *Adaptationist punishment in humans.* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1368784##

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior.* Advance online publication. doi:10.1016/j.evolhumbehav.2011.11.002

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior, 28,* 75–84. doi:10.1016/j.evolhumbehav.2006.06.001

Kurzban, R., Dukes, A., & Weeden, J. (2010). Sex, drugs and moral goals: Reproductive strategies and views about recreational drugs. *Proceedings of the Royal Society B: Biological Sciences, 277,* 3501–3508. doi: 10.1098/rspb.2010.0608

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127,* 187–208. doi:10.1037/0033-2909.127.2.187

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 15387–15392. doi:10.1073/pnas.251541498

Levy, L. W. (1993). *Blasphemy: Verbal offense against the sacred, from Moses to Salman Rushdie.* New York, NY: Knopf.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology, 5,* 289–305.

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society B: Biological Sciences, 270,* 819–826. doi:10.1098/rspb.2002.2290

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature, 445,* 727–731. doi:10.1038/nature05510

Lieberman, D., & Lobel, T. (2012). Kinship on the kibbutz: Coresidence duration predicts altruism, personal sexual aversions and moral attitudes among communally reared peers. *Evolution and Human Behavior, 33,* 26–34. doi:10.1016/j.evolhumbehav.2011.05.002

Marlatt, G. A. (1996). Harm reduction: Come as you are. *Addictive Behaviors, 21,* 779–788. doi:10.1016/0306-4603(96)00042-1

Marr, D. (1982). *Vision.* New York, NY: Freeman.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge, England: Cambridge University Press.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2010). Evolved mechanisms for revenge and forgiveness. In P. R. Shaver & M. Mikulincer (Eds.), *Human aggression and violence: Causes, manifestations, and consequences* (pp. 221–239). Washington, DC: American Psychological Association. doi:10.1037/12346-012

McCullough, M. E., Kurzban, R., & Tabak, B. A. (in press). The evolution of revenge and forgiveness. *Behavioral and Brain Sciences.*

McWilliams, S. A. (1996). Accepting the invitational. In B. M. Walker, J. Costigan, L. L. Viney, & B. Warren (Eds.), *Personal construct theory: A psychology for the future* (pp. 57–78). Melbourne, Australia: Australian Psychological Society.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34,* 57–74. doi:10.1017/S0140525X10000968

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11,* 143–152. doi:10.1016/j.tics.2006.12.007

Minsky, M. L. (1985). *The society of mind.* New York, NY: Simon & Schuster.

Mivart, S. G. (1871). *On the genesis of species.* London, England: Macmillan.

Mock, D. W. (2004). *More than kin and less than kind: The evolution of family conflict.* Cambridge, MA: Oxford University Press.

Newton-Fisher, N. E. (2002). Relationships of male chimpanzees in the Budongo Forest, Uganda. In C. Boesch, G. Hohmann, & L. F. Marchant (Eds.), *Behavioral diversity in chimpanzees and bonobos* (pp. 125–137). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511606397.013

Nichols, S. (2002). On the genealogy of norms: A case for the role of emotion in cultural evolution. *Philosophy of Science, 69,* 234–255. doi:10.1086/341051

Nozick, R. (1974). *Anarchy, state, and utopia.* New York, NY: Basic Books.

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin, 135,* 303–321. doi:10.1037/a0014823

Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology, 47,* 223–243. doi:10.1016/0022-5193(74)90111-8

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296,* 1321–1323. doi:10.1126/science.1070223

Phillips, S., & Cooney, M. (2005). Aiding peace, abetting violence: Third parties and the management of conflict. *American Sociological Review, 70,* 334–354. doi:10.1177/000312240507000207

Piazza, J., & Bering, J. M. (2008). The effects of perceived anonymity on altruistic punishment. *Evolutionary Psychology, 6,* 487–501.

Pinker, S. (1997). *How the mind works.* New York, NY: Norton.

Pinker, S. (2007). *The stuff of thought: Language as a window onto human nature.* New York, NY: Viking Press.

Posada, R., & Wainryb, C. (2008). Moral development in a violent society: Colombian children's judgments in the context of survival and revenge. *Child Development, 79,* 882–898. doi:10.1111/j.1467-8624.2008.01165.x

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

Rekart, M. L. (2005). Sex-work harm reduction. *Lancet, 366,* 2123–2134. doi:10.1016/S0140-6736(05)67732-X

Riddle, J. M. (1997). *Eve's herbs: A history of contraception and abortion in the West.* Cambridge, MA: Harvard University Press.

Ridley, M. (1996). *The origins of virtue.* London, England: Viking.

Robinson, P. H., & Kurzban, R. (2007). Concordance and conflict in intuitions of justice. *Minnesota Law Review, 91,* 1829–1907.

Robinson, P. H., Kurzban, R., & Jones, O. D. (2007). The origins of shared intuitions of justice. *Vanderbilt Law Review, 60,* 1631–1688.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605. doi:10.1016/0010-0285(75)90024-9

Ross, H. S., & Den Bak-Lammers, I. M. (1998). Consistency and change in children's tattling on their siblings: Children's perspectives on the moral rules and procedures of family life. *Social Development, 7,* 275–300. doi:10.1111/1467-9507.00068

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research, 15,* 165–184. doi:10.1023/A:1019923923537

Rozin, P. (1999). The process of moralization. *Psychological Science, 10,* 218–221. doi:10.1111/1467-9280.00139

Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science, 8,* 67–73. doi:10.1111/j.1467-9280.1997.tb00685.x

Sarhan, A., & Burke, J. (2009, September 13). How Islamist gangs use Internet to track, torture and kill Iraqi gays. *The Guardian.* Retrieved from http://www.guardian.co.uk/world/2009/sep/13/iraq-gays-murdered-militias

Schelling, T. C. (1960). *The strategy of conflict.* Cambridge, MA: Harvard University Press.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin, 34,* 1096–1109. doi:10.1177/0146167208317771

Schülke, O., Bhagavatula, J., Vigilant, L., & Ostner, J. (2010). Social bonds enhance reproductive success in male macaques. *Current Biology, 20,* 2207–2210. doi:10.1016/j.cub.2010.10.058

Seyfarth, R. M., & Cheney, D. L. (2012). The evolutionary origins of friendship. *Annual Review of Psychology, 63,* 153–177. doi:10.1146/annurev-psych-120710-100337

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "Big Three" of morality (autonomy, community, and divinity), and the "Big Three" explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York, NY: Routledge.

Sigmund, K. (2007). Punish or perish? Retaliation and collaboration among humans. *Trends in Ecology and Evolution, 22,* 593–600.

Silk, J. B., Alberts, S. C., & Altmann, J. (2004). Patterns of coalition formation by adult female baboons in Amboseli, Kenya. *Animal Behaviour, 67,* 573–582. doi:10.1016/j.anbehav.2003.07.001

Simoons, F. J. (1994). *Eat not this flesh: Food avoidances from prehistory to the present.* Madison: University of Wisconsin Press.

Singer, P. (2009). *The life you can save: Acting now to end world poverty.* New York, NY: Random House.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E. Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439,* 466–469. doi:10.1038/nature04271

Smetana, J. G., & Braeges, J. L. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly, 36,* 329–346.

Smetana, J. G., Schlagman, N., & Adams, P. W. (1993). Preschool children's judgments about hypothetical and actual transgressions. *Child Development, 64,* 202–214. doi:10.2307/1131446

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations.* London, England: Strahan & Cadell.

Smith, M. S., Kish, B. J., & Crawford, C. B. (1987). Inheritance of wealth

as human kin investment. *Ethology and Sociobiology, 8,* 171–182. doi:10.1016/0162-3095(87)90042-2

Snyder, G. H. (1984). The security dilemma in alliance politics. *World Politics, 36,* 461–495. doi:10.2307/2010183

Snyder, G. H. (1997). *Alliance politics.* Ithaca, NY: Cornell University Press.

Tetlock, P. (2000). Cognitive biases and organizational correctives: Do both disease and cure depend on the politics of the beholder? *Administrative Science Quarterly, 45,* 293–326. doi:10.2307/2667073

Tetlock, P. E. (2002). Social-functionalist metaphors for judgment and choice: The intuitive politician, theologian, and prosecutor. *Psychological Review, 109,* 451–471. doi:10.1037/0033-295X.109.3.451

Thomas, C. W., Cage, R. J., & Foster, S. C. (1976). Public opinion on criminal law and legal sanctions: An examination of two conceptual models. *Journal of Criminal Law and Criminology, 67,* 110–116. doi:10.2307/1142462

Tiger, L. (1969). *Men in groups.* New York, NY: Random House.

Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human morality and sociality: Evolutionary and comparative perspectives* (pp. 191–234). New York, NY: Palgrave Macmillan.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46,* 35–57. doi:10.1086/406755

Turiel, E. (1998). The development of morality. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–932). Hoboken, NY: Wiley.

United Nations. (2000). *Civil and political rights, including questions of disappearances and summary executions.* New York, NY: Commission on Human Rights, United Nations.

van Prooijen, J.-W. (2006). Retributive reactions to suspected offenders: The importance of social categorizations and guilt probability. *Personality and Social Psychology Bulletin, 32,* 715–726. doi:10.1177/0146167205284964

Waldmann, M. R., & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science, 18,* 247–253. doi:10.1111/j.1467-9280.2007.01884.x

Weeden, J. (2003). *Genetic interests, life histories, and attitudes towards abortion* (Unpublished doctoral dissertation). University of Pennsylvania.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16,* 780–784. doi:10.1111/j.1467-9280.2005.01614.x

Whitehouse, M. E. A., & Jaffe, K. (1996). Ant wars: Combat strategies, territory and nest defense in the leaf-cutting ant *Atta laevigata. Animal Behaviour, 51,* 1207–1217. doi:10.1006/anbe.1996.0126

Whiten, A., & Byrne, R. W. (Eds.). (1997). *Machiavellian intelligence II: Evaluations and extensions.* Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511525636

Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi Bushmen: A case of strong reciprocity? *Human Nature, 16,* 115–145. doi:10.1007/s12110-005-1000-9

Williams, G. C. (1966). *Adaptation and natural selection.* Princeton, NJ: Princeton University Press.

Wright, R. (1994). *The moral animal: Why we are, the way we are: The new science of evolutionary psychology.* New York, NY: Pantheon.

Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition, 119,* 166–178. doi:10.1016/j.cognition.2011.01.004

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16,* 80–84. doi:10.1111/j.1467-8721.2007.00480.x

Zimbardo, P. (2007). *The Lucifer effect: Understanding how good people turn evil.* New York, NY: Random House.

Received November 1, 2011
Revision received March 5, 2012
Accepted May 8, 2012 ■