



## Mysteries of morality

Peter DeScioli\*, Robert Kurzban

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, United States

### ARTICLE INFO

#### Article history:

Received 3 October 2008

Revised 5 May 2009

Accepted 15 May 2009

#### Keywords:

Third-party condemnation

Evolution of morality

Moral impartiality

Moral judgment

Moralistic punishment

### ABSTRACT

Evolutionary theories of morality, beginning with Darwin, have focused on explanations for altruism. More generally, these accounts have concentrated on conscience (self-regulatory mechanisms) to the neglect of condemnation (mechanisms for punishing others). As a result, few theoretical tools are available for understanding the rapidly accumulating data surrounding third-party judgment and punishment. Here we consider the strategic interactions among actors, victims, and third-parties to help illuminate condemnation. We argue that basic differences between the adaptive problems faced by actors and third-parties indicate that actor conscience and third-party condemnation are likely performed by different cognitive mechanisms. Further, we argue that current theories of conscience do not easily explain its experimentally demonstrated insensitivity to consequences. However, these results might be explicable if conscience functions, in part, as a defense system for avoiding third-party punishment. If conscience serves defensive functions, then its computational structure should be closely tailored to the details of condemnation mechanisms. This possibility underscores the need for a better understanding of condemnation, which is important not only in itself but also for explaining the nature of conscience. We outline three evolutionary mysteries of condemnation that require further attention: third-party judgment, moralistic punishment, and moral impartiality.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Current evolutionary theories of morality address why people are motivated to perform certain actions – such as helping others – and to avoid certain actions – such as incest (e.g., Haidt, 2007). These theories, by themselves, do not explain why people think that others should be punished for moral violations. They do not explain third-party moral judgment, moralistic punishment, and moral impartiality, as symbolized by the scale, sword, and blindfold of Lady Justice. These patterns of human cognition are species-typical and complex (see Sections 6–8), suggesting that they serve some evolved function. However, the adaptive functions of moral condemnation remain mysterious.

Darwin (1871) initiated the evolutionary biology of morality in *The Descent of Man*, devoting two chapters to the subject. Darwin claimed that the “foundation-stone” of morality is sympathy, a social instinct also found in non-human animals. He argued that sympathy evolved into morality by group selection – groups with morality out-competed groups without morality. Darwin recognized that an important problem for his theory was the wide variety in moral rules, especially useless and harmful rules, such as Hindu food taboos. Darwin concluded that morality was designed to benefit the group and that detrimental rules are errors.

Modern biological accounts of morality have made several key revisions of Darwin’s theory. First, while some researchers continue to appeal to group selection (e.g., Haidt, 2007), other theorists have shifted focus to other evolutionary pathways to altruism (e.g., de Waal, 1996; Ridley, 1996; Wright, 1994), especially kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971). In one important variety of altruism-based theories, the evolution

\* Corresponding author. Present address: Economic Science Institute, Chapman University, Orange, CA 92866, United States. Tel.: +1 215 913 8569; fax: +1 215 898 7301.

E-mail addresses: [descioli@psych.upenn.edu](mailto:descioli@psych.upenn.edu), [pdescioli@yahoo.com](mailto:pdescioli@yahoo.com) (P. DeScioli).

of cooperation hinges on punishment, occurring when groups with norm-enforcing punishers out-compete other groups (e.g., Boyd & Richerson, 1992, 2005; Fehr, Fischbacher, & Gächter, 2002; Gintis, 2000). Second, several researchers have departed from Darwin, arguing that the diversity of moral rules points to multiple, independently evolved psychological systems underlying morality (Haidt & Joseph, 2004, 2008; Hauser, 2006; Krebs & Janicki, 2004; Miller, 2007; Stich, 2006; Wilson, 1993), including systems associated with suffering, hierarchy, reciprocity, honesty, self-control, violence, fairness, and incest avoidance.

We contribute an adaptationist analysis of morality, focusing on moral condemnation. Adaptationism (Williams, 1966) leverages the tight link between structure and function in biological systems. This link allows inferences about structure from function (e.g., predicting the microstructure of water strider legs from their function, Gao & Jiang, 2004), and also allows inferences about function from structure (e.g., material properties of the Achilles tendon show design for running, not walking; Bramble & Lieberman, 2004). In Section 2, we consider whether moral cognition appears sufficiently distinct and coherent that it makes sense to investigate what functions the system performs. In Sections 3 and 4, we present two main adaptationist arguments about the functional organization of moral cognition. First, inferring structure from function, we draw on the strategic dynamics among perpetrator, victim, and third-party condemner to argue for two distinct components of moral cognition: One subsystem regulates one's own behavior (conscience) while another mechanism is specialized for judging others (condemnation). Second, inferring function from structure, we consider empirical evidence showing that insensitivity to welfare consequences – nonconsequentialism – is pervasive in moral judgment. Briefly, we will argue that nonconsequentialism challenges the prevailing view that moral conscience is designed to promote the welfare of family, friends, or groups.

In Section 5, motivated by the preceding arguments, we outline a specific functional relationship between conscience and condemnation components of moral cognition. In particular, we suggest that condemnation mechanisms causally precede conscience, and that conscience functions, at least in part, as a defense system designed to avoid attacks from third-party condemners. This view implies that moral conscience cannot be fully explained independent of condemnation.

If so, then much about the evolution of morality remains mysterious. There is little work and still less agreement on the function of moral condemnation systems (but see Boyd & Richerson, 1992, 2005). Sections 6–8 describe three problems of moral condemnation – third-party judgment, punishment, and impartiality.

## 2. The moral dimension

Philosophers and laypeople alike debate whether particular actions are right or wrong. These inquiries concern the positions of actions along the moral dimension from good to evil. For instance, Plato (4th century BC/2004) condemned suicide as “wrong” while Hume (1783) positioned

suicide on the “right” side of the moral spectrum. These moral debates take for granted the moral dimension itself (Macnamara, 1991), i.e., the cognitive capacity to assign moral values to actions.

This paper examines “morality,” meaning phenomena surrounding the concepts “right” and “wrong.” Specifically, we focus on the moral dimension rather than the positions of actions along moral lines. The evolutionary explanations for why particular actions are assigned particular moral values have received considerable attention (e.g., Alexander, 1987; Darwin, 1871; de Waal, 1996; Haidt & Joseph, 2004, 2008; Hauser, 2006; Krebs & Janicki, 2004; Lieberman, Tooby, & Cosmides, 2003, 2007; Ridley, 1996; Wilson, 1993; Wright, 1994). However, why incest is perceived as “wrong” and reciprocity “right” is a different question from why human minds possess a right-wrong spectrum upon which events like incest and reciprocity might fall.

The moral dimension refers broadly to the capacity to experience the world as morally textured and differentiated. What functions might be served by the cognitive systems that generate moral distinctions? Before addressing this question, this section considers whether the cognitive systems underlying the moral dimension are sufficiently distinct and coherent that they might perform some evolved function.

### 2.1. Irreducible Moore-ality

Is morality a distinct phenomenon? If right and wrong are synonymous with or derived from other concepts (benefit/harm, obedience/disobedience, etc.) then the explanation of morality will necessarily be tied to explanations of these phenomena. Some, for example, regard morality as equivalent to concern for welfare. The moral philosopher Gert (2005) wrote, “It should be apparent that by an evil, I mean a harm... by a good, I mean a benefit” (p. 91). In stark contrast, Moore (1903) argued that right and wrong are among the “innumerable objects of thought which are themselves incapable of definition, because they are the ultimate terms of reference to which whatever is capable of definition must be defined” (p. 10, emphasis original). Similarly, some cognitive psychologists regard moral rightness and wrongness as conceptual primitives (Macnamara, 1991; Mikhail, 2007).

The idea that moral concepts are derived from welfare concepts is inconsistent with important moral phenomena, including many sexual prohibitions (e.g., Stengers & Van Neck, 2001) and food taboos (Douglas, 1966; Fessler & Navarrete, 2003; Simoons, 1994) that are not explained by welfare gains. The ethnographic record is replete with moral rules that are understood not in terms of welfare or harm but in relation to purity, authority, divinity, loyalty, etc. (Haidt, 2007; Haidt & Joseph, 2008; Shweder, Much, Mahapatra, & Park, 1997). Even when people explain their moral judgments in terms of welfare, these justifications are not always accurate. Laboratory evidence shows that, at least in some cases, stated welfare concerns do not drive moral judgments but instead reflect post hoc rationalization (Haidt, 2001). In studies of harmless violations (e.g., eating one's dog following a natural death), judgments of harmfulness did not predict moral judgments

(Haidt & Hersh, 2001; Haidt, Koller, & Dias, 1993). More starkly, people maintained that violations like incest were wrong even in situations when no harm could result, appearing “morally dumbfounded,” unable to explain their own judgments (Haidt, 2001). Similarly, Tetlock (2000) presented subjects with policies such as markets for body organs; even after participants’ welfare objections were satisfied with additional provisions, roughly half continued to morally condemn the policies. These findings caution that welfare justifications cannot be simply taken at face value: People generate welfare justifications even when their welfare judgments do not explain their moral judgments.

Further, actions that yield clear welfare gains are sometimes rejected for moral reasons. In the footbridge dilemma (see below), 90% of people thought it was impermissible to kill one person to save five others (Hauser, 2006). Similarly, participants rejected welfare gains produced by trade-offs involving “protected values,” such as making money by destroying natural forests (Baron & Spranca, 1997; Tetlock, 2003). These observed dissociations among moral and welfare judgments show that the concept “morally wrong” cannot be simply equivalent to “harmful.”<sup>1</sup>

Right and wrong also do not seem to be derived from authority concepts like permissible/forbidden, or lawful/unlawful (Macnamara, 1991). Key evidence comes from developmental psychology. Contradicting early claims by Piaget (1932) and Kohlberg (1981), children are skeptical of authority in the moral realm. Children as young as three regard moral rules (but not social conventions) as independent of authority, custom, and explicit rules (Smetana & Braeges, 1990; Smetana, Schlagman, & Adams, 1993; Turiel, 1998). Children often disobey authority when commands will result in harm (Laupa & Turiel, 1986). They also reject the legitimacy of rules applied to personal issues such as hairstyle (Nucci, 1981). Further, children do not view violators of unjust laws as blameworthy (Helwig & Jasiobedzka, 2001). Even young children appear to appreciate the motto of the University of Pennsylvania, *Leges sine moribus vanae*: Laws without morals are useless.

In short, moral concepts are not reducible to welfare concepts or authority concepts. People can conceive of welfare-increasing wrongs, welfare-decreasing rights, wrongful obedience, and rightful disobedience (unlike unimaginable entities such as a four-sided triangle). The concepts of right and wrong, being irreducible, require explanation. Understanding these concepts, and the underlying computational systems, is a central aim of moral psychology.

## 2.2. Coherence across different types of violations

Cross-culturally, moral prohibitions concern diverse phenomena, including violence, sex, food, communication,

and witchcraft (e.g., Barton, 1919; Shweder, Mahapatra, & Miller, 1987; Shweder et al., 1997). In turn, different moral rules surround different biological problems (e.g., dominance hierarchies, mate guarding, foraging), which in many species are each handled by specialized cognitive systems (Krebs & Davies, 1993). This might suggest that “morality” is an umbrella term for a collection of different psychological systems (e.g., Haidt & Joseph, 2004, 2008; Hauser, 2006; Krebs & Janicki, 2004; Miller, 2007; Stich, 2006; Wilson, 1993). “Morality” might be an artificial grouping rather than a natural kind, potentially explaining why moral concepts seem irreducible.

However, there is coherence in moral cognition that cuts across content domains. Moral condemnation, in particular, shows the same fundamental properties across diverse offenses *viz.*, third-party judgment, moralistic punishment, and (the pretense of) moral impartiality (reviewed in Sections 6–8). Also, the causal and intentional structure of moral actions seems to be evaluated similarly across offense types, perhaps indicating a common representational format (Mikhail, 2007). Further, moral cognition differs from other types of normative judgments (e.g., convention, authority, precaution) in patterns of reasoning (Fiddick, 2004; Turiel, 1998) and, more visibly, in emotional and behavioral consequences. Last, research on the emotions elicited by moral violations (e.g., guilt, shame, righteous anger) has found that “it is not the events, *per se*, that determine which emotion is experienced but rather how events are appraised” (i.e., appraisals about causality and actors’ dispositions; Tracy & Robins, 2006, p. 1339; see also Tangney, Stuewig, & Mashek, 2007). The emotion of guilt, for instance, can be elicited by assault, theft, lying, infidelity, etc. (Keltner & Buswell, 1996).

Coherence seems apparent when the moral value of an action changes, and the many accoutrements of moral judgment are activated or deactivated in a coordinated fashion. “Moralization,” when “objects or activities that were previously morally neutral acquire a moral component” (Rozin, Markwith, & Stoess, 1997, p. 67), and the reverse, “amoralization,” show that actions undergo discrete shifts in moral status. For instance, a precautionary rule might shift to a moral rule, or vice versa (e.g., smoking, Rozin & Singh, 1999). Importantly, newly minted rules recruit the full suite of moral machinery, including prohibition, outrage, censure, overjustification, internalization, and enhanced parent-to-child transmission (Rozin, 1999). It is as though moral cognition has an “insert here” parameter, processing diverse moral rules with the same computational architecture (cf. Pinker, 1999).

Features of moral cognition that cohere *across* content domains suggest limitations to theories that tie the evolved function of morality to *particular* content domains. Previous work has focused on some particular subset of moral phenomena, dismissing moral variety as error (e.g., Darwin, 1871, pp. 95–99). Most often theorists concentrate on moral rules about altruism (Alexander, 1987; Darwin, 1871; de Waal, 1996; Ridley, 1996; Wright, 1994), or a collection of domains, such as incest, trade, honesty, and adultery (e.g., Haidt & Joseph, 2004, 2008; Hauser, 2006; Krebs & Janicki, 2004; Lieberman et al., 2003, 2007; Miller, 2007; Wilson, 1993). If, however, there are mechanisms that

<sup>1</sup> The observed distinction between moral and welfare concepts does not, by itself, imply that moral cognition is insensitive to (conscious or unconscious) information about harm (e.g., the deleterious effects of incest on reproduction), but only that the *concept* of harm turns on different information from moral concepts. Insensitivity to welfare in moral cognition is addressed in Section 4.

operate across domains, then they might have evolved functions that cut across domains. These overarching functions would be explicable not in terms of violation-specific adaptive problems (e.g., inbreeding avoidance), but in terms of adaptive problems shared in common across the variety of moral judgments.

In sum, the moral dimension is what remains constant across moral judgments about altruism, trade, fighting, sex, food taboos, drug use, black magic, etc. Across domains, moral cognition features a rich psychology surrounding moral judgment of oneself and others; condemnation, for instance, includes monitoring, gossip, judgment, punishment, and impartiality (see Sections 6–8). Focusing on the moral dimension itself, rather than the objects on which it operates, highlights evolutionary questions about the distinct and coherent features of moral cognition.

### 3. The problem of morality

A better understanding of the adaptive problems surrounding moral interactions can help guide research to uncover the cognitive mechanisms that evolved to handle moral situations. Moral interactions involve multiple individuals who each make decisions that affect their own and others' outcomes. This type of adaptive problem can be described in terms of *game theory*, the study of "games," or strategic environments in which individuals ("players") can affect each others' outcomes. At its core, game theoretic analysis involves taking the perspective of each player in an interaction and assessing how players can pursue their goals, given information about how the other players are likely to behave. Game theory has proved extremely useful for understanding the evolution of mechanisms for strategic behavior in humans and non-human species (Krebs & Davies, 1993; Maynard Smith, 1982). What strategic problems, or "games," shaped the cognitive systems that humans use to handle moral situations? Further, given the broad outline of moral interactions, what can be inferred about the structure and composition of moral cognition?

Consider this scenario: *A tall beauty is attacked by a hungry horde of grubby onlookers, and her screams attract a sky-soaring savior who lays waste to the villainous worms.* Although the scenario could be about human crime-fighting, it actually describes the strategic dynamic among maize, worm, and wasp. When attacked by moth larvae, maize emits a specialized chemical signal to alert parasitic wasps (Takabayashi & Dicke, 1996). Wasps attack the worms by injecting eggs that soon hatch into wasp larvae, which eat worms from the inside out.

This example illustrates that strategic games can shape adaptations (Maynard Smith, 1982). Specifically, the scenario involves a three-player game with asymmetric positions, each requiring different adaptations (see also Sinervo & Lively, 1996). Different strategic problems lead us to expect maize to have signaling mechanisms and wasps to have signal reception systems, rather than vice versa. Indeed, this is correct. Maize even emits different chemical compositions for different worm species, each signal drawing different specialist wasps (De Moraes, Lewis, Pare, Alborn, & Tumlinson, 1998).

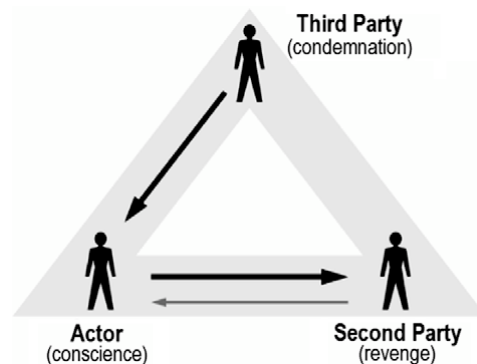
In line with adaptationism (Williams, 1966), the reliable link between strategic function and mechanism structure implies that information about adaptive problems can be used to make inferences about mechanism structure. Illustrating this reliable link, the three-player game above caused functionally equivalent devices to evolve (independently) in tobacco, cotton, strawberry, and cucumber, as well as in several species of wasps and mites (Takabayashi & Dicke, 1996). In this section, we draw on the strategic dynamic among perpetrator, victim, and third-party condemner to make inferences about the structure of moral cognition.

#### 3.1. Morality as (at least) a three-player game

Moral interactions frequently involve three roles or players: We will refer to these individuals as "actors," "second parties," and "third parties" (see Fig. 1). In this framework, actors affect second parties, and third parties judge actors and sometimes intervene. Third parties are recurrent elements of moral interactions. For instance, US crime statistics (1993–1999) show that in cases of assault – a common moral violation – third parties were usually present (72% of cases) and roughly half became involved in the situation (Planty, 2002).

The presence of third parties in moral interactions sets up a strategic game among actors, second parties, and third parties. Each role can be regarded as a different position in a moral game. These positions are defined by their respective problems, in the same way that tennis servers and receivers confront different problems (Walker & Wooders, 2001). Further, individuals might change positions (as in tennis), taking, in turn, the role of actor, second party, or third party.

We begin from the perspective of third parties, i.e., bystanders who observe and judge an actor whose behavior affects a second party. As previously mentioned, the broad adaptive functions served by third-party interven-



**Fig. 1.** Diagram of a strategic interaction among actor, second party, and third party. One important case is among perpetrator, victim, and condemner. Arrows depict the activity of one individual toward another as described in the parentheses. We focus on cognitive processes for managing the actor position (conscience) as distinct from mechanisms for managing the third party position (condemnation). For example, conscience might apply moral rules to questions such as "Should I take money from a victim?" or "Should I lie to a victim?" whereas condemnation would apply moral rules to ask "Should I condemn an actor who takes money from a victim?" or "Should I condemn an actor who lies to a victim?"

tion are unclear. Here we describe proximate problems that third-parties confront and resolve, leaving aside for the moment the ultimate goals that intervention supports.

Third parties condemn actors' moral wrongs and they also praise actors' moral virtues. Here we focus on third-party condemnation rather than praise.<sup>2</sup> We propose that one of the central moral games is the interaction among third-party condemner, perpetrator (actor), and victim (second party).

In this game, third-party condemners confront the problem of detecting, judging, and punishing perpetrators' violations, while recruiting others' support and avoiding counter-accusations. Condemners' judgments are primarily retrospective, occurring after the fact, and they usually involve uncertainty about actors' motives and behavior. Moreover, perpetrators typically respond to condemnation with denial, counter-accusations, and retaliation. In deciding whether to condemn, third parties must consider the probability of guilt, whether an accusation will be believed by others, and whether their own innocence is demonstrable (to defend against counter-accusations). If third parties choose to condemn a perpetrator, they must further decide the severity of the offense, how much punishment is deserved, and whether or not to perform punishment themselves. If condemners decide to punish, they face the problem of minimizing the costs associated with perpetrator retaliation (see Section 7). Third parties might reduce punishment costs by recruiting others' help or by delivering sanctions opportunistically, e.g., when the violator's kin are absent (see Knauft, 1987).

Actors, or potential perpetrators, face different problems. Actors must compute the costs and benefits of their potential actions and choose the actions that best achieve their goals. Some actions might involve moral violations against a second party, which could draw revenge from victims or sanctions from third parties. Whereas third parties seek to detect violations, perpetrators confront the opposite problem: avoiding detection by concealing violations. Whereas third parties make moral judgments retrospectively, actors make moral judgments prospectively, assessing the wrongness of a behavior before selecting a course of action.<sup>3</sup> Whereas third parties have incomplete

information about actors' intentions, perpetrators are certain of their own intentions (although they might try to hide intentions from condemners). Whereas third parties compute how much punishment is deserved, actors compute the punishments they expect to incur, and they attempt to avoid sanctions, sometimes by inhibiting wrongful behavior.

Finally, consider the perspective of the second party, i.e., a potential victim who is affected by the behavior of a perpetrator. Victims must avoid costs imposed by perpetrators, while recruiting aid from condemners. Victims can try to avoid being wronged by evading perpetrators, by mounting a defensive stand, or by engaging in revenge after the infraction to deter future violations. These tasks are not specific to moral violations (e.g., revenge can be used to deter behavior that is not morally wrong). However, in order to recruit help from third parties, victims are aided by moral judgments and accusations against their foes. Note that this suggests that second party moral judgment is derivative: Its force depends on the willingness of third parties to condemn moral violators.

The strategic dynamic among perpetrator, victim, and condemner might have shaped human adaptations. Recurrent adaptive problems can shape content-rich cognitive systems over evolutionary time (Tooby & Cosmides, 1992). The moral game might have favored cognitive systems that include representations of each position (perpetrator, victim, and condemner) as well as computations performed over these representations. Crucially, individuals in each position confront different problems, implying that they will require different computations to solve them.

Understanding moral cognition will be aided by a careful dissection of the relevant strategic problems. Broadly, perpetrators confront the problem of avoiding detection and punishment by condemners, and revenge by victims. Victims must avoid costs imposed by perpetrators, while recruiting aid from condemners. Condemners confront the problem of detecting and punishing perpetrators' violations, while recruiting others' support and avoiding counter-accusations. Because these very different problems require different computational procedures, they are likely handled by different specialized cognitive mechanisms (Marr, 1982; Sherry & Schacter, 1987; Tooby & Cosmides, 1992).

### 3.2. *Conscience and condemnation as different components of moral cognition*

Inferring structure from function, we suggest that the three-player moral game points to two components of moral cognition. The first is a system that uses moral concepts to guide one's actions; we will refer to this set of mechanisms as *conscience*. The second is a system that uses moral concepts to judge and punish a perpetrator; we will refer to this system as *condemnation*. While conscience operates in actors, condemnation operates in third-party condemners, as they evaluate actors.

Note that our usage of the terms "conscience" and "condemnation" stems from the three-player game among perpetrator, victim, and condemner. We classify as "conscience" the mechanisms in actors that select among

<sup>2</sup> Some broad stylized facts point to a special role for condemnation. Condemnation far exceeds praise in moral discourse among children (Ross & den Bak-Lammers, 1998) and adults (Wiessner, 2005). Mythic moral heroes, like superheroes, deliver punches to villains rather than rewards to the virtuous (DeScioli & Kurzban, 2008). Human legal systems sentence criminals to prison, rather than "sentencing" upstanding citizens to Disney World vacations. Looking at moral concepts, one common usage of "right" refers to "not condemnable" rather than "praiseworthy" – violent assault can be regarded as "right" if provoked, and abandoning a spouse can be viewed as "right" if precipitated by infidelity. However, the reverse is not true for "wrong" which means "condemnable" and not merely "not praiseworthy." Thus, moral concepts might especially hinge on condemnation. Finally, people praise altruism, but it is unclear how much praise is given to moral behavior, which is different from altruism (see Section 3).

<sup>3</sup> Third parties also make prospective moral judgments, e.g., to threaten potential violators or to negotiate candidate moral rules. However, these judgments are an *elaboration* on a basically retrospective design: prospective condemnation is meaningful only in terms of potential retrospective judgment and punishment after the violation occurs. Analogously, prospective conscience mechanisms are elaborated by retrospective guilt which is meaningful in connection with subsequent prospective judgments.

potential actions based on moral values computed for each action. We classify as “condemnation” the mechanisms in third parties that deploy accusations and punishment based on moral values computed for actors’ past actions. This distinction fits reasonably well with common usage of these terms.<sup>4</sup>

The inference that conscience and condemnation are distinct is supported by basic differences in the adaptive problems surrounding them. Actor conscience is primarily prospective, includes complete information about intentions, and competes with other motivations. Third-party condemnation is retrospective, informed by only incomplete information about intentions, and can be more detached from other motives (but see below about partiality, Section 8). Different functional demands can be used to infer different underlying cognitive mechanisms (Marr, 1982; Sherry & Schacter, 1987; Tooby & Cosmides, 1992). We can expect conscience and condemnation to be underlain by different specialized computational programs, just as we expect different mechanisms for worm signaling and wasp receiving.

The conscience/condemnation distinction is supported by work on moral emotions (Tangney, Stuewig, & Mashek, 2007). This literature distinguishes the *self-conscious* emotions of shame, embarrassment, and guilt (Tangney et al., 2007), from the *other-critical* emotions of contempt, anger, and disgust (Rozin, Lowery, Imada, & Haidt, 1999). These systems might differ in their targets (self vs. other) because they are designed for deployment in different strategic positions (actor vs. third-party).

Understanding conscience and condemnation requires a characterization of the functional relationship between them. Consider an analogy with human language. Language production and comprehension are served by functionally distinct systems, both neuroanatomically and at the level of gross morphology (ears as opposed to vocal cords and tongue, etc.). Further, language production and comprehension have a specific functional relationship. In general, theories of the evolution of communication regard reception/comprehension systems as driving the evolution of signaling/production systems (Dawkins & Krebs, 1978; Maynard Smith, & Harper, 2003). That is, the properties of signal reception systems cause the properties of signal production systems, rather than the opposite, because a signal cannot affect the receiver’s behavior unless the receiver has a system capable of interpreting the signal. Analogously, condemnation mechanisms might have caused the details of conscience, or vice versa.

The possibility that conscience and condemnation are different component mechanisms draws attention to questions about their functional relationship. The issue is as basic as whether the nose is designed to hold spectacles, or spectacles are designed to fit noses (see Voltaire’s *Candide*, 1759). To use an evolutionary example, warblers and cuckoos have matching colored eggs, leading to questions about

whether the color of cuckoo eggs is shaped by the color of warbler eggs or vice versa. Research favors a “warbler-centered” theory holding that warbler eggs are mimicked by cuckoo eggs, so that warblers will be tricked into caring for cuckoo offspring (Brooke & Davies, 1988). Below, we examine the functional relationship between actor conscience and third-party condemnation. Is conscience the core of morality, with condemnation emerging as a byproduct? Or, does condemnation form the moral core, with conscience evolving as a downstream consequence?

#### 4. Is morality conscience-centered?

Historically, evolutionary theorists have proposed “conscience-centered” explanations for moral cognition. That is, theorists have taken the problem of morality to be explaining the cognitive mechanisms that operate in actors to motivate their own behavior, answering questions such as, “why do people avoid incest?” Much less attention has been given to the mechanisms that cause third-party condemnation, answering questions such as, “why do people condemn others for incest?” Often, this focus is implicit: Theorists do not use the word “conscience,” but simply refer to mechanisms for self-regulation as “morality,” leaving condemnation out of view. There are several important exceptions that distinguish self-regulation from third-party condemnation. These accounts generally regard condemnation as a byproduct or spillover from conscience mechanisms or other psychological mechanisms, such as disgust (e.g., Fessler & Navarrete, 2004; Greene, 2008; Haidt & Joseph, 2004; Lieberman et al., 2003). In sum, theories of morality have generally been “conscience-centered” in that they either focus exclusively on systems for moral self-regulation, assuming that conscience can be explained independent of condemnation, or they explicitly argue that third-party condemnation is a downstream consequence of conscience mechanisms.

The historical focus on conscience mechanisms explains why theories of morality have tended to invoke altruism models. Altruism mechanisms motivate an actor’s behavior to benefit others. Darwin (1871) invoked a group selection model to explain morality, arguing that conscience is designed to benefit the group. Modern theories continue to regard morality as an altruism device, although the evolutionary processes thought to be responsible have expanded beyond group selection, and now include kin selection (e.g., Wright, 1994), reciprocity (e.g., Ridley, 1996), cultural group selection (e.g., Boyd & Richerson, 2005), and sexual selection (Miller, 2007). In *The Moral Animal*, Wright (1994) says of morality, “Darwin’s sometimes diffuse speculations about the ‘social instincts’ have given way to theories firmly grounded in logic and fact, the theories of reciprocal altruism and kin selection” (p. 328). In *The Origins of Virtue*, Ridley (1996) focuses on reciprocity, discussing the Prisoner’s Dilemma, public goods, and gains in trade, closing with the line: “We must encourage social and material exchange between equals for that is the raw material of trust, and trust is the foundation of virtue” (p. 265). In *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*, de Waal (1996) reviews primate

<sup>4</sup> In one example from fiction, Jiminy Cricket (in Disney’s *Pinocchio*) was assigned to be “Pinocchio’s conscience, Lord High Keeper of the knowledge of right and wrong... and guide along the straight and narrow path.” Acting as conscience, Jiminy guided Pinocchio’s own actions (versus urging him to condemn or accuse others for wrongdoing).

altruism, including chapters entitled “Sympathy,” “Quid pro Quo,” and “Getting Along.” De Waal opens the book: “In addition to being human, we pride ourselves on being humane. What a brilliant way of establishing morality as the hallmark of human nature – by adopting our species name for charitable tendencies!” (p. 1). Following Darwin, these accounts use theories of altruism to explain morality. More generally, they take conscience (mechanisms operating in actors) to be the key explanandum of morality.

This section presents an alternative view. We look at the structure of moral conscience to make inferences about its function. A central feature of moral cognition – pervasive nonconsequentialism – poses difficulties for altruism theories of conscience, and more generally, challenges the assumption that conscience can be explained independent of condemnation.

#### 4.1. Compliance with moral rules can be caused by non-moral systems

Conscience, on the present view, is a cognitive system that uses the moral concepts “right” and “wrong” to guide actor behavior. The effects of conscience are difficult to isolate because non-moral cognitive systems often cause morally compliant behavior (Kant, 1785/1993). For example, many birds are faithful mating partners but this behavior is (presumably) not caused by cognitive systems that use the moral concepts “right” and “wrong.” Mate fidelity in humans might also be caused, in part, by non-moral systems, in addition to moral motivations. The challenge is to pick out conscience against a background of morally compliant behavior caused by other systems.

Altruism is an important example of how morally compliant behavior is often caused by non-moral systems. Altruism can evolve by multiple pathways, including kin selection, reciprocity, and mutualism. Helping is widespread in organisms from bacteria (Griffin, West, & Buckling, 2004) to insects (Holldobler & Wilson, 1990) to non-human primates (de Waal, 1996; Schino, 2007). Likewise, many human altruism devices, such as parental care mechanisms, evolved prior to moral cognition and continue to operate independent of moral motives (e.g., umbilical cords, mammary glands). The development of altruism is revealing. By age one, children comfort distressed individuals with hugs and kisses, and by age two, many prosocial behaviors emerge (Eisenberg & Fabes, 1998; Zahn-Waxler, Radke-Yarrow, & Wagner, 1992). For instance, Warneken and Tomasello (2006) found that 18-month-olds helped experimenters reach an object dropped accidentally (but not intentionally). However, moral cognition seems to begin to develop around age 3 and continues to change throughout childhood (Darley & Shultz, 1990). During this time, children increasingly show patterns typical of adult moral judgment, such as distinguishing intentional and accidental violations or foreseeable and unforeseeable violations (Darley & Shultz, 1990). Importantly, these changes occur after a large repertoire of prosocial behavior is already in place. Taken together, phylogenetic and ontogenetic evidence indicates that altruistic behavior can occur independent of moral conscience (see also Batson, Klein, Highberger, & Shaw, 1995; Kant, 1785/1993).

The issue is more general than altruism. For example, while Freud (1918) attributed incest avoidance to fear of moralistic punishment, recent evidence indicates that childhood co-residence causes sexual aversion (Fessler & Navarrete, 2004; Lieberman et al., 2003, 2007). Compliance with incest prohibitions might be primarily caused by specific incest aversion mechanisms rather than conscience. Similarly, Kant argued that suicide was morally wrong but most people “preserve their lives, to be sure, in accordance with duty, but not from duty” (Kant, 1785/1993, p. 10).

Isolating the specific effects of conscience is important for understanding its function. Evolutionary theories of moral conscience generally take for granted that compliance with moral rules can be attributed to the operation of conscience. However, it is unclear to what extent conscience causes moral compliance over and above compliance produced by non-moral systems designed for altruism, fidelity, honesty, disease-avoidance, etc. Indeed, experiments that control for alternative motives (e.g., reputation) show surprisingly little evidence that moral judgment motivates moral behavior. Instead, moral judgment often motivates efforts to *appear* moral (Batson, 2008).

#### 4.2. Nonconsequentialism in moral cognition

Moral *consequentialism* is the philosophical view that “whether an act is morally right depends only on consequences” (Sinnott-Armstrong, 2006). Usually, consequentialist theories focus on how actions directly affect *welfare* outcomes (e.g., utilitarianism), and we will use this sense of the term. That is, consequentialists judge an action solely based on the welfare consequences that the action is expected to produce.<sup>5</sup> In contrast, moral *nonconsequentialism* is the view that moral rightness does not depend only on consequences. Commonly, theories in this genre focus on absolute rules of behavior, such as Kant’s “categorical imperatives”. Consequentialism regards actions as completely instrumental, to be judged solely according to their expected effects. In contrast, nonconsequentialism allows judgment on the basis of the properties of an action (e.g., whether it involves lying) rather than only on the basis of its effects.

The debate between these views extends beyond philosophy. Legal professionals debate whether the “rule of reason” (consequentialism) or “per se rules” (nonconsequentialism) should be used to decide cases (Arthur, 2000). In the policy arena, these approaches often compete

<sup>5</sup> Philosophers have proposed many forms of consequentialism (Sinnott-Armstrong, 2006). To avoid any potential confusion, we will specify further the type of consequentialism we have in mind: The relevant consequences are *welfare* consequences that are expected (versus actual) and direct (versus indirect). By “welfare,” we remain general by including all possible weightings among self and others’ welfare as potential grounds for consequentialist judgment. The focus on “expectation” further implies that greater weight is placed on immediate proximate effects, because more distant effects should be discounted in proportion to uncertainty. The focus on “direct” effects means that the relevant consequences are the expected welfare consequences attached to the action under evaluation; the relevant effects *do not* include consequences resulting from the moral judgment itself, such as reputation effects for the judge.

to shape decision-making (Baron, 1998), as signaled by the title of an important book on the subject, *Fairness Versus Welfare* (Kaplow & Shavell, 2002). For example, a debate about HIV policy concerns whether programs should focus on behavior per se (abstinence) by denying services to active sex workers and drug users, or should serve these populations to reduce HIV transmission and the resulting harm (Marlatt, 1996; Rekart, 2005).

Experiments have found that people's moral judgments are sometimes consequentialist and sometimes nonconsequentialist (reviewed in Hauser, 2006). For example, in one version of the trolley problem (the switch dilemma), people generally made consequentialist decisions: most participants (90%) thought it was permissible to flip a switch to redirect a trolley, thereby killing one person to save five. However, in the footbridge dilemma, people generally made nonconsequentialist decisions: most people (90%) said it was impermissible to save five people by pushing one person off of a footbridge in front of the trolley. These results show that most people are nonconsequentialists (but not strict deontologists).

Nonconsequentialism is widespread and complex. One line of research found that people maintained that certain behaviors were wrong and should be punished even when no harmful consequences would result (e.g., non-reproductive consensual incest, Haidt, 2001; Haidt & Hersh, 2001; Haidt et al., 1993; Tetlock, 2000). Others have shown that the intentional and causal structure of behavior influences nonconsequentialist decisions in complex ways – sufficiently complex that some researchers regard moral judgments as comparable in intricacy to the universal grammar of natural language (for ideas about a universal moral grammar, see Hauser, 2006; Mikhail, 2007).

A narrow focus on behavior per se, rather than consequences, is a signature feature of moral cognition. Reasoning surrounding the concepts “right” and “wrong” often involves judgments based on behavior per se rather than the instrumental effects of behavior. No other decision-making domain seems to share this peculiarity. For example, people often use rules in precautionary reasoning, such as “Do not drive through stop signs,” but they hardly maintain such rules even when this will lead to much greater harm (e.g., a truck skidding out of control from behind). Outside of morality, a narrow concentration on behavior per se, rather than consequences, is regarded as pathological (obsessive-compulsive disorder, see Boyer & Lienard, 2006).

#### 4.3. Nonconsequentialism undermines consequentialist explanations

Evolved mechanisms are organized to bring about particular consequences, thereby allowing organisms to protect, repair, and replicate themselves. Behavioral adaptations work by evaluating the environment and reconfiguring the organism in such a way that certain consequences (e.g., predatory attack) are more or less likely. These mechanisms tailor behavior to fit the changing world. Thermoregulation devices, for instance, track temperature changes and activate behaviors (e.g., shivering, sweating) that bring about specific consequences (target

body temperature). Mate choice mechanisms track information about others' sex/age/health and activate behaviors (e.g., courtship, copulation) that bring about specific consequences (fertilization). In short, evolution generally produces consequentialist mechanisms. This does not, of course, imply conscious calculation of consequences (e.g., sweating), but only designs that are well organized to bring about adaptive outcomes (Dawkins, 1976).

Moral nonconsequentialism in humans is puzzling because the underlying mechanisms focus on behavior per se rather than on consequences. This decision procedure seems to undermine the very *raison d'être* of behavior: to dynamically respond to different problems. Rules of behavior that are insensitive to welfare consequences such as “never kill” or “never exchange sex for money” (compare to “never shiver”) are, by definition, insensitive to context, which can lead to damaging – sometimes disastrous – consequences. What explains why moral judgment rejects certain behaviors, even when the costs are severe?

Nonconsequentialism raises doubts about altruism theories of moral conscience. The a priori prediction of altruism theories seems clear: Moral conscience should be consequentialist. In non-human animals, altruism (e.g., parental care) is regulated by outcome-driven processes in which the performance target is increasing others' welfare (e.g., offspring; Krebs & Davies, 1993). Kin selection favors mechanisms that increase inclusive fitness by helping relatives. When “dilemmas” arise, such as food shortages, many animals show their consequentialism by neglecting or even killing their offspring or siblings in order to allocate resources more efficiently (Mock, 2004). Similarly, other pathways to altruism such as reciprocity or group selection should yield consequentialist mechanisms, not inflexible rules. Group selection, for instance, should favor a design that sacrifices one group member to save five.

More broadly, the candidate adaptive problems that surround theorizing in the morality literature – altruism, disease avoidance, infidelity, etc. – ought, barring exceptional circumstances, to lead to adaptations designed to bring about a solution to the adaptive problem (delivering benefits, avoiding pathogens, etc.). The striking and persistent deviations from consequentialism in moral conscience undermine theories that posit direct consequentialist functions.

#### 4.4. Nonconsequentialism: error or evidence?

A riposte to the above argument is that what appears to be nonconsequentialism is actually error, arising from a system implementing *rule consequentialism*. That is, simple heuristics such as “never kill” increase welfare on balance, even though they produce some errors. These heuristic rules might be implemented by conscious reasoning (Baron, 1994; Sunstein, 2005) or by unconscious emotions (Gigerenzer, 2007; Greene, 2008, p. 60). Further, the artificial dilemmas used in experiments might exaggerate errors, giving a misleading impression of their magnitude. Natural selection can favor fast and frugal heuristics when time, knowledge, or computational abilities are limited (Gigerenzer & Goldstein, 1996). Do moral rules function as heuristics for welfare?



We think this error explanation is unlikely. What special limitations on time, knowledge, or computation explain why moral problems require extremely simple rules (more than other decision domains)? All real world problems are complex, but computational systems usually manage complexity by evaluating a few key dimensions (rather than none). A persuasive argument that moral rules are heuristics will have to specify the properties of moral problems that explain why extremely simple rules are required to solve them. The current arguments are cast so broadly that if they were correct, then all cognitive functions would be performed with commandment-like rules. Given that this is not the case, these arguments need to be reconsidered.

Consider, for example, a design that focuses on several computations. Conscience mechanisms could compute for nearby individuals the physical harm likely to occur in the immediate future – this system would favor killing one person to save five. These computation requirements do not seem exceptionally burdensome, involving calculations that people appear to already perform (in precautionary reasoning) and that seem readily available in individuals' welfare judgments. Indeed, in most of the relevant experiments, welfare costs were high and easy to compute. Often, participants were asked to assess both welfare and moral wrongness (Haidt, 2001; Haidt & Hersh, 2001; Haidt et al., 1993; Tetlock, 2000), and, unlike moral judgments, welfare judgments accurately tracked welfare outcomes. Given that people are able to compute welfare in these situations, simple moral rules offer no computational advantage.

Further, moral conscience is not, in fact, cognitively simple. On the contrary, experimental findings show that people are sensitive to many dimensions of actor behavior, even when consequences are ignored (Hauser, 2006; Robinson & Darley, 1995; also see Section 6). As mentioned above, the complexity of moral judgment has led to theories surrounding a universal moral grammar (Hauser, 2006; Mikhail, 2007). Conscience is no cognitive miser, so why are so little computational resources devoted to evaluating welfare outcomes when making moral judgments?

Evaluated for flight, penguins' wings appear flawed and "suboptimal." At some point, it makes sense to consider wing features not as defects but rather as evidence of an alternative function, like swimming. If moral nonconsequentialism points to an alternative function for conscience, then attempts to explain conscience independent of condemnation are like trying to explain a penguin wing out of water. Because the evolutionary function of morality is unknown, patterns in moral cognition cannot be safely dismissed as error. Instead, consistent features are best viewed as potential clues to the adaptive problems that shaped the system.

#### 4.5. Summary

Current theories of moral conscience are problematic. First, non-moral systems can cause moral compliance, making it difficult to know the extent to which conscience is responsible for altruism, fidelity, incest avoidance, etc.

(Kant, 1785/1993). In fact, research has turned up surprisingly little evidence that moral judgment motivates morally compliant behavior; instead, it often motivates efforts to *appear* moral (Batson, 2008). Second, conscience is often nonconsequentialist, focusing narrowly on behavior per se rather than consequences. Because this is peculiar to conscience, it seems safe to associate nonconsequentialism with moral concepts, whereas consequentialist judgments are more difficult to trace to moral as opposed to non-moral mechanisms. Last, a decision procedure that leads one to absolutely avoid certain behaviors does not appear well designed to accomplish any direct actor goal, undermining theories that posit direct consequentialist functions.

We think that the key to unlocking the problem of non-consequentialism is the strategic dynamic among perpetrator, victim, and condemner. Decision procedures that are never advisable in individual decision-making are sometimes advantageous in multi-player strategic interactions because *they influence others' behavior* (Schelling, 1960). For example, commitment mechanisms (e.g., environmentalists chaining themselves to threatened trees) reduce agents' own freedom of action, but this can influence other agents' behaviors in desirable ways (e.g., Frank, 1988; Hirshleifer, 1987). Perhaps the structure of moral conscience can be illuminated by considering how actors interact with third-party condemners.

### 5. Condemnation-centered morality

We now turn to an exploration of the potential value of examining moral systems starting with condemnation rather than conscience.

#### 5.1. Conscience does not explain condemnation but condemnation explains conscience

Theories proposed to explain actor conscience do not, without additional assumptions, concurrently explain third-party condemnation. For example, it has been proposed that incest aversion functions to guide the organism to avoid inbreeding costs (e.g., Lieberman et al., 2003). This does not, by itself, explain why humans want *others* punished for incest.<sup>6</sup> In fact, individuals with a disposition to *praise* others' fitness-reducing moral violations, such as incest or suicide, could in theory gain an advantage by manipulating competitors to engage in self-destructive behavior.

There is at present no well-developed model that provides a direct pathway from conscience to condemnation. That is, there is no reason to expect that moral systems designed to avoid certain acts (conscience) will cause representations that others should avoid those acts and be punished if they do them (condemnation). One possible conscience-to-condemnation relationship is that once conscience mechanisms evolved, this created selection favoring individuals with condemnation systems. However, it

<sup>6</sup> For instance, in the United States legal system, "all fifty states and the District of Columbia have some variation of a prohibition of incest on the books" (Inbred Obscurity, 2006, p. 2469). In Massachusetts, consensual incest is punishable by up to 20 years in prison.

is difficult to see how the presence of conscience mechanisms might favor mutations associated with condemnation mechanisms. Another possible approach would be to assume that conscience mechanisms are designed to regulate one's own behavior but they sometimes overgeneralize to regulate others' behavior. However, such extreme errors – failure to distinguish one's own behavior from others' behavior – imply an exceptionally poor design (Cosmides & Tooby, 2000) and have been observed only in cases of severe pathology (e.g., self-monitoring deficits in schizophrenia; Frith, 1996). Moreover, whereas overgeneralization could cause individuals to experience vicarious aversion to others' violations, it seems insufficient to explain why people want violators to be punished. Moralistic punishment is a new feature not present in conscience.

Whereas conscience does not explain third-party condemnation, *given third-party condemnation, an explanation for conscience straightforwardly follows*. In a population of people who condemn others for certain behaviors, selection will favor defense systems that guide individuals to avoid those behaviors (unless the benefits outweigh the costs of being punished). Indeed, models show that, given certain assumptions, punishment can favor the evolution of *any* behavioral system (Boyd & Richerson, 1992, 2005). Actors can estimate condemnation costs by representing the computations of third-party condemners, *whatever those computations might be*. In this scenario, conscience is designed to apply the concepts of right and wrong to regulate one's own behavior to defend against third-party condemners.

Mencken (1949) wrote, "Conscience is the inner voice that warns us somebody may be looking" (p. 617). If conscience functions, in part, as a defense system, then this might help explain nonconsequentialist decisions. Conscience might focus on behavior per se rather than consequences because third-party condemners focus on behavior. Of course, this does not solve the problem of nonconsequentialism, but it potentially consolidates two puzzles – nonconsequentialism in both actors and condemners – into one puzzle – nonconsequentialism in condemners.

### 5.2. Conscience as a defense system: evidence from moral hypocrisy

The idea that conscience functions as a defense system might explain a ubiquitous feature of human life: moral hypocrisy, the gap between individuals' moral standards and their behavior (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997). People often engage in behavior that they believe is morally wrong. What explains this? One possibility is that immoral behavior results from a failure of self-regulation mechanisms, perhaps due to competing motives. This account posits errors in system integration between conscience and other cognitive mechanisms.

Another possibility is that conscience is a defense system that allows immoral behavior when condemnation is unlikely. For example, when persuasive justifications have been identified, conscience might facilitate immoral behavior; this could explain why justifications can cause

"moral disengagement" (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). After behaving immorally, maintaining a representation of the behavior as morally wrong can facilitate concealment or, in the event of detection, deployment of the necessary justifications or reparations. Moral hypocrisy, therefore, might simply reflect the functioning of conscience as a system for defense rather than direct behavior regulation.

Moral hypocrisy has been demonstrated in the laboratory (Batson & Thompson, 2001; Batson, Thompson, Seufferling, Whitney, & Strongman, 1999; Batson et al., 1997). In a typical experiment, participants assigned themselves and another individual to different tasks, one more desirable than the other. Participants could flip a coin to choose randomly or they could simply choose. Retrospectively, nearly all participants judged flipping the coin as morally right. Roughly half of participants flipped the coin, but whether they used the coin or not, 90% assigned themselves to the desirable task, showing that participants typically ignored the coin flip. Batson and Thompson (2001) concluded that "at least some individuals want to appear moral while, if possible, avoiding the cost of actually being moral" (p. 54).

Similarly, studies of children's tattling show they eagerly report others' wrongs but ignore their own transgressions (den Bak & Ross, 1996; Dunn & Munn, 1985). Talwar, Lee, Bala, and Lindsay (2002) found no relationship between children's understanding of lying and truth-telling and their own lying behavior. Most children correctly distinguished lies from truth, judged a fictional liar as wrong, recommended that other children should tell the truth, and claimed that they would tell the truth. However, most of these same children lied to conceal their own cheating. In a similar study (Talwar, Lee, Bala, & Lindsay, 2004), children lied to hide their parent's transgression – unless they themselves might be implicated in the infraction.

### 5.3. Lady Justice

In contrast to the historical focus on actor conscience in theories of morality, we propose that third-party condemnation is a key – perhaps *the* key – explanandum of morality. If conscience functions, in part, as a defense system, then its structure cannot be understood independent of the structure of condemnation, which conscience should closely parallel.

The potential centrality of condemnation has important implications for empirical research. For example, consider the action-omission effect, which leads people to view killing someone as more wrong than letting someone die. This phenomenon can be approached from two perspectives: (1) How does the action-omission effect benefit actors? or, (2) How does the action-omission effect benefit third-party condemners? The condemnation-centered view foregrounds the latter question.

This returns us to the threefold mystery of morality symbolized by Lady Justice's scale, sword, and blindfold (see Fig. 2). First, why do people care, at all, about moral violations occurring among unrelated others? Second, why do people punish moral violators, even when punishing is costly? Third, why do people try to appear impartial,

## THREE MYSTERIES OF MORALITY




<b>Moral Judgment</b>	Humans monitor and evaluate interactions among unrelated others according to moral rules of behavior.	
<b>Moral Punishment</b>	Humans impose costs on individuals who violate moral rules, even if this might draw retaliation.	
<b>Moral Impartiality</b>	Humans damage valuable social relationships in order to enforce moral rules impartially.	

Fig. 2. Three mysteries of morality.

claiming to neglect relationships with kin, friends, and allies in their moral decisions and actions? Just as sexual reproduction initially appeared biologically bizarre with seemingly insuperable costs (Hamilton, 2001; Williams, 1975), third-party judgment, punishment, and impartiality represent biological mysteries of the first order.

### 6. Moral judgment

In the biological world, organisms monitor others' interactions only when the benefits exceed the costs. Third-party monitoring can sometimes help organisms find food (e.g., wasps, see above), quality mates (mate copying, White & Galef, 1999), or avoid costly fights (using transitive inference, Doutrelant, McGregor, & Oliveira, 2001). A condemnation-centered view of morality highlights the question: What benefits explain why people monitor others' moral behavior, even when they themselves are unaffected?

Interest in moral infractions is enormous. The press draws audiences with reports of celebrity debauchery, political corruption, and corporate fraud. Medieval public executions and modern superhero movies (DeScioli & Kurzban, 2008) attest to widespread desire to see transgressors punished, whether in fact or fiction.

Research on children's tattling shows that moral interest begins early in development (den Bak & Ross, 1996; Dunn & Munn, 1985; Ross & Den Bak-Lammers, 1998). Even before age two, children call attention to others' violations (while ignoring their own). Roughly half of the time, the tattler is unaffected by the violation. Further, tattling persists even when parents react negatively to snitching. Finally, tattling precedes, and is uncorrelated with, positive talk, leading Ross and den Bak-Lammers (1998) to conclude that "tattling does not emerge from the more general experience of sharing information about the sibling with the parent, but is an earlier, largely unrelated form of parent-child discourse" (p. 294).

Continuing into adulthood, moral judgment is a staple of everyday conversation. Gossip, often about wrongdoing, is cross-culturally ubiquitous (Barkow, 1992; De Backer, 2005; Dunbar, 1996; Gluckman, 1963; Haviland, 1977). Roughly 65% of conversation is about social topics (Dunbar, 2004). As in children's tattling, moral discourse is disproportionately negative. This pattern has been observed cross-culturally. For instance, among the Ju/'hoansi Bush-

men, 56% of conversation contained moralistic criticism, whereas 7% contained praise (Wiessner, 2005).

Moral surveillance is an active process. People seek evidence, interrogate claimants, and interview eyewitnesses, whose testimony is given more credence than secondary sources (Wilson, Wilczynski, Wells, & Weiser, 2000). People are sophisticated judges, comparing facts from multiple sources, attributing greater credibility to facts repeated by multiple independent sources, and discounting information from sources with vested interests (Hess & Hagen, 2002).

Humans seek, gather, evaluate, and communicate information about wrongdoing among others. What benefits offset the costs of these time-consuming behaviors?

#### 6.1. What is moral judgment good for?

One possibility is that moral judgment functions to evaluate others' altruism/selfishness. Broadly, this predicts that moral judgment should be a function of intended benefit or harm. But this seems inconsistent with observed insensitivities to welfare outcomes (e.g., the footbridge dilemma, Section 4). If moral judgment measures altruism, then why don't welfare outcomes dominate judgment? Similarly confusing are findings showing that the severity of unintended consequences can influence moral judgment (Rucker, Polifroni, Tetlock, & Scott, 2004). Participants judged a hypothetical carjacker who harmlessly ejected a passenger as less wrong than a carjacker who inadvertently caused a heart attack. As Rucker et al. noted, "Surely, we learn nothing new about the moral character of the carjacker in the high-severity versus the low-severity condition" (p. 673). These observations suggest that moral judgment performs some alternative or additional functions other than assessing others' altruism.

Rather than consequences, moral judgment seems to be particularly attuned to properties of perpetrator behavior, including whether actors: (1) act or fail to act, (2) act intentionally or accidentally, (3) act with contact or at a distance, and (4) violate directly or as a byproduct (see Hauser, 2006). First, acts are judged more harshly than inaction, even when the latter causes more harm (Baron & Ritov, 1994, 2004; Cushman, Young, & Hauser, 2006; Haidt & Baron, 1996; Ritov & Baron, 1990, 1992, 1995, 1999; Spranca, Minsk, & Baron, 1991). For instance, recommending a salad dressing with cayenne pepper to an allergic tennis competitor was seen as more wrong than

remaining silent when the competitor chose himself (Spranca et al., 1991). The effect has also been found in children (8 years: Baron, Granato, & Spranca, 1993; 11–17 years: Keltikangas-Järvinen & Lindeman, 1997).

Second, intended violations are perceived as more wrong than accidental ones (Robinson & Darley, 1995), holding goals constant. Pizarro, Uhlmann, and Bloom (2003) looked at “causal deviance,” in which an agent intends to cause an outcome but causes the outcome in a manner other than the intended one. Participants evaluated a protagonist who planned to stab his enemy and either did so, or, was bumped and stabbed him accidentally. The accidental killing was viewed as less blameworthy. The intention/accident distinction has been found as early as age three (Ferguson & Gail, 1988; Grueneich, 1982; Leon, 1984; Nuñez & Harris, 1998; Sedlak, 1979).

Third, transgressions involving physical contact are seen as more immoral. Recall that when an individual facing the Trolley Problem can flip a switch to divert the trolley (rather than push the man), killing one to save five, judgments change dramatically: Only 10% viewed this as wrong (vs. 90% for pushing the man; Hauser, 2006).

Fourth, violations occurring as a foreseen byproduct of another act are seen as less wrong than violations used as a means to an end. In another variation of the Trolley Problem, there is a sidetrack that loops back to the original track. Diverting the trolley to the sidetrack can save five people, but only if there is a heavy object on the sidetrack. When the object was a heavy man, 50% of participants judged diverting the trolley as impermissible. When the object was a weight – but a man standing in front of the weight will be killed – only 25% judged diverting the trolley as impermissible (Hauser, 2006; see also Royzman & Baron, 2002; Waldmann & Dieterich, 2007).

In sum, moral judgment is especially attuned to properties of actor behavior. Moreover, people are often unaware that these factors play a role in their decisions (Cushman, Young, & Hauser, 2006). What function might explain these cognitive processes? It might help to think about the problems confronted by third-party condemners. One problem, for instance, is recruiting others' support, which often requires substantiating evidence. Perhaps, for example, the action-omission distinction tracks the likelihood of a persuasive moral accusation. Importantly, unpersuasive accusations can draw dangerous retaliation from the accused and their allies. If providing evidence for a wrongful act is easier than demonstrating a violating omission, then third-party condemners might benefit by feeling less moral outrage for omissions relative to actions (DeScioli, 2008). This is one way that the action-omission distinction might contribute to a condemnation function.

## 6.2. Scales of justice

Other features of moral judgment require explanation. First, the ontology of legitimate perpetrators or victims (animals, inanimate objects, etc.) shows striking variation (Kadri, 2005; Singer, 1981). Next, people are motivated to remain consistent in their moral judgments, leading to strong order effects in experiments (Hauser, 2006). Also, people dislike moral diversity, preferring group members

to agree with them (Haidt, Rosenberg, & Hom, 2003). Further, individuals readily compare violations across domains, and severity rankings sometimes show remarkable consistency (Robinson & Kurzban, 2007).

Moral judgment is complex. While moral judgments are not uniquely determined by welfare, authority, custom, precedent, etc., each of these factors can influence judgment. Emotions such as disgust and anger can exert strong influences (Haidt & Hersh, 2001; Haidt et al., 1993; Rozin et al., 1999). Also intriguing but infrequently discussed is the historically widespread practice of trial by ordeal or combat (Kadri, 2005), in which surviving fire, boiling oil, or dueling determines moral rightness. Why can moral judgment be influenced by arbitrary contests?

These complexities suggest that the psychological scales of moral judgment are tipped by many factors. Moral dilemmas might occur when welfare is pitted against authority, custom contradicts emotion, and so on. Cultural variation is expected because it is unlikely that precisely the same balance of forces will be struck for all moral issues across different societies (Shweder et al., 1997). Variation should taper when factors reach extremes or converge (welfare and disgust both oppose boiling babies for fun). The reason that moral judgment possesses these sensitivities should be clarified by asking the question: What adaptive functions are served by monitoring others' moral violations?

## 7. Moralistic punishment

Focusing on third-party condemnation highlights the distinction between moralistic punishment and second-party revenge (see also Nozick, 1981). Morality and vengeance intersect in moral rules that place limits on retaliation (e.g., *lex talionis*, e.g., “eye for an eye”). However, moral rules about revenge do not imply an equivalence between these forms of punishment. An evolutionary perspective underscores important differences between revenge and moralistic punishment.

Second-party punishment is widespread in nature (Clutton-Brock & Parker, 1995). Design for damaging others, even when costly, can yield an evolutionary advantage by altering others' behavior. Punishment occurs in a variety of contexts, including dominance relationships, competition for resources, mating interactions, parent-offspring conflict, and cooperative breeding (Clutton-Brock & Parker, 1995). Egret chicks punish their smaller siblings for competing to get food brought by the parent (Mock, 2004). Mother elephant bite unrelated pups who try to drink their milk (Reiter, Stinson, & Le Boeuf, 1978). In several primate species, males punish females who refuse their attempts to mate (Smuts & Smuts, 1993). These examples illustrate how animals use punishment in a variety of contexts to deter or coerce others.

In contrast, third-party punishment is rare among non-humans, with several possible exceptions. Social insects kill workers for laying eggs (Foster & Ratnieks, 2001; Gobin, Billen, & Peeters, 1999), hyenas intervene in others' fights (Engh, Siebert, Greenberg, & Holekamp, 2005), and chimpanzees intervene on behalf of unrelated allies (de Waal,

1996). Among pigtail macaques, Flack, de Waal, and Krakauer (2005) found that third parties intervened in 72% of conflicts; 13% of interventions drew subsequent retaliation.

Explanations for second-party punishment are relatively straightforward (e.g., relative advantage, deterrence, extortion), but third-party punishment is generally more difficult to explain (kin selection for social insects; perhaps alliance-building for some mammalian cases). The problem is especially difficult because third-party condemners suffer costly retaliation from perpetrators. What benefits offset the costs of third-party punishment?

### 7.1. Field and laboratory observations

Moralistic punishment varies in intensity and costs.<sup>7</sup> Dissemination of information about others' infractions damages violators' reputations (Hess & Hagen, 2002). Even this form of punishment can be costly. Among the Ashanti of Ghana, "tale-bearing concerning the private affairs of a Chief was often punished by cutting off the lips" (Ratray, 1929, p. 327). Publicly ridiculing, shaming, or complaining about transgressors risks undermining valuable relationships (Kaplan & Hill, 1985). Miller (2003) investigated third-party sanctions of people who illegally parked in handicapped spaces. Punishment ranged from cold stares to tire slashing and was sometimes met with retaliation; one condemner was physically assaulted by three men. Among the Gebusi of New Guinea, Knauff (1987) documented 82 executions of alleged disease-causing sorcerers; 5% of the executioners were themselves killed in retaliation by the sorcerer's kin. Among the Ju/'hoansi Bushmen of Botswana, Wiessner (2005) found that punishment increased in intensity as violations were repeated, beginning with mockery and insults, followed by harsh criticism and finally violence. Most punishments were administered by coalitions of three or more individuals, and 8% of these group punishments yielded transparent costs to punishers, usually by interrupting reciprocal sharing; 3% ended in vicious fights.

Research in social psychology on bystander intervention (Latané & Nida, 1981) shows that people are often willing to incur costs to stop others' violations. Studies have found high rates of third-party intervention for wrongdoing, including assault (65%, Shotland & Straw, 1976; 44%, Fischer, Greitemeyer, Pollozek, & Frey, 2006), rape (65%, Harari, Harari, & White, 1985), theft (57%, Howard & Crano, 1974; 28%, Gelfand, Hartmann, Walder, & Page, 1973), and graffiti and littering (49% and 63%, respectively, Chekroun & Brauer, 2002).

More recently, researchers have used economic games to examine third-party punishment. Turillo, Folger, Lavelle, Umphress, and Gee (2002) found that third parties paid to punish, 15% of the time, individuals who unfairly divided money with another individual. Fehr and Fischbacher (2004) found that 60% of third parties were willing to pay at least a small amount to punish "unfair" individuals. Also,

third-party parties spent 8% of their endowment to punish people who defected when their partner cooperated in a Prisoner's Dilemma game. Carpenter and Matthews (2005) examined punishment in a Public Goods Game (Ledyard, 1995). They found that 10% of participants showed third-party punishment of non-contributors, spending an average of only \$0.10 total on third-party sanctions. Henrich et al. (2006) looked at third-party punishment for unfair divisions in 15 cultures around the world. Overall, 66% of third parties were willing to pay to punish the most unfair division (100% allocated to oneself), with this figure ranging from 28% to over 90% across cultures.

While the previous studies examined the prevalence of third-party punishment, economic games can also be used to look at the influences of key variables. Kurzban, DeScioli, and O'Brien (2007) examined the influence of anonymity. Third-party participants could pay to punish individuals who violated reciprocity in a sequential prisoner's dilemma. When participants knew they would announce their decisions to other third parties, they spent roughly three times more on punishment (average \$3.17) than (different) participants in an anonymous condition (average \$1.06). Interestingly, in post-experiment surveys, not a single subject in the public condition mentioned the audience as a factor in their decision. Audience effects might be specific to third-party punishment, as second-party revenge has been observed to be insensitive to anonymity (Bolton & Zwick, 1995).

### 7.2. Accounting for cost

The costs of moralistic punishment imply an important distinction between: (1) wanting wrongdoers to be punished, and (2) wanting to punish wrongdoers. This distinction is often neglected (e.g., Singer et al., 2006). Because the underlying systems should be sensitive to cost, the feeling that a wrongdoer *deserves* punishment is distinct from motivation to *perform* punishment oneself (Robinson, Kurzban, & Jones, 2007). Historically, public executions have drawn large audiences, reflecting a strong desire to see wrongdoers punished; this does not show observers' willingness to perform costly punishment themselves.

In conclusion, while second-party punishment is biologically common, third-party punishment is rare in non-humans and presents special theoretical challenges. Precisely what benefits explain why people have cognitive systems that perform costly third-party punishment?

## 8. Moral impartiality

Lady Justice's blindfold symbolizes that condemnation should not depend on who is helped or harmed. Moral impartiality requires condemners to ignore kinship, friendship, and group loyalty. Sometimes people show moral impartiality, like when David Kaczynski turned in his brother Theodore, the "Unabomber," for killing three people. Sometimes people are partial, like when politician William Bulger refused to help authorities find his brother, wanted for 19 murders. Here we consider the theoretical challenges posed by moral impartiality. We do not, of

<sup>7</sup> Elsewhere we have investigated whether third-party punishment is driven by reputation benefits, arguing that third-party punishment is rare in anonymous laboratory environments. Here we focus, more generally, on the theoretical challenges posed by third-party punishment, independent of whether it occurs in public or private.

course, assume that people are always impartial, or even that they are usually impartial. Instead, we ask why people are *ever* impartial to any degree at all, why they often claim to be impartial, why they praise impartiality, and why humans possess the concept of moral impartiality. The mystery of moral impartiality surrounds its inevitable conflicts with loyalty.

Focusing on third-party condemnation shifts how impartiality itself is construed. From the perspective of conscience, impartiality is the opposite of *selfishness*, i.e., actors' own preferences should not influence moral judgments. The view from third-party condemnation is markedly different. For third parties, with little or no direct interest in the interaction, moral impartiality's opposite is *favoritism*. From the third-party perspective, "partiality" is not selfishness, but loyalty to kin, friends, and groups.

Third-party impartiality is a frictionless machine of biology – seemingly impossible. Theories of altruism require *partiality* (toward kin, reciprocating partners, or groups), which is made possible by specialized discrimination mechanisms (Axelrod & Hamilton, 1981; Cosmides & Tooby, 1992; Trivers, 1971). That is, partiality is a specific strategy underpinned by specialized cognitive machinery. *Impartiality undermines this machinery*, threatening valuable relationships. Even the pretense of impartiality could weaken trust and damage relationships.

Third-party punishment among non-human animals tends to be partial. Among hyenas, when third parties joined fights among unrelated others, they sided with the higher status individual 94% of the time (Engh et al., 2005). Chimpanzees possess a "system of revenge" in which individuals side against those who side against them (de Waal & Luttrell, 1988). There are, however, some exceptions: Some high-ranking individuals in primate groups take a "control role" in managing conflicts (see de Waal, 1996). Among pigtail macaques, 29% of third-party interventions were impartial: the third party targeted both combatants or stood neutrally between them (Flack et al., 2005).

Impartiality is rare in nature and poses evolutionary challenges. Importantly, moral impartiality is not explained by theories of altruism. In fact, altruism poses the problem: Why do people ever put moral impartiality above family, friends, and coalitions?

### 8.1. The concept of moral impartiality

The concept of impartiality is distinct from being unbiased or consistent (Gert, 2005). A baseball umpire might be biased toward a narrow strike zone, but can still be impartial with respect to the two teams. An inconsistent umpire, randomly switching strike zones, can also be impartial. Instead, the concept of impartiality requires the specification of the relevant group and the respect in which the actor is impartial (Gert, 2005). Gert (2005) proposed the definition: "A is impartial in respect R with regard to group G if and only if A's actions in respect R are not influenced by which member(s) of G benefit or are harmed by these actions" (p. 132). Moral impartiality is the specific case in which moral rules are applied impartially to the group of moral agents (Gert, 2005).

### 8.2. Field and laboratory observations

When US soldiers were asked whether they would report a unit member for intentional and unnecessary violence against innocent foreign civilians, 45% said they would not (Morgan, 2007). On the view presented here, what requires explanation is the 55% who claimed they *would* report a unit member. Why would someone condemn fellow group members for violations against strangers?

Moral impartiality is reflected in various cultural forms. In philosophy, it is seen in Kant's (1785/1993) universalizability criterion, Smith's (1759) impartial spectator, and Rawls' (1971) veil of ignorance. Impartiality is reflected in age-old principles of justice such as *audi alteram partem* ("hear both sides"). The Bible states, "Wrongdoers will pay for the wrong they do; there will be no favoritism" (Colossians 3:25). Modern superheroes are portrayed as impartial, even saving evildoers to turn them over to the authorities (despite the fact that leaving villains alive endangers humanity).

The ideal of moral impartiality is present cross-culturally. Judges among the Lozi of Zambia announced to litigants that they decide, "not by *sobozi* (partiality or prejudice), but by *bupaki* (evidence), *lisupo* (indications, probabilities, presumptions), and *libaka* (reasons, reasoning)" (Gluckman, 1967, p. 105). In Gebusi sorcery inquests, sickness-causing sorcerers were identified by spiritual mediums who, to prevent partiality, would often be drawn from an outside community (Knauff, 1987).

The costs of impartiality are visible in cultural means of mitigating them. Among the Ashanti of Ghana, people believe that the Chief's decisions channel the ancestral spirits, rather than reflecting favoritism. Rattray (1929) reports, "His judgments were regarded, not as emanating from his own mouth, but as being the decisions of his dead ancestors. He was thus placed largely beyond the risk of opprobrium, which a just decision, but one displeasing to a certain and possibly powerful faction, might have occasioned" (p. 289). Among the Azande of Central Africa, chiefs avoid close blood-brotherhood relationships to help maintain impartiality: "Chiefs have to settle cases and dispense justice and direct administration. An alliance of blood would militate against the fairness of their judgments and paralyze their execution" (Evans-Pritchard, 1933, p. 374). Finally, impartiality is sometimes used for only the most severe offenses. Among the Ifugao of the Philippines, people are obliged to support kin in all disputes, but "the only exception" to this rule is sorcery, an offense so severe that people turn against family members (Barton, 1919, p. 70).

Turning to laboratory research, Lieberman and Linke (2007) examined the influence of kinship and group membership on third-party judgments (wrongness and punishment). Wrongness judgments were not influenced by social category. However, perpetrators of theft were judged as deserving less punishment (but not less wrong) when they were kin (versus nonkin) and when they were ingroup (versus outgroup) members. This might indicate that wrongness judgments tend to be more impartial than punishment decisions. However, wrongness ratings were

close to the maximum value, so this result might simply reflect a ceiling effect.

Van Prooijen (2006) found that impartiality is affected by the certainty of guilt. Participants judged ingroup and outgroup perpetrators for violations in which guilt was uncertain or certain. When guilt was uncertain, participants showed less retributive emotion for ingroup than outgroup violators (ingroup bias), but the reverse occurred (outgroup bias) when guilt was certain.

Bernhard, Fischbacher, and Fehr (2006) looked at ingroup partiality among two indigenous groups in Papua New Guinea. In the first stage of an experimental game, Player A divided money with Player B (Dictator Game). In the second stage, a third party could punish player A for making an unfair division; for each monetary unit spent on punishment, Player A's payoff was reduced by three units. There were four conditions manipulating ingroup membership: The punisher was either in the same group as the (potential) perpetrator and victim, the perpetrator only, the victim only, or neither. Interestingly, perpetrators' group memberships did not affect punishment decisions. People did punish more, however, when the victim was an ingroup member (versus outgroup member).

### 8.3. Loyalty vs. moral impartiality

Accounts that link morality with loyalty and personal ties (e.g., Fletcher, 1993; Gilligan, 1977) are undermined by the inevitable conflict between altruism and moral impartiality (see also Batson et al., 1995). Loyalty and personal caring are *alternatives* to moral impartiality. This tension is illustrated by the Brazilian maxim, "For friends, everything; for enemies, the law," and the colloquialism, "My country right or wrong." One cannot be both a good altruist and a good (impartial) moralist. For this reason, it would seem that people should prefer partial rather than impartial partners. The possibility of false accusation highlights the crucial difference. With ambiguous evidence, loyal altruists, not impartial condemners, will grant the benefit of the doubt. In fact, the potentially lethal threat of false accusation – an inevitable consequence of moral judgment – should favor the evolution of intensified loyalties to family, friends, and groups to counteract the threat of impartial condemnation.

Despite this, however, humans across cultures advocate neglecting loyalties when making moral judgments. When a moralist condemns a perpetrator, the moralist is judged by the yardstick of impartiality. Further, condemners sometimes act on the ideal, damaging vital relationships, even endangering their groups. Why do people praise moral impartiality, and, moreover, why do they sometimes put morality above their relationships?

## 9. Conclusion

Morality has been the focus of deep meditation for millennia, from Aristotle to Kant to modern theorists. The rise of Darwinism added new questions about morality. "What actions are morally wrong?" and, "What is the nature of

morality?" expanded to include, "What are the evolved functions of moral adaptations?"

In answering this last question, we think it is helpful to distinguish morality from altruism. Books that purport to explain the human *Origins of Virtue*, or why humans are *Moral Animals*, should not be judged by (the words on) their covers. Explanations of prosociality or altruism do not explain – or minimally, do not necessarily explain – all moral phenomena. Indeed, moral judgment is often nonconsequentialist, casting doubt on theories that posit direct (non-strategic) consequentialist functions, including altruism. More generally, the problem of morality has been framed in terms of actor conscience. We argue that the condemnation-centered perspective potentially holds additional value.

The moral dimension of the human mind centers around the irreducible concepts of right and wrong. The surrounding cognitive architecture remains constant while the content of moral rules varies across time and culture. People constantly negotiate which moral rules to observe, but the meaning of these debates depends on the unique and universal implications of moral judgment (third-party monitoring, gossip, punishment, etc.; actor concealment, justification, apology, etc.). The consistent operation of moral cognition across diverse offense types (bestiality, black magic, cannibalism, etc.) might reflect overarching functions that are not directly tied to rule content.

Moral condemnation often involves three players: perpetrator, victim, and condemner. Consideration of this strategic dynamic shows that actors and third-parties face distinct adaptive problems. Conscience and condemnation, therefore, are likely performed by different cognitive systems. Further, if moral condemnation is performed by mechanisms for implementing strategic moves in a multi-player game, then understanding this strategic dynamic will illuminate the functions of condemnation.

In conclusion, we suggest that humans have systems specifically designed for condemnation and, in turn, conscience functions as a defense system, mirroring condemnation computations to anticipate and avoid punishment. The condemnation-centered view focuses research on the question: *How does moral cognition benefit third-party condemners?* More specifically, *how do the cognitive systems underlying judgment, punishment, and impartiality work together to yield the benefits that offset their costs?* In short, how can we explain the psychology symbolized by Lady Justice's scale, sword, and blindfold? Asking these questions might help solve the multiple mysteries of morality.

## Acknowledgments

We would like to thank Bryan Bencomo, Jon Haidt, Greg Hall, Jessamyn Haupt, Dennis Krebs, Nicole Ruedy, Brendan Ryan, Elske Straver, and Chris Taylor for helpful comments. The writing of this manuscript was supported by a predoctoral fellowship to the first author from the International Foundation for Research in Experimental Economics (IFREE).

## References

- Alexander, R. A. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Arthur, T. C. (2000). A workable rule of reason: A less ambitious antitrust role for the federal courts. *Antitrust Law Journal*, 68, 337–389.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71, 364–374.
- Barkow, J. H. (1992). Beneath new culture is old psychology: Gossip and social stratification. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 627–637). New York: Oxford University Press.
- Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences*, 17, 1–10.
- Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making*. Oxford: Oxford University Press.
- Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes*, 59, 475–498.
- Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, 94, 74–85.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70, 1–16.
- Baron, J., Granato, L., & Spranca, M. (1993). Decision-making biases in children and early adolescents: Exploratory studies. *Merrill-Palmer Quarterly*, 39, 22–46.
- Barton, R. F. (1919). *Ifugao law*. Berkeley: University of California Press.
- Batson, C. D. (2008). Moral masquerades: Experimental exploration of the nature of moral motivation. *Phenomenology and the Cognitive Sciences*, 7, 51–66.
- Batson, C. D., & Thompson, E. R. (2001). Why don't moral people act morally? Motivational considerations. *Current Directions in Psychological Science*, 10, 54–57.
- Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology*, 68, 1042–1054.
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice. Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72, 1335–1348.
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H., & Strongman, J. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, 77, 525–537.
- Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442, 912–915.
- Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior*, 10, 95–121.
- Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171–195.
- Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.
- Boyer, P., & Lienard, P. (2006). Why ritualized behavior? Precaution systems and action-parsing in developmental, pathological and cultural rituals. *Behavioral and Brain Sciences*, 29, 1–56.
- Bramble, D. M., & Lieberman, D. E. (2004). Endurance running and the evolution of Homo. *Nature*, 432, 345–352.
- Brooke, M. de L., & Davies, N. B. (1988). Egg mimicry by cuckoos *Cuculus canorus* in relation to discrimination by hosts. *Nature*, 335, 630–632.
- Carpenter, J., & Matthews, P. H. (2005). *Norm enforcement: Anger, indignation, or reciprocity?* Unpublished manuscript.
- Chekroun, P., & Brauer, M. (2002). The bystander effect and social control behavior: The effect of the presence of others on people's reactions to norm violations. *European Journal of Social Psychology*, 32, 853–866.
- Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, 373, 209–216.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind*. New York: Oxford University Press.
- Cosmides, L., & Tooby, J. (2000). Consider the source. The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 53–115). NY: Oxford University Press.
- 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.
- Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, 41, 525–556.
- Darwin, C. (1871). *Descent of man, and selection in relation to sex*. New York: D. Appleton and Company.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioral ecology* (pp. 282–309). Sunderland, MA: Sinauer Associates.
- De Backer, C. (2005). *Like Belgian chocolate for the universal mind: Interpersonal and media gossip from an evolutionary perspective*. Unpublished doctoral dissertation, Ghent University, Ghent, Belgium.
- De Moraes, C. M., Lewis, W. J., Pare, P. W., Alborn, H. T., & Tumlinson, J. H. (1998). Herbivore-infested plants selectively attract parasitoids. *Nature*, 393, 570–573.
- de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.
- de Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology*, 9, 101–118.
- den Bak, I. M., & Ross, H. S. (1996). I'm telling! The content, context and consequences of children tattling on their siblings. *Social Development*, 5, 292–309.
- DeScioli, P. (2008). *Investigations into the problems of moral cognition*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- DeScioli, P., & Kurzban, R. (2008). Cracking the superhero's moral code. In R. Rosenberg (Ed.), *The psychology of superheroes* (pp. 245–259). Dallas, TX: BenBella Books.
- Douglas, M. (1966). *Purity and danger: An analysis of the concepts of pollution and taboo*. London: Routledge & Kegan Paul.
- Doutrelant, C., McGregor, P. K., & Oliveira, R. F. (2001). The effect of an audience on intrasexual communication in male Siamese fighting fish, *Betta splendens*. *Behavioral Ecology*, 12, 283–286.
- Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. Cambridge: Harvard University Press.
- Dunbar, R. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, 8, 100–110.
- Dunn, J., & Munn, P. (1985). Becoming a family member: Family conflict and the development of social understanding in the second year. *Child Development*, 56, 480–492.
- Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology* (pp. 701–778). Toronto, Ontario: Wiley.
- Engh, A. L., Siebert, E., Greenberg, D., & Holekamp, K. E. (2005). Patterns of alliance formation and post-conflict aggression indicate spotted hyenas recognize third-party relationships. *Animal Behaviour*, 69, 209–217.
- Evans-Pritchard, E. E. (1933). Zande blood-brotherhood. *Africa*, 6, 369–401.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25, 63–87.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13, 1–25.
- Ferguson, T. J., & Gail, B. (1988). Children's evaluations of retaliatory aggression. *Child Development*, 59, 961–968.
- 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.
- Fessler, D. M. T., & Navarrete, C. D. (2004). Third-party attitudes toward sibling incest: Evidence for Westermarck's Hypotheses. *Evolution and Human Behavior*, 25, 277–294.
- Fiddick, L. (2004). Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literatures. *The Quarterly Journal of Experimental Psychology*, 57A, 447–474.
- Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology*, 36, 267–278.
- Flack, J. C., de Waal, F. B. M., & Krakauer, D. C. (2005). Social structure, robustness, and policing cost in a cognitively sophisticated species. *American Naturalist*, 165, E126–E139.
- Fletcher, G. P. (1993). *Loyalty: An essay on the morality of relationships*. Oxford: Oxford University Press.
- Foster, K. R., & Ratnieks, F. L. (2001). Convergent evolution of worker policing by egg eating in the honeybee and common wasp. *Proceedings of the Royal Society B*, 268, 169–174.



- Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: W.W. Norton & Co..
- Freud, S. (1918). *Totem and taboo*. New York: Vintage Books.
- Frith, C. (1996). The role of the prefrontal cortex in self-consciousness: The case of auditory hallucinations. *Philosophical Transactions of the Royal Society of London B*, 351, 1505–1512.
- Gao, X., & Jiang, L. (2004). Water-repellent legs of water striders. *Nature*, 432, 36.
- Gelfand, D. M., Hartmann, D. P., Walder, P., & Page, B. (1973). Who reports shoplifters? A field-experimental study. *Journal of Personality and Social Psychology*, 25, 276–285.
- Gert, B. (2005). *Morality: Its nature and justification*. New York: Oxford University Press.
- Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking Press.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gilligan, C. (1977). In a different voice. Women's conceptions of self and of morality. *Harvard Educational Review*, 47, 481–517.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.
- Gluckman, M. (1963). Gossip and scandal. *Current Anthropology*, 4, 307–316.
- Gluckman, M. (1967). *The judicial process among the Barotse of Northern Rhodesia*. Manchester: Manchester University Press.
- Gobin, B., Billen, J., & Peeters, C. (1999). Policing behaviour towards virgin egg layers in a polygynous ponerine ant. *Animal Behaviour*, 58, 1117–1122.
- 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.
- Griffin, A. S., West, S. A., & Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature*, 430, 1027–1035.
- Grueneich, R. (1982). The development of children's integration rules for making moral judgments. *Child Development*, 53, 887–894.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.
- Haidt, J., & Baron, J. (1996). Social roles and the moral judgment of acts and omissions. *European Journal of Social Psychology*, 26, 201–218.
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 62, 221–227.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus* (Fall), 55–66.
- 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, pp. 367–391). New York: Oxford University Press.
- 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.
- Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology*, 33, 1–36.
- Hamilton, W. (1964). The genetic evolution of social behaviour. *Journal of Theoretical Biology*, 7, 1–52.
- Hamilton, W. D. (2001). *Narrow roads of gene land. The evolution of sex* (Vol. 2). Oxford: Oxford University Press.
- Harari, H., Harari, O., & White, R. V. (1985). The reaction to rape by American male bystanders. *The Journal of Social Psychology*, 125, 653–658.
- Hauser, M. D. (2006). *Moral minds*. New York: HarperCollins.
- Haviland, J. B. (1977). *Gossip, reputation, and knowledge in Zinacantan*. Chicago: University of Chicago Press.
- Helwig, C. C., & Jasiobedzka, U. (2001). The relation between law and morality: Children's reasoning about socially beneficial and unjust laws. *Child Development*, 72, 1382–1393.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770.
- Hess, N.C., & Hagen, E.H. (2002). *Informational warfare*. Unpublished manuscript.
- Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. In J. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 307–326). Boston: MIT Press.
- Holldobler, B., & Wilson, E. (1990). *The ants*. Cambridge: Harvard University Press.
- Howard, W., & Crano, W. D. (1974). Effects of sex, conversation location, and size of observer group on bystander intervention in a high risk situation. *Sociometry*, 37, 491–507.
- Hume, D. (1783). *Essays on suicide and the immortality of the soul*. London: Kearsley.
- Inbred obscurity: Improving incest laws in the shadow of the "sexual family". (2006). *Harvard Law Review* 119, 2464–2485.
- Kadri, S. (2005). *The trial*. New York: Random House.
- Kant, I. (1993). *Grounding for the metaphysics of morals*. (J.W. Ellington, Trans.) Indianapolis: Hackett Publishing Company. (Original work published 1785).
- Kaplan, H., & Hill, K. (1985). Food sharing among ache foragers: Tests of explanatory hypotheses. *Current Anthropology*, 26, 223–245.
- Kaplow, L., & Shavell, S. (2002). *Fairness versus welfare*. Cambridge: Harvard University Press.
- Keltikangas-Järvinen, L., & Lindeman, M. (1997). Evaluation of theft, lying, and fighting in adolescence. *Journal of Youth and Adolescence*, 26, 467–483.
- Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, 10, 155–171.
- Knauff, B. M. (1987). Reconsidering violence in simple human societies: Homicide among the Gebusi of New Guinea. *Current Anthropology*, 28, 457–500.
- Kohlberg, L. (1981). *The philosophy of moral development*. San Francisco: Harper & Row.
- Krebs, J. R., & Davies, N. B. (1993). *An introduction to behavioral ecology* (3rd Ed). Oxford: Blackwell Scientific Publications.
- Krebs, D. L., & Janicki, M. (2004). The biological foundations of moral norms. In M. Schaller & C. Crandall (Eds.), *Psychological foundations of culture* (pp. 125–148). Hillsdale, NJ: Erlbaum.
- Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75–84.
- Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, 89, 308–324.
- Laupa, M., & Turiel, E. (1986). Children's conceptions of adult and peer authority. *Child Development*, 57, 405–412.
- Ledyard, J. (1995). Public goods: A survey of experimental research. In A. Roth & J. Kagel (Eds.), *Handbook of experimental economics*. New York: Princeton University Press.
- Leon, M. (1984). Rules mothers and sons use to integrate intent and damage information in their moral judgments. *Child Development*, 55, 2106–2113.
- 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*, 270, 819–826.
- Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, 445, 727–731.
- Macnamara, J. (1991). The development of moral reasoning and the foundations of geometry. *Journal for the Theory of Social Behaviour*, 21, 125–150.
- Marlatt, G. A. (1996). Harm reduction: Come as you are. *Addictive Behaviors*, 21, 779–788.
- Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. New York: Cambridge University Press.
- Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford: Oxford University Press.
- Mencken, H. L. (1949). *A Mencken chrestomathy*. New York: Knopf.
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, 11, 143–152.
- Miller, G. F. (2007). Sexual selection for moral virtues. *Quarterly Review of Biology*, 82, 97–125.
- Miller, G. P. (2003). Norm enforcement in the public sphere: The case of handicapped parking. *George Washington Law Review*, 71, 895–933.
- Mock, D. W. (2004). *More than kin and less than kind*. Cambridge, MA: Harvard University Press.
- Moore, G. E. (1903). *Principia ethica*. New York: Barnes and Noble Publishing.

- Morgan, D. (2007). US Marines unlikely to report civilian abuse: Study. *Reuters*.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, 52, 114–121.
- Nuñez, M., & Harris, P. L. (1998). Psychological and deontic concepts: Separate domains or intimate connection? *Mind & Language*, 13, 153–170.
- Piaget, J. (1932). *The moral judgment of the child*. London: Routledge & Kegan Paul.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: HarperCollins.
- Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39, 653–660.
- Planty, M. (2002). *Third-party involvement in violent crime, 1993–1999 (US Bureau of Justice Statistics Special Report NCJ 189100)*. Washington, DC: US Department of Justice.
- Plato. (2004). *Laws* (B. Jowett, Trans.). Whitefish, MT: Kessinger Publishing. (Original work published 4th century BC).
- Rattray, R. S. (1929). *Ashanti law and constitution*. Oxford: Clarendon Press.
- Rawls, J. (1971). *A theory of justice*. Oxford: Oxford University Press.
- Reiter, J., Stinson, N. L., & Le Boeuf, B. J. (1978). Northern elephant seal development: The transition from weaning to nutritional independence. *Behavioral Ecology and Sociobiology*, 3, 337–367.
- Rekart, M. L. (2005). Sex-work harm reduction. *Lancet*, 366, 2123–2134.
- Ridley, M. (1996). *The origins of virtue*. London: Viking, Penguin Books.
- Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3, 263–277.
- Ritov, I., & Baron, J. (1992). Status-quo and omission bias. *Journal of Risk and Uncertainty*, 5, 49–61.
- Ritov, I., & Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and Human Decision Processes*, 64, 119–127.
- Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes*, 79, 79–94.
- Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law*. San Francisco: Westview Press.
- Robinson, P., & Kurzban, R. (2007). Concordance and conflict in intuitions of justice. *Minnesota Law Review*, 91, 1829–1907.
- Robinson, P., Kurzban, R., & Jones, O. D. (2007). The origins of shared intuitions of justice. *Vanderbilt Law Review*, 60, 1631–1688.
- 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.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15, 165–184.
- Rozin, P. (1999). The process of moralization. *Psychological Science*, 10, 218–221.
- Rozin, P., & Singh, L. (1999). The moralization of cigarette smoking in the United States. *Journal of Consumer Psychology*, 8, 339–342.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76, 574–586.
- 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.
- Rucker, D. D., Polifroni, M., Tetlock, P. E., & Scott, A. L. (2004). On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin*, 30, 673–684.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Schino, G. (2007). Grooming and agonistic support: A meta-analysis of primate reciprocal altruism. *Behavioral Ecology*, 18, 115–120.
- Sedlak, A. J. (1979). Developmental differences in understanding plans and evaluating actors. *Child Development*, 50, 536–560.
- Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, 94, 439–454.
- Shotland, R. L., & Straw, M. K. (1976). Bystander response to an assault: When a man attacks a woman. *Journal of Personality and Social Psychology*, 34, 990–999.
- Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–83). Chicago: University of Chicago Press.
- 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. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York: Routledge.
- Simoons, F. J. (1994). *Eat not this flesh: Food avoidances from the prehistory to the present*. Madison: University of Wisconsin Press.
- Sinervo, B., & Lively, C. M. (1996). The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, 380, 240–243.
- Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. New York: Farrar, & Giroux.
- 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.
- Sinnott-Armstrong, W. (2006). Consequentialism. *Stanford Encyclopedia of Philosophy*. <<http://plato.stanford.edu>>.
- 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. (1993). Preschoolers' judgments about hypothetical and actual transgressions. *Child Development*, 64, 202–214.
- Smith, A. (1759). *The theory of moral sentiments*. London: A. Millar, A. Kincaid & J. Bell.
- Smuts, B. B., & Smuts, R. W. (1993). Male aggression and sexual coercion of females in nonhuman primates and other mammals: Evidence and theoretical implications. *Advances in the Study of Behavior*, 22, 1–63.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27, 76–105.
- Stengers, J., & Van Neck, A. (2001). *Masturbation: The history of a great terror*. New York: Palgrave/St. Martins.
- Stich, S. (2006). Is morality an elegant machine or a kludge? *Journal of Cognition and Culture*, 6, 181–189.
- Sunstein, C. (2005). Moral heuristics. *Behavioral and Brain Sciences*, 28, 531–543.
- Takabayashi, J., & Dicke, M. (1996). Plant-carnivore mutualism through herbivore-induced carnivore attractants. *Trends in Plant Science*, 1, 109–113.
- Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law and Human Behavior*, 26, 395–415.
- Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law and Human Behavior*, 28, 411–435.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, 58, 345–372.
- Tetlock, P. E. (2000). Coping with trade-offs: Psychological constraints and political implications. In S. Lupia, M. McCubbins, & S. Popkin (Eds.), *Political reasoning and choice*. Berkeley: University of California Press.
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Science*, 7, 320–324.
- Tooby, J., & Cosmides, L. (1992). Psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 19–136). New York: Oxford University Press.
- Tracy, J. L., & Robins, R. W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin*, 32, 1339–1351.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.
- Turiel, E. (1998). The development of morality. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology* (pp. 95–130). New York: Academic Press.
- Turillo, C. J., Folger, R., Lavelle, J. J., Umphress, E. E., & Gee, J. O. (2002). Is virtue its own reward? Self-sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes*, 89, 839–865.
- 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.
- 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.
- Walker, M., & Wooders, J. (2001). Minimax play at Wimbledon. *The American Economic Review*, 91, 1521–1538.
- Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 31, 1301–1303.

- White, D. J., & Galef, B. G. Jr., (1999). Mate choice copying and conspecific cueing in Japanese quail, *Coturnix coturnix japonica*. *Animal Behaviour*, 57, 465–473.
- Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi Bushmen: A case of strong reciprocity? *Human Nature*, 16, 115–145.
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton: Princeton University Press.
- Williams, G. C. (1975). *Sex and evolution*. Princeton: Princeton University Press.
- Wilson, J. Q. (1993). *The moral sense*. New York: Free Press.
- Wilson, D. S., Wilczynski, C., Wells, A., & Weiser, L. (2000). Gossip and other aspects of language as group-level adaptations. In C. Heyes & L. Huber (Eds.), *The evolution of cognition* (pp. 347–365). Cambridge: MIT Press.
- Wright, R. (1994). *The moral animal*. New York: Pantheon.
- Zahn-Waxler, C., Radke-Yarrow, M., & Wagner, E. (1992). Development of concern for others. *Developmental Psychology*, 28, 126–136.