Psychological Bulletin

Deontological Coherence: A Framework for Commonsense Moral Reasoning

Keith J. Holyoak and Derek Powell

Online First Publication, October 6, 2016. http://dx.doi.org/10.1037/bul0000075

Psychological Bulletin

0033-2909/16/$12.00 http://dx.doi.org/10.1037/bul0000075

Deontological Coherence: A Framework for Commonsense Moral Reasoning

Keith J. Holyoak and Derek Powell
University of California, Los Angeles

We review a broad range of work, primarily in cognitive and social psychology, that provides insight into the processes of moral judgment. In particular, we consider research on pragmatic reasoning about regulations and on coherence in decision making, both areas in which psychological theories have been guided by work in legal philosophy. Armed with these essential prerequisites, we sketch a psychological framework for how ordinary people make judgments about moral issues. Based on a literature review, we show how the framework of *deontological coherence* unifies findings in moral psychology that have often been explained in terms of a grab-bag of heuristics and biases.

Keywords: coherence-based reasoning, deontology, dual-process theories, moral judgment, pragmatic reasoning schemas

In this article we review a wide range of literature that bears on the question of how ordinary people make judgments about moral issues. We do not aim to provide a full review of the field of moral psychology, both because the scope of the field is prohibitively broad, and because a number of excellent recent reviews are available (Curry, 2016; Greene, 2014; Haidt & Kesebir, 2010; Machery & Mallon, 2010; Waldmann, Nagel, & Wiegmann, 2012). Rather, we selectively review work that bears on a framework for understanding "commonsense" moral reasoning based on what we refer to as *deontological coherence*. This framework originates in research that was not explicitly directly at moral reasoning, but rather at deductive reasoning and complex decision making. In addition to work in psychology, the approach is rooted in moral and legal philosophy. The scope of our article is constrained accordingly. We will first review work in psychology and philosophy that provides a background for the framework of deontological coherence, and then research directly in moral psychology that serves to test predictions derived from the framework. Based on our literature review, we argue that the framework of deontological coherence unifies findings in moral psychology that have often been explained in terms of a grab-bag of heuristics, biases, and errors.

Our framework is a *descriptive* one, meant to explain how ordinary people think about moral questions. This descriptive project has been guided and informed by work on *normative* theories, which have the deeper aim of characterizing what constitutes moral choices and actions. Normative ethical theories can potentially guide descriptive psychological theories of moral judgment in two ways: (a) they might provide a normative standard against which human moral judgments can be compared, and (b) they might offer a conceptual vocabulary or framework for a descriptive account. We argue that no specific normative theory has been rationally established as truth, ruling out (a). Accordingly, we argue that moral psychology should abandon any claim to be comparing human moral judgments against some rationally established normative standard, instead adopting the stance of *methodological atheism*. However, in accord with (b), we argue that a normative perspective termed *moderate deontology*, which stresses the need for adjudication between competing rights and duties (along with other considerations), can provide a useful conceptual framework for understanding moral judgment.

This framework, based on *deontological coherence*, makes at least three sets of predictions about human behavior in moral judgment tasks: (a) people hold (potentially sophisticated and complex) deontic moral rules that inform their moral decisions, (b) these rules are generally not inviolable but instead provide soft constraints that can be overridden by other rules or by considerations related to consequences, and (c) resolving conflict related to moral concerns is achieved through coherence-based reasoning, which yields systematic coherence shifts in relevant attitudes and evaluations.

We review evidence for these predictions as we consider deontology as a conceptual framework for morality, and coherence-based reasoning as a mechanism by which moral judgments are produced. In most cases alternative explanations have been offered (often based on proposed heuristics and biases), but our aim here is to show that deontological coherence offers a unifying explanation. We do not take

Keith J. Holyoak and Derek Powell, Department of Psychology, University of California, Los Angeles.

Zachary Horne, Dan Simon, Derek Penn, David Uttal, and three anonymous reviewers provided helpful comments on earlier drafts.

Derek Powell is now at the Department of Psychology, Stanford University.

Correspondence concerning this article should be addressed to Keith J. Holyoak, Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Los Angeles, CA 90095-1563. E-mail: holyoak@lifesci.ucla.edu

any individual study, or even the entire collection, as definitive evidence in support of our proposal. Our limited goal is to establish that the framework merits further exploration.

Normative Ethical Theories

Although many important philosophical approaches to morality can be distinguished (e.g., virtue ethics, Pence, 1991; Doris, 1998; and intuitionism, Dancy, 1991), we focus on two major views that have dominated moral psychological discussions. Following the terminology of Rawls (1971), one basic approach is *teleological*—it aims to establish what is *good* to achieve, and then defines the *right action* as that which brings about the maximal good. The alternative approach is *nonteleological*—it denies that the good is prior to, and determinant of, the right. The teleological approach is clearly simpler. We will consider the most influential of its many variants, *utilitarianism.* The nonteleological approach leads to variants of *deontology* (from the Greek, "study of duty"), which grounds moral reasoning not in maximization of the good, but in the interlocking concepts of rights and duties.

Utilitarianism

Teleological theories of morality assume that the only factor that ultimately determines the rightness of an act is its consequences, a position dubbed *consequentialism*. Theories differ in their assumptions about the relevant consequences (e.g., pleasure, happiness, or simply the satisfaction of preferences) and how they are best distributed (e.g., ethical egoism seeks the greatest good for *me*). Here we sketch the version that has had the greatest impact on moral psychology, *utilitarianism*.[1] A product of 19th-century moral philosophy, utilitarianism originated with Bentham (1823/2009), and was subsequently advanced by John Stuart Mill (1861/2004), Henry Sidgwick (1907/1981), and many other more recent philosophers (e.g., Singer, 1979, 2005). There are countless other more nuanced variants, but at its simplest utilitarianism defines the right act as that which maximizes some utility measure summed across an entire group of moral patients (i.e., those deserving of moral concern). In the context of research on moral psychology, utility is typically operationalized in terms of straightforward concerns such as the potential loss of human life (typically in situations where unknown groups of strangers are at risk), thereby allowing moral psychologists to sidestep more nuanced questions about how "utility" ought to be defined and distributed.

As the term "utility" connotes, utilitarianism is closely linked to modern economic theories based on "expected utility," and to the concept of cost-benefit analysis. Utilitarianism's close linkage to economic theory has bolstered its attraction as a normative theory. It is amenable to mathematical analyses of moral decisions as optimization problems (e.g., a state of affairs may be defined as "Pareto optimal" if it is impossible to make any individual better off without making at least one individual worse off). Within the field of judgment and decision making, research on moral decisions has been largely treated as an extension of work on nonmoral decision making. For individual decision making (in the absence of apparent moral concerns), an economic theory (expected utility) has been treated as the normative theory, against which background much of actual human decision making has been characterized in terms of apparent deviations from optimality, reflecting the use of suboptimal "heuristics and biases" (for a review see Griffin et al., 2012).

By analogy, a number of influential moral psychologists (e.g., Baron, 1994; Greene, 2008) have treated utilitarianism as the normative moral standard against which human moral reasoning is to be evaluated. Where human moral judgments deviate from utilitarianism, researchers have argued that these judgments must be based on heuristics (Sunstein, 2005), which induce attendant biases and errors. Some have not only accepted this interpretation, but have viewed this psychological account of deontological judgments as an indictment of that alternative normative theory. Greene (2008, p. 36) proposed to put normative moral philosophy to empirical test: "I will argue that deontological judgments tend to be driven by emotional responses, and that deontological philosophy, rather than being grounded in moral *reasoning*, is to a large extent an exercise in moral *rationalization*. . . . [I]f these empirical claims are true, they may have normative implications, casting doubt on deontology as a school of normative moral thought."

Deontology

As a nonteleological approach to morality, deontology is most generally defined as the denial of utilitarianism (i.e., deontology asserts the good is *not* always prior to and determinant of the right). Deontology in some form can be traced back four millennia (to the Babylonian Code of Hammurabi, and later the Ten Commandments of Moses); however, its modern version owes much to the towering 18th-century philosopher Immanuel Kant (1785/1953). This view (see collection edited by Darwall, 2002; also Nagel, 1986; Rawls, 1971; Zamir & Medina, 2010) insists that actions can be right or wrong in and of themselves, rather than their moral value being solely dependent on their consequences. That is, the right does not necessarily maximize the good. Deontology affords even more variants than utilitarianism, as the content of deontological moral values may differ enormously. However, as we will review below, human moral judgments often evince features typical of deontological ethics, suggesting that concepts rooted in deontology provide a natural basis for a descriptive framework.

Utilitarianism is *agent-neutral*, positing that what is right for one is what is right for all in the group. In contrast, deontology is *agent-relative*—each person is responsible, first and foremost, for the moral value of their own actions. However laudable the ends, there are some actions that it would be *wrong for me to take*. Deontology is based on a folk theory of human voluntary action, according to which people consciously form goals that then direct their actions. Thus *intentions* are often crucial. To *aim* to harm someone is worse than to harm them as a side effect of promoting some permissible end, even if the bad side effect was foreseen (the famous "doctrine of double effect," originally laid out by the 13th-century Catholic philosopher Thomas Aquinas). Deontology thus implies that bombing a city to deliberately kill civilians (terror bombing) is morally worse than bombing the city to destroy munitions factories (strategic bombing), even if the same number

[1] More specifically, we focus on a version termed *maximizing act-utilitarianism* (Portmore, 2011). This variant, which is the direct moral analog of expected utility theory in economics, has had the greatest influence on moral psychology.

of civilian deaths occurs (and was foreseen) in the two cases. The doctrine of double effect does not imply that an unintended but foreseeable harm is morally good—just that it is less bad than an otherwise equivalent intended harm.

Deontology thus urges the moral agent to avoid aiming to do harm. To do good is good, but to avoid doing evil is paramount (Nagel, 1986). This is why the Hippocratic oath is often paraphrased as, "First do no harm," and Google's corporate motto is, "Don't be evil." It follows that in case of doubt, the agent should refrain from doing an immoral act (what we term the "no-action default"). The deontological emphasis on not doing harm leads to a focus on negative obligations ("Thou shalt not kill," and the like). Positive moral obligations also exist (e.g., the mariner's duty to rescue people in distress on a nearby ship; the duty of parents to provide for their children). Nonetheless, the guiding principle is that *a moral agent refrains from intentionally doing wrong acts* (see Cushman, Young, & Hauser, 2006).

The content of deontological moral values may differ enormously. Since the Enlightenment, core Western values have included human autonomy, justice, liberty, and truth, (and many other cultures have affirmed at least some of these values). These values have priority over promoting the common good (although the latter may itself be a value). For example, the value of autonomy implies that *your* interests and projects—including your concern for your family, friends, and immediate community—are granted a special status for *you*. If your life's dream is to climb Mount Everest, you may (morally) save up your money, devote much of your time to training, and go climb the mountain—you are not obliged to forego your personal (and expensive) dream to donate all your savings to charity. At the same time, other people who do not share (and may not even approve of) your life's goal need not assist you. Nobody is obliged to maximize the happiness of anyone.

In the commonsense morality of deontology, there are certain things one must not do, or must do, to stay within acceptable moral bounds. Beyond that, each of us is relatively free to lead our lives as we see fit. Some may "go beyond the call of duty" in their beneficence toward others (i.e., performing what are termed *supererogatory* acts), and that is commendable; but it is their individual choice. Unlike utilitarianism, deontology recognizes each individual person's unique moral position.

Methodological Atheism

Comparisons with normative standards have often guided the development of psychological theories at the computational level (Marr, 1982). However, it seems that no moral theory can claim the degree of rational support underlying, for example, probability theory. Among ethicists, there is little consensus on the status of utilitarian ethical theories. For example, the attractive theoretical concept of Pareto optimality, if construed as a normative claim as to what constitutes a stable moral system, implies that in a society based on institutionalized slavery not a single slave may be freed (because to do so would impose negative utility on a slave master). Critics of utilitarianism have argued that the view is suspect because it lacks any role for justice (e.g., McCloskey, 1957) or fairness (e.g., Rawls, 1971; Nagel, 1986), and because of its severe demands for self-sacrifice and its failure to respect special relationships between individuals, such as family ties (Kagan, 1989; Scheffler, 1982). Many articulate responses have been offered to these and other criticisms (e.g., see collection edited by Eggleston & Miller, 2014; Pettit, 1991; Portmore, 2011). In a similar vein, serious objections have been levied against many forms of deontology (e.g., Mill, 1861/2004; Nietzsche, 1887/1996).

Unless and until some normative ethical theory can be rationally established, we argue that moral psychologists are well-advised to adopt a version of *methodological atheism* (following Berger, 1967) or agnosticism toward normative moral claims, whether these originate in theology or in secular philosophy. Individual psychologists certainly may hold personal ethical positions, but their place in scientific inquiry should be duly constrained. Though it may be painful for researchers to make do without any clear normative standard, it is surely worse to employ one that is ill-founded.

Utilitarianism and Deontology as Conceptual Frameworks

Although we argue that none can lay claim to be the normative standard for comparisons with human behavior, normative ethical theories can still provide conceptual frameworks to guide psychological hypotheses. Importantly, these frameworks can and should include concepts derived from both deontological and utilitarian theories. There is an apparent incommensurability (normatively speaking) between teleological and nonteleological approaches to morality. In particular, deontological theories employ moral concepts (notably rights and duties) that are simply not a part of utilitarian theories. However, there is reason to believe humans possess and utilize concepts related to both deontology and to consequences. Indeed, both approaches suggest that a moral judgment may be defined as one that takes into account the value of others (e.g., Nagel, 1986). We advocate for a moral psychological framework in which the processes of moral judgment are based on domain-general principles, and the special quality of moral judgment comes from their content: deontic rules and valuations driven by concern for the well-being of others.

Despite the normative problems facing all moral theories, including utilitarianism, moral psychologists have tended to treat utilitarianism as a normative ethical standard (though for dissenting views see, e.g., Bennis, Medin, & Bartels, 2010; Mikhail, 2011), effectively relegating deontological concepts to secondary status. Deontological judgments have been depicted as errors, the occurrence of which ought to be explained, but about which relatively little else needs to be said. For example, according to the influential dual-process theory of moral judgment (Greene et al., 2001, 2004), utilitarian and deontological judgments are produced by two separate cognitive systems, the conflicting outputs of which are mediated by cognitive control centers in the brain. In this view, deontological judgments are produced by fast and evolutionarily primitive affective processes, whereas utilitarian judgments are produced by slower and evolutionarily newer cognitive processes. Deontological judgments are interpreted as moral errors (Greene, 2008) produced by lapses in cognitive control, which allow fast affective processes to win out over slower but more rational cognitive processes. This theory has stimulated a great deal of research on the role of affect and of cognitive control in moral judgment (e.g., Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Crockett, Clark, Hauser, & Robbins, 2010; Greene et al.,

2004; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Koenigs et al., 2007; Moretto, Làdavas, Mattioli, & Di Pellegrino, 2010; Cushman et al., 2012; Prehn et al., 2015; Shenhav & Greene, 2014; Treadway et al., 2014).

Unfortunately, this theory seems also to have led researchers to neglect questions about how deontological concepts, such as duties, rights, and rules, are represented and used. Many researchers apparently assume these constructs are psychologically uninteresting, as they are not considered to be the objects of cognitive reasoning processes, but instead of affective processes. However, the empirical evidence available suggests this neglect of deontological concepts by psychologists has been a misstep. The framework of deontological coherence, which we will now introduce, offers a new perspective on the psychological underpinnings of moral judgments. In accord with the rule-based structure of deontological ethics, our framework will examine the possibility that moral judgments are driven (at least in part) by rule-based reasoning. We will review rule-based reasoning in general, the structure of schemas that support intuitive rule-based inferences, as well as more specific evidence concerning rule-based reasoning in moral judgment. Our review is focused through the lens of deontological ethical theory, in an effort to counteract the prevailing bias of the field against this approach (but without, we again emphasize, making any claim for its normativity).

Deontological Coherence in Moral Judgment

Although some researchers have emphasized the complexity of moral judgment (e.g., DeScioli & Kurzban, 2012; Mikhail, 2007), within moral psychology deontological rules have often been limited to simple imperatives, such as "Do not kill." It has been argued that the application of such rules is effortless and automatic (e.g., Greene & Haidt, 2002; Haidt, 2001; Shenhav & Greene, 2014). In contrast, the framework of deontological coherence emphasizes how the concepts of rights and duties produce complex systems of moral rules and systematic relationships among those rules.

Rights and Duties

Although deontology may be formulated in various ways, we adopt the conceptualization we consider to be most psychologically natural, based on the concepts of *rights* and *duties* (see Almond, 1991), which are interdefinable with *permissions* and *obligations*. A *right* grants permission, and expands options; it says we *may* do something (or not). A *duty* imposes obligation, and constrains options; it says we *must* not do (or do) something. The modals "may" and "must" are *deontic* concepts, a term that shares its Greek root (*déon*, duty or obligation) with "deontology." The core Western values mentioned above collectively comprise what are often termed *human rights*—the rights each person is born with, simply by virtue of being a person. A rights is typically conveyed by some sort of authority for some reason, which we will refer to as its *grounds* (Almond, 1991). What authority (if any) provides the grounds for human rights is a matter of debate—notable possibilities include God, an implicit social contract among people, or a democratic government. These alternatives provide different *foundational* theories of rights, justifying why they exist. For the purposes of our descriptive framework, however, it is sufficient to posit that people accept certain rights as moral *factors* (Zamir & Medina, 2010). An important source of moral agreement among people is that alternative foundational theories often support the same moral factor. Whether we trace our values to God, Thomas Hobbes, Thomas Paine, or what our mothers taught us, we can all agree that we should not harm a fellow citizen without cause.

Rights and duties are inherently relational concepts. If others deserve the same consideration as myself, then I have a duty not to violate their human rights (their claim to such rights being as strong as mine). The concept of rights can readily be extended to more specific, contextual bonds between individuals, such as promises, contracts, and commercial transactions. A useful aspect of deontology, when formulated in terms of rights and duties, is that it makes contact with legal philosophy. Morality is far from coextensive with the law—immoral laws have certainly been imposed, and many moral principles lack legal force. Still, there certainly is overlap, and a legal system cannot deviate too far from the commonsense morality of its culture if it is to be respected and obeyed (Zamir & Medina, 2010). Hohfeld (1919) provided a classic analysis of the legal concepts of right and duties. In Hohfeld's analysis, rights and duties are *correlatives*, such that one person's right implies another's duty. For example, the right of a property owner to control access to the land they own implies my duty to avoid trespassing on it.[2]

Similarly, the framework of deontological coherence views people's moral rules as systematic products of interlocking conceptions of rights and duties. Our everyday concepts of rights and duties are closely linked to people's understanding of social regulations, which appear to be rooted in *pragmatic reasoning schemas* (Cheng & Holyoak, 1985, 1989; Cheng et al., 1986; Holyoak & Cheng, 1995a, 1995b). Pragmatic reasoning schemas are mental representations of deontic rules used to draw inferences. Such schemas are characterized as "pragmatic" because they correspond not to the formal rules of normative inference systems (such as propositional logic), but rather are attuned to accomplishing everyday goals, such as maintaining and regulating social regulations.

The theory of pragmatic reasoning schemas was originally developed to explain the puzzling effects of content on people's apparent ability to reason deductively in Wason's (1966) famous "selection task." In this task, people are asked to evaluate a conditional rule, in the form *If p then q*, by identifying which of four cards (respectively showing the cases p, q, $\sim p$ and $\sim q$) must be examined to assess whether the rule holds. According to standard propositional logic, the normative choice is to check the p card (to be sure it has q on the reverse) and the $\sim q$ card (to be sure it does *not* have p on the reverse). When the rule has arbitrary content (e.g., "If a card has a vowel on one side, then it has an even number on the other"), college students routinely fail to see the

[2] Hohfeld's (1919) analysis makes finer distinctions than just that between rights and duties. For example, a *privilege* is a conditional right that can potentially be revoked; a *power* is the right to create a further right (e.g., the power to write a will that passes an inheritance to one's heirs); and an *immunity* is a right to protection (e.g., an employee may join a union without retaliation from the employer). Although these distinctions are important ones, for our present purposes we will gloss over them, and simply use "rights" and "duties" in their most general senses.

relevance of a card showing an odd number (the $\sim q$ card). In contrast, people perform more normatively for familiar social regulations, such as "If a person is to drink alcohol, then they must be over 21 years of age" (D'Andrade, 1982).

Rather than reasoning according to conditionals as prescribed by propositional logic, people appear to be reasoning according to *permission schemas*. Schemas for reasoning about permissions and similar social regulations appear to be typically acquired without any formal training. A permission schema is not simply a list of social rules that apply in particular circumstances, but rather a general and abstract conceptual structure employed in reasoning. Cheng and Holyoak (1985) showed that specific prior knowledge of a rule is not necessary to obtain facilitation in reasoning. Facilitation can be obtained for rules interpreted as social regulations even when their content is highly abstract or not familiar. For example, the unfamiliar rule, "If the form says 'ENTERING' on one side, then the other side includes cholera among the list of diseases" yielded good performance if people were simply provided with a rationale (i.e., grounds) for the regulation (entering a country at its airport requires proof the person has obtained a cholera vaccination). These findings were interpreted in terms of pragmatic schemas for reasoning about conditional permissions and obligations.

Holyoak and Cheng (1995a) adapted Hohfeld's (1919) analysis of rights and duties as correlates to formalize the permission and obligation schemas, and to specify the relationship between them. Thus, a right of party X against party Y with respect to action A,

right (of X, against Y, re A),

implies a correlative duty of Y toward X with respect to that action,

duty (of Y, toward X, re A).

The theory of pragmatic reasoning schemas predicts that this interdefinability of rights and duties will cause the interpretation of an ambiguous rule to be agent-relative. For example, the rule, "If an employee works on the weekend, then that person gets a day off during the week" will be interpreted by the employer as a conditional permission to the employee: fulfilling the prerequisite (working on weekend) means the employee *may* have a day off. The employee, by contrast, will interpret the same rule as a conditional obligation imposed on the employer (if the prerequisite is satisfied, then the employer *must* allow a day off). And in fact, depending on the point of view assumed by the reasoner, this ambiguous rule led to an opposite pattern of choices in the selection task, with each side preferentially checking the two cases in which *their* perceived rights were at risk (see also Gigerenzer & Hug, 1992; Politzer & Nguyen-Xuan, 1992).

These findings suggest that rather than being the sole province of legal scholars, an intuitive concept of rights and duties underlies laypeople's understanding of formal and informal social regulations. Indeed, Chao and Cheng (2000) showed that even preschool children readily interpret unfamiliar conditionals as permissions (given a simple rationale), and reason accordingly in a selection-type task. Recent developmental work provides evidence that children as young as three years old understand ownership rights (Kanngiesser & Hood, 2014; see also Noles et al., 2012). Preschoolers (but not chimpanzees; Riedl, Jensen, Call, & Tomasello, 2012) respond to perceived unfairness, and will often intervene when an object is unfairly taken from a third party, preferring to return the object to its rightful owner rather than keeping it themselves (Riedl, Jensen, Call, & Tomasello, 2015).

Research on pragmatic reasoning schemas suggests that far from being a source of cognitive errors, the application of rules derived from rights and duties (e.g., the permission schema) is a fundamental function of human reasoning. These rules are applied by general and sophisticated reasoning processes. Thus the notion (pervasive in the field of moral psychology) that deontological concepts and moral rules are always unconditional imperatives, such as "Do not kill," is a gross oversimplification.

Deontological Principles Affect Moral Judgments

The deontological coherence framework assumes that moral judgments evoke psychological rules that are grounded in concerns about moral rights and duties. Indeed, people's moral judgments are influenced by a number of factors that constitute important moral concerns in most forms of deontological ethics. We review evidence that (a) people's moral judgments distinguish between doing and allowing, (b) and also between intending and merely foreseeing harm, and that (c) people show agent-relative preferences for people with whom they have close relationships.

Doing versus allowing. Deontological ethics typically assign positive rights and corresponding negative duties. For example, John's right to life obligates Jim to refrain from any action that would place John's life at risk, but does not obligate Jim to save John's life if the latter is otherwise threatened. In this respect deontological ethics generally agree with the traditional Catholic doctrine of "doing versus allowing," which holds that actively harming is worse than allowing harm to occur (Foot, 1967; Kamm, 1994; Moore, 2008; Quinn, 1989). In contrast, utilitarian ethics typically treats the distinction between doing and allowing harm as irrelevant (Kagan, 1989; Rachels, 1975). In fact, people view actively harming as morally worse than passively allowing harms to occur, even when the harm was foreseen (Baron & Ritov, 2004, 2009; Borg et al., 2006; Cushman et al., 2006, 2012; DeScioli et al., 2011; Shultz et al., 1981; Spranca et al., 1991). This is true even when controlling for perceptions of the agent's intentions (DeScioli et al., 2011; Spranca et al., 1991) and for the perceived causal force of their behavior (DeScioli et al., 2011). Though this tendency has sometimes been called "omission bias," its source appears to be a deontological principle that prohibits directly causing harm (Baron & Ritov, 2004, 2009); hence we prefer the term "no-action default."

Consider a moral dilemma in which an even trade can be made: acting will save one person but allow another person to die. For a utilitarian, the decision of whether or not to act in this situation should amount to a coin flip—there is nothing in the consequences that favors either action or inaction. In contrast, Borg et al. (2006) found that people overwhelmingly chose to refrain from acting in even-trade dilemmas, choosing to act only 9% of the time. These judgments appear to be mediated by activity in the dorsolateral prefrontal cortex (DLPFC). This area of the prefrontal cortex (sometimes accompanied by the more anterior rostrolateral area, RLPFC) is the hub of a domain-general cognitive control and working memory network also involving the parietal cortex (Duncan, 2010). Essentially, greater activation of the DLPFC means

someone is "thinking harder." The DLPFC is preferentially engaged by moral situations that contrast active and passive harms (Borg et al., 2006; Cushman et al., 2012). Interestingly, although people are often unaware of the bases of their moral (Cushman et al., 2006; Haidt & Hersh, 2001; Haidt, Koller, & Dias, 1993) and nonmoral decisions (Nisbett & Wilson, 1977), many people are able to articulate the "doing versus allowing" principle when asked to explain differences in their judgments (Cushman et al., 2006).

The claim that the no-action default is the product of a deontological principle is also supported by its exceptions. Some social roles assign positive duties to agents, requiring them to ensure the welfare of those for whom they are responsible. In situations where one agent has a duty of responsibility for others (e.g., a parent, or the conductor on a train), the no-action default is eliminated (Baron, 1994; Haidt & Baron, 1996), sometimes being replaced by a preference for action (Ritov & Baron, 1994). Rather than being the product of a simple "bias," the no-action default appears to be a systematic consequence of a deontological moral code.

Intentions. Deontology (in agreement with most legal codes) typically claims that agents should be judged not just by their actions but also by their intentions. An action *intended* to do harm is morally wrong even if it does not achieve its end, and an intentionally harmful action is morally worse than an action that produces unintended but foreseen harm. People's moral judgments turn out to be exquisitely sensitive to agents' intentions. In fact, *unsuccessful* attempted harms are judged more harshly than accidental harms that actually occur (Moran et al., 2011; Young & Saxe, 2009; Young et al., 2007). In addition, consistent with the Catholic "doctrine of double effect," people generally feel it is worse to do harm intentionally than as an unintended but foreseen consequence (Borg et al., 2006; Greene et al., 2009; Hauser et al., 2007; Moore et al., 2008; Young et al., 2007; Young & Saxe, 2009, 2011). Although alternative explanations have been advanced (e.g., Greene et al., 2001; Waldmann & Dieterich, 2007), it seems that adherence to the doctrine of double effect (or some similar rule) at least partially explains the well-known divergence of participants' judgments between the "standard trolley" and "footbridge" dilemmas (Borg et al., 2006; Greene et al., 2009; Hauser et al., 2007; Moore et al., 2008). In these dilemmas, a runaway trolley threatens to kill a number of people (usually five) who can only be saved if another person is sacrificed. Most people approve of sacrificing in the "standard trolley" dilemma, where the sacrifice is made by redirecting the threat (without intent to kill), but disapprove in the "footbridge" version, where the sacrifice is made by using the victim's body to stop the trolley (thereby intending harm).

Even young children grasp the importance of intention in moral judgment. From three years of age, children are capable of distinguishing between intentional and accidental harms when evaluating actions (Yuill & Perner, 1988), though young children generally tend to rely on outcome information more strongly than do older children. After developmental advances in their understanding of the mental states of others (theory of mind; Wimmer & Perner, 1983), intentions play an increasingly large role in children's moral judgments (Baird & Astington, 2004; Fu et al., 2014; Shultz et al., 1986; Yuill & Perner, 1988; Zelazo et al., 1996). By age six or seven, children are able to make judgments driven primarily by evaluations of intentions (Baird & Astington, 2004).

Similarly, neuroimaging studies support the role of theory-of-mind processing during adults' moral judgments. When participants made judgments about attempted harms, Young et al. (2007) observed increased activation in the right temporo-parietal junction (RTPJ), a brain area associated with theory-of-mind reasoning. Participants' tendency to forgive unintentional harms was also correlated with increased activity in the RTPJ (Young & Saxe, 2009).

Agent-relative preferences for kin, friends, and ingroup members. Utilitarianism generally does not permit moral agents to privilege the lives or well-being of their kin, friends, or other ingroup members over the lives and well-being of strangers. Deontology, by contrast, acknowledges that moral agents have special duties toward those who are closest to them, duties that often demand that these people be given preferential treatment. Once again, laypeople's moral judgments are overwhelmingly in accord with deontology. Numerous studies have shown that people are more likely to engage in altruistic behavior toward kin than toward strangers (Burnstein et al., 1994). Indeed, altruistic behavior is correlated with relatedness even among kin (Burnstein et al., 1994; Korchmaros & Kenny, 2001; Neyer & Lang, 2003; Stewart-Williams, 2007; Webster, 2003), such that more closely related kin (e.g., siblings) receive greater help than more distant relatives (e.g., cousins). However, people also value close nonkin relationships—they are often just as or more willing to help friends than kin (Cialdini et al., 1997; Kruger, 2003), posing difficulty for some evolutionary accounts of altruistic behavior, (e.g., Hamilton, 1964).

When asked to consider harmful actions, people are much less likely to approve when the victim is a family member, friend, or other ingroup member (Bleske-Rechek et al., 2010; Cikara et al., 2010; O'Neill & Petrinovich, 1998; Petrinovich, O'Neill, & Jorgensen, 1993; Swann et al., 2010; Uhlmann et al., 2009). Although the overwhelming majority of people endorse sacrificing in the standard trolley dilemma, people are much less likely to do so when the one to be sacrificed is a family member or romantic partner (Bleske-Rechek et al., 2010). Cikara et al. (2010) examined participants' judgments in variations of the "footbridge" dilemma that pitted the lives of ingroup and outgroup members against one another. Princeton students were more willing to sacrifice an extreme outgroup member (a homeless person) than an ingroup member (a fellow student). In this dilemma people generally disapprove of sacrificing strangers, but 84% of participants said it was permissible to sacrifice an extreme outgroup member to save ingroup members.

Interestingly, decisions about extreme outgroup members engaged areas of the brain associated with cost-benefit analysis, such as the ventromedial PFC (VMPFC) and also DLPFC, to a greater extent than decisions in which an ingroup member was to be sacrificed, or where outgroup members were saved. These findings are rather paradoxical from a utilitarian standpoint: presumably the difference in the tradeoff is greatest for extreme cases, making cost-benefit calculations easier. However, Cikara et al.'s (2010) interpretation fits well with the framework of deontological coherence: Cost–benefit analyses are generally not even performed when considering the sacrifice of a valued ingroup member, because of the strong deontological prohibition against harming such people. Conversely, prohibitions against sacrificing a devalued

outgroup member are weaker, allowing the action to be considered seriously and hence triggering cost-benefit analyses.

Although people are generally less willing to sacrifice in-group members than out-group members, Kurzban, DeScioli, and Fein (2012) found that people were actually *more* willing to sacrifice one brother to save five. One potential account of this finding arises out of the different positive and negative duties owed to kin and strangers. Whereas we primarily have a negative duty not to harm strangers, we have strong positive duties to help kin members (and in that context, saving *more* kin could be favored by deontological as well as consequentialist considerations).

Moderate Deontology

Deontological "rules" have often been taken to be absolutist—"thou shalt not" does not lightly admit exceptions. Kant's (1785/1953) "Categorical Imperative"—"act only in accordance with that maxim through which you can at the same time will that it become a universal law"—also suggests that moral rules are absolute. This absolutism has consequences that appear unreasonable to many. Not only does it seem to imply that no prohibited act can ever be taken (even though in some particular circumstance it would generate tremendous common good), but it makes it difficult or impossible to adjudicate among conflicting duties. In a standard example, imagine that in Nazi Germany you are sheltering several Jews in your attic, and a Nazi soldier knocks on the door and asks you if there are any Jews in your home. An absolutist interpretation of a duty to tell the truth appears to imply that you must answer in the affirmative, even though you thereby betray your guests and send them to their deaths (but see O'Neill, 1991, for a discussion of Kant's views and responses to critics).

Although moral psychologists have typically taken absolutism to be a fundamental property of deontology, various modern deontologists have in fact urged a *moderate* standard for moral rules (Ross, 1930; Zamir & Medina, 2010; see Davis, 1991). Moderate rules might better be characterized as moral factors, constraints, or concerns. This move is essential to the application of deontological concepts to the development of a descriptive framework for moral reasoning. People are forever faced with conflicts between the multiple duties to which they are subject; and even if they are not utilitarians, they do not always refuse to take a "wrong" action if it would lead to great benefits for many. A moral concern is something that should be taken seriously, and can serve as a valuable default rule (cf. Holland et al., 1986). But when the full context is considered in the dilemma just described, lying is a lesser wrong than betrayal—thus it is right to lie to the Nazi soldier.

An important point, and one that has often been missed in moral psychology, is that the move to moderate deontology is entirely consistent with the standard use of "rule" as a psychological construct. In a classic analysis of the concept of psychological rules, Smith, Langston, and Nisbett (1992) proposed several criteria for psychological "rule-following." These include being applicable to abstract and to unfamiliar content, and enabling transfer to new content domains after training—criteria met by the rules that form the permission and obligation schemas (Nisbett et al., 1987). Notably absent from Smith et al.'s set of psychological criteria is any notion that a rule must be absolute.

Duties Serve as Soft Constraints on Moral Judgments

Different moral duties receive different weights. The deontological coherence framework assumes, in accord with moderate forms of deontology, that moral rules may be given different weights, leading some moral duties to be more valued than others. Thus, deontological coherence implies that moral acts are not simply categorized as "permissible" or "impermissible" in binary fashion. Rather, the severity of different moral violations is expected to lie along a continuum.

In an early psychological study of moral judgment, Thurstone (1927) asked participants to make paired comparisons of moral violations by choosing which of a pair was more severe. The derived representation of moral "wrongness" formed a unidimensional magnitude scale. More recent research on people's "moral foundations" has shown that moral values are given different weights both within and across individuals (e.g., Haidt & Graham, 2007). For instance, Graham et al. (2009) found that political liberals and conservatives both place high weights on harm and fairness, whereas authority and purity are more strongly weighted by conservatives than liberals.

Moral duties may be violated to maximize utility or honor stronger duties. People are often willing to violate moral prohibitions when doing so would lead to sufficiently better consequences. A natural way of explaining their willingness to do so is to imagine that this willingness depends on the strength of the rule and on the difference in consequences associated with violating versus adhering to it. Famously, people are less willing to sacrifice in the "footbridge" dilemma than in the standard trolley dilemma, even when the consequences are equated across the two variants (e.g., Hauser et al., 2007). In the standard trolley dilemma, the saving of five versus one usually provides a sufficient net benefit to justify violating a rule against doing foreseeable harm, whereas in the footbridge version the same net benefit is usually *not* sufficient to violate the stronger rule prohibiting intentional killing. Yet people are more willing to sacrifice the one when the number of people to be saved is increased, even in the footbridge dilemma (Bartels, 2008; Rai & Holyoak, 2010; Trémolière & Bonnefon, 2012). The weighting of different moral concerns can also be manipulated by situational factors. For example, focusing participants' attention on moral rules ("do not kill" or "save lives") in one moral situation can carry-over and affect subsequent judgments (Broeders et al., 2011).

If duties have different weights, then one moral rule may be violated in favor of satisfying another, stronger rule. The famous experiments by Milgram (1963, 1965, 1974; Burger, 2009) provide an especially dramatic example: people are often willing to violate a moral rule prohibiting harm to satisfy a perceived duty to obey an authority figure. Kohlberg (1963) examined how children and adults adjudicate between conflicting duties (e.g., being honest vs. keeping secrets). Kohlberg argued that the ability to override certain moral obligations in favor of others is an important marker of moral development.

Rai and Fiske (2011) hypothesized that moral reasoning serves largely to regulate different social relationships, and review anthropological evidence indicating that the relative priority agents place on different moral rules often depends upon their relationship with the relevant moral patients. For example, rules about fairness and equality are relaxed when one person is in a role of

authority. Hoffman et al. (1994) found that players in an ultimatum game were more likely to make and accept less generous offers when their roles had been determined by their relative performance on an earlier task. Participants perceived those who had done well as having greater authority; hence it was acceptable for them to keep more for themselves. More generally, Rai and Fiske's review accords well with the framework of deontological coherence, which proposes that moral rules arise out of interlocking rights and duties. Their review highlights the complexity of deontological moral rules: it appears unlikely that any blanket statement could constitute a suitable rule for fairness (e.g., "goods should be distributed equally") unless it also contained provisions for its exceptions (e.g., "except when one person has done more work"). However, appropriate prescriptions for conduct in a given situation can be derived from consideration of the rights of each agent.

Deontological Coherence as Constraint Satisfaction

Decision problems that raise moral issues are often challenging. Although we have argued that moral judgments involve rule-based reasoning according to deontological principles, this is clearly not all there is to moral judgment. Resolving moral decision problems may depend on multiple factors, including (a) one or more moral constraints (both deontological and consequential), (b) causal knowledge of how various possible actions would lead to various possible outcomes, (c) mental state attributions, and (d) emotional responses elicited by consideration of options. These and other relevant factors may often conflict with one another. Fortunately, people have available a domain-general decision mechanism that can take a hard problem of this sort and make it easy (or at least easier): *coherence-based reasoning*. The basic idea is straightforward: in the course of reaching a decision, a reasoner will shift their interrelated attitudes and beliefs so that they cohere with the emerging decision. Importantly, our emphasis is on how coherence emerges during reasoning (often transiently). We do not claim that people always hold coherent prior beliefs; rather, local coherence can emerge during reasoning even if the person holds beliefs that are globally incoherent. A similar distinction between coherence in prior beliefs and coherence that emerges during reasoning has been made in discussions of conceptual change in children (e.g., Chi & Ohlsson, 2005; DiSessa, 1982).

Bidirectional Inferences

Coherence-based reasoning is a domain-general mechanism that applies to moral reasoning as a special case. Its operation has been observed in a variety of complex decisions in which moral issues arise—legal cases, attitudes to war, attributions of blame and responsibility. However, its potential to provide a general framework for moral reasoning has not been fully realized (but see Clark, Chen, & Ditto, 2015; Uhlmann et al., 2009). A stumbling block, we suspect, is that coherence-based reasoning rests on a principle that directly contradicts a near-universal assumption within both philosophical and psychological work on moral judgment. Deontologists (including moderates) and utilitarians alike have generally assumed that when a moral issue arises, people approach it with specific predetermined values—beliefs about their rights and duties, and/or about utilities (both their own and those of others) associated with possible outcomes. People's values may change over time, but not (it is assumed) within the seconds or minutes spent pondering a typical moral decision. Despite the sharp conflicts about the normative and descriptive procedures for making such decisions, the general principle that inferences are *unidirectional* has been almost universally assumed—based on their entering values and beliefs (taken as fixed), the person makes an appropriate calculation (e.g., of whether an action would violate a duty, or promote utility) that leads to a decision about what ought to be done.

In stark contrast, coherence-based reasoning is *bidirectional*—rather than values and beliefs being fixed over the course of the reasoning episode, they may change to increase their coherence with the emerging decision. The outcome of decision making is not simply the choice of an option, but rather a restructuring of the entire package of values, attitudes, beliefs and emotions that relate to the selected option.

Operation of Coherence Models

A brief history. The history and scope of coherence theories is a broad topic, and we will not attempt a full discussion here (for more thorough reviews of work in psychology see Simon & Holyoak, 2002; Simon, Stenstrom, & Read, 2015; for applications to law see Simon, 2004, 2012). Early Gestalt psychologists promoted the view that attitudes and beliefs (including immediate perception) are governed by bidirectional interactions among their constituent elements, which act to promote a form of cognitive consistency (e.g., Heider, 1946, 1958; Lewin, 1938; Wertheimer, 1923/1967). Most famously, Festinger's (1957) cognitive dissonance theory postulated that attitudes and beliefs are altered retroactively to maintain consistency with one's actions.

After a couple of decades of relative neglect, the basic ideas underlying coherence theories reemerged in the principles of parallel distributed processing, embodied in neural-network models (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). A standard type of neural network consists of a set of nodes interconnected by weighted links. In the general case, links (either excitatory or inhibitory) can pass signals in either direction (i.e., feeding forward from one node to another, or back in the reverse direction). Beginning from a start state of initial activation levels on each node, activation is passed from each node to those to which it is directly connected, with the connecting links acting as multipliers. Each unit updates its activation based on its summed inputs. This process is repeated cyclically. The updating algorithm performs *constraint satisfaction*, so that the entire network eventually settles into a stable state in which the solution consists of a set of active nodes that "agree" with one another, whereas those nodes that "disagree" with the solution are deactivated.

Inspired by the success of constraint satisfaction models as an account of important perceptual phenomena (e.g., the top-down impact of context on word recognition), Holyoak and Thagard (1989) began to develop constraint-satisfaction models for tasks involving higher-level reasoning, notably analogy. Thagard (1989) created a model of "explanatory coherence," ECHO, to address the problem of evaluating the relative merits of competing explanations. Spellman, Ullman, and Holyoak (1993) used a simpler variant of ECHO, called Co3 ("*Co*herence Model of *Co*gnitive *Co*nsistency"), to explain how attitudes and beliefs can interact to jointly determine a decision about an important political issue

(whether to support a war). Thagard and colleagues continued to extend the scope of coherence-based processing, developing similar constraint-network models to explain processes such as impression formation (Kunda & Thagard, 1996; Thagard, 2000) and the interplay between cognition and emotion (Thagard, 2006; Thagard & Nerb, 2002; see also Dalege et al., 2015).

Although coherence-based reasoning has generally been formulated in terms of neural-network models, many of its basic phenomena can also be captured by Bayesian formulations (Jern, Chang, & Kemp, 2014). As noted by Simon (2004, p. 518), coherence-based reasoning resembles the philosophical view that moral principles and moral judgments can be gradually reconciled by an iterative process of *reflective equilibrium* (Rawls, 1971).

Shifting moral attitudes toward war. As an illustration of how coherence-based reasoning may relate to fluidity in moral values, we will describe an application of the Co3 model (Spellman et al., 1993) to explain shifting attitudes toward a war. These investigators applied their model (schematized in Figure 1) to a shift in public opinion that occurred in the United States over a 2-week period at the beginning of the first Gulf War in 1991, when George H. W. Bush initiated a U.S.-led counterattack against the Iraqi forces of Saddam Hussein, which had invaded and occupied Kuwait (Spellman & Holyoak, 1992). A survey was administered to a group of American college students on two occasions: during the first two days after the counterattack began, and two weeks later. Factor analysis was applied to extract six factors (the same factors for both survey administrations) related to people's war-related attitudes.

These factors appear as nodes in Figure 1. One of these, GEN, reflected direct support for the U.S. entering a war, and was treated as the outcome variable. The other factors measured beliefs and attitudes that predicted support for the war. For three of these, positive values were associated with opposition to the war. PAC was related to pacifism (defined by agreement with statements such as, "War is never justified"). ISO was a measure of isolationism ("The U.S. should not get involved in regional politics"). TER assessed degree of belief that U.S. military action might

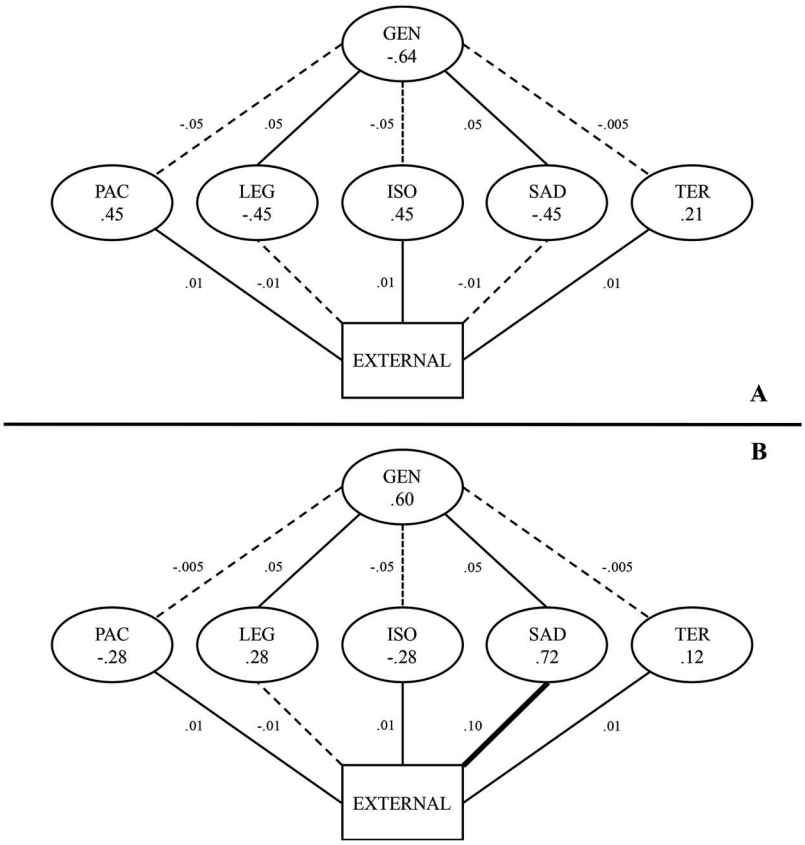

Figure 1. Network showing Co3 units and links representing the relations between attitudes about six constructs related to the Persian Gulf War. Panel 1A shows original weights on links and activations of units after 100 cycles. Panel 1B shows activations of units after the weight on the link from EXTERNAL to SAD is changed from −.01 to +.10 and network has run for 1000 cycles. Solid lines are excitatory links; dotted lines are inhibitory links; thickness of the lines indicates the weight on the link. GEN = general support, PAC = pacifism, LEG = legitimacy, ISO = isolationism, SAD = Saddam Hussein, TER = terrorism. EXTERNAL represents external influences on each of the constructs. From "A Coherence Model of Cognitive Consistency," by B. A. Spellman, J. B. Ullman and K. J. Holyoak, 1993, *Journal of Social Issues, 49,* p. 158. Copyright 1993 by Wiley Blackwell. Reprinted by permission.

provoke terrorism at home. For the other two factors, positive values were associated with support for U.S. military action. LEG assessed the moral legitimacy of the war ("The U.S. is acting to stop aggression and defend principals of international law and sovereignty"). SAD was a measure of the degree to which Saddam Hussein posed a future danger.

The top and bottom graphs in Figure 1 show the network and weights assumed for a Co3 simulation of initial and posttest attitudes, respectively. The weight structure of the network is virtually identical for the two time periods, with positive values on links from the LEG and SAD nodes to GEN, and negative values on links from PAC, ISO, and TER to GEN. The five predictor factors are not directly connected to one another, but are linked indirectly through the GEN node. The EXTERNAL node is a computational device (activation value clamped to 1) that feeds constant activation to each of the five predictor factors (reflecting the assumption that all factors receive constant attention during the decision process). Activations of all units (except EXTERNAL) are initialized at 0. The activation of a unit j may range from -1 to 1 and is computed on each cycle using a standard updating equation:

$$a_j(t+1) = a_j(t)(1-\theta) + \begin{cases} net_j(\max - a_j(t)) & \text{if } net_j > 0 \\ net_j(a_j(t) - \min) & \text{if } net_j < 0 \end{cases}$$

where $a_j(t)$ is the activation of unit j on cycle t, θ is a decay parameter (set at .05) that decrements the activation of each unit on every cycle, *max* and *min* are, respectively, maximum (1) and minimum (-1) activations, and net_j is the net input to a unit, defined as

$$net_j = \Sigma_i w_{ij} a_i(t).$$

The top part of Figure 1 shows the asymptotic state of the network after 100 cycles. The activation values of the nodes qualitatively capture the results of the first survey administered by Spellman et al. (1993). Overall, initial opinion was mildly opposed to U.S. intervention (GEN is negative), with positive activations for PAC, ISO and TER, and negative activation for LEG and for SAD (i.e., little concern that Saddam is a future threat). A marked opinion shift took place by the time of the second survey, two weeks later, with sentiment switching to support of intervention. The most salient news events that might have triggered this shift involved actions taken by Saddam Hussein (e.g., mistreating captured American prisoners of war, firing missiles at Israel). To model the impact of this recent news, a single weight was changed in the Co3 network: the link from EXTERNAL to SAD was changed from weakly negative (-.01) to moderately positive (.10). The network was then allowed to run an additional 900 cycles, achieving a new asymptotic state, as shown in the bottom of Figure 1.

The impact of the single weight change on the new state of the network was dramatic. Not only did the activation of SAD become highly positive (the direct impact of the change), but so did the activation of GEN. Moreover, other predictor factors also changed: PAC and ISO became negative, whereas LEG became positive. All of these changes in the network's asymptotic activation mirrored changes in people's attitudes as assessed by the two surveys. Of particular importance with respect to moral reasoning, coherence-based decision making yielded marked changes in attitudes on what would seem to be key moral issues, notably degree of pacifism (PAC) and perceived legitimacy of military action (LEG). Thus, new information indicating that Saddam Hussein was a "bad guy" not only triggered a reversal in support for the war, but also led to major changes in moral factors relevant to this decision. Attitudes change, but coherence is maintained.

Coherence phenomena. The Spellman et al. (1993) study was naturalistic; however, a considerable body of experimental work has provided a more detailed picture of the empirical phenomena associated with coherence-based decision making. These phenomena (see Table 1), which collectively serve as the signature of this reasoning mechanism, are consistent with a yet more extensive body of research in the field of judgment and decision making showing that people's preferences are not simply passively stored in memory, but rather are actively *constructed* (see collection edited by Lichtenstein & Slovic, 2006). Coherence shifts have been extensively studied in the context of economic decisions that do not involve obvious moral concerns (e.g., Ariely, Loewenstein, & Prelec, 2003; Chael, 2015; Russo, 2014; Russo et al., 2008; Russo, Meloy, & Medvec, 1998; Simon et al., 2008; Simon, Krawczyk, & Holyoak, 2004). Here we will focus on studies of simulated legal decision making, using problems that *do* involve moral concerns.

Holyoak and Simon (1999) showed that in the context of a legal trial, students changed their minds about multiple factors, including moral values (their belief in freedom of speech), so as to achieve consistency with their final decision. Furthermore, these "coherence shifts" occurred in the process of reaching that decision. The investigators had college students play the role of judges to decide the verdict in a realistic legal case. The case of "Quest v. Smith" focused on the then emerging technology of the Internet. Quest, an Internet company, has gone bankrupt and is now suing Smith, a former investor who posted negative statements about the company's management on an electronic bulletin board, alleging his message caused the company's collapse. Six conflicting arguments were made by each side, involving issues such as whether the message was true, whether it was indeed the cause of the bankruptcy, and whether Smith's motive was malicious or altruistic (intended to warn other investors). One issue of general moral significance involved free speech (whether Smith had the right to state his opinions openly, or whether "harmful" speech should be

Table 1
Major Phenomena Accompanying Coherence-Based Reasoning

1. Sharply divided decisions are accompanied by high confidence for each individual decision maker.
2. Attitudes and beliefs that are at first only loosely correlated become strongly correlated in the process of reaching a decision.
3. A coherence shift largely takes place prior to commitment to a decision.
4. A coherence shift can be triggered by any task set that encourages attention and comprehension (even if no decision is required).
5. Manipulating one decision-relevant factor changes not only the decision, but also evaluations of other factors.
6. People are largely unaware that they have shifted their views.
7. Constraint-based reasoning involves motives and emotions as well as cognitive beliefs.
8. A coherence shift can "prime" the choice of option in a different problem presented shortly afterwards.
9. However, shifts in attitudes tends to be transient.

controlled). The competing arguments were closely balanced so that the case was highly ambiguous.

A series of assessments were used to track the course of participants' attitudes and beliefs related to the case. All participants first completed a pretest before hearing about the legal case. In the guise of an opinion survey, people were asked to rate their degree of agreement or disagreement for issues (presented individually) that would later bear on the case. Next, the legal case was presented, and participants were asked to be "fair and just" in deciding it. Half the participants received an "interim" assessment of their "leaning" regarding the case (expecting to receive additional information later). The same issues as had been rated on the pretest were rated again (in a new random order), now set in the context of the legal case. Finally, all participants were asked to give their final verdict (without receiving any additional information) and to rate their confidence in it, after which they completed a final posttest.

Figure 2 shows the distribution of verdicts, with degrees of confidence, in this initial experiment. The pattern illustrates a basic signature of coherence-based reasoning. Though the case was indeed highly ambiguous, as evidenced by a near-even split of verdicts favoring Quest versus Smith, the great majority of individual participants reported great confidence in their verdict. Figure 3 shows the accompanying shift in attitudes and beliefs, aggregated across all arguments involved in the case, broken down according to the final verdict that was reached. Ratings for each argument were coded so that a higher score represents greater favorability to Quest's position, and then averaged to create a "Q"-score. On the pretest, Q-scores did not reliably differ as a function of the eventual verdict. By the interim test a strong and reliable difference had emerged, which became slightly larger (though not reliably so) on the posttest. This coherence shift took place for each of the individual issues related to the case. This was true even for attitudes toward freedom of speech, which might have been considered a deeply entrenched moral value, as those who decided in favor of Quest became less supportive of free speech in the posttest.

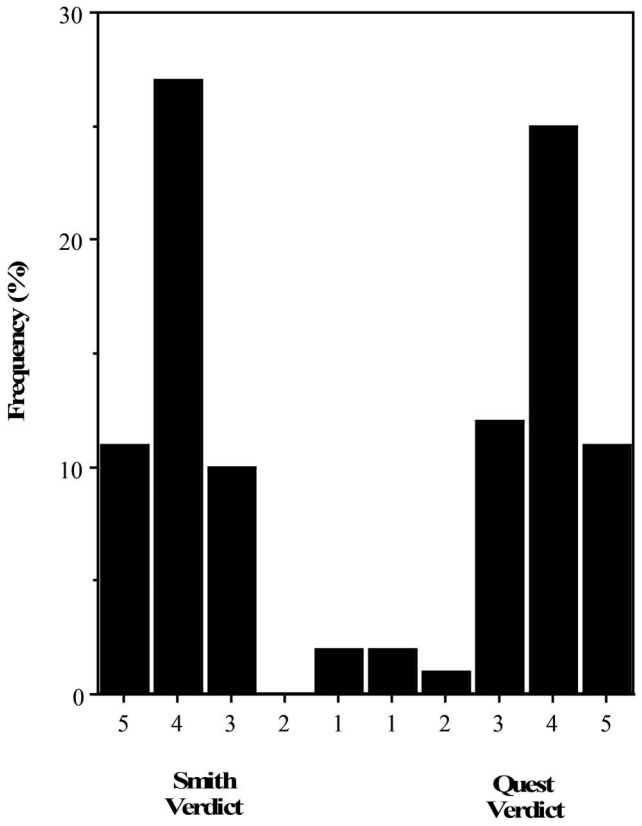

Figure 2. Distribution of confidence ratings for final verdict in "Quest v. Smith" case (Holyoak & Simon, 1999, Experiment 1). Ratings are on a 5-point scale, ranging from 1 (*minimal confidence*) to 5 (*maximal confidence*). Scale is reversed for the two verdicts. From "Bidirectional Reasoning in Decision Making by Constraint Satisfaction," by K. J. Holyoak and D. Simon, 1999, *Journal of Experimental Psychology: General, 128,* p. 6. Copyright 1999 by the American Psychological Association. Reprinted by permission.

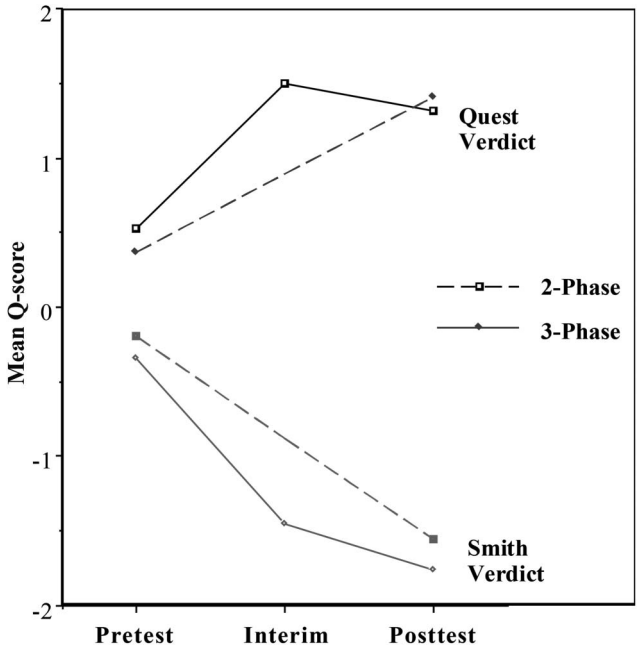

Figure 3. Shifts in Q-scores (favorability to Quest's position) across tests as a function of eventual verdict for "Quest v. Smith" case (Holyoak & Simon, 1999, Experiment 1). From "Bidirectional Reasoning in Decision Making by Constraint Satisfaction," by K. J. Holyoak and D. Simon, 1999, *Journal of Experimental Psychology: General, 128,* p. 7. Copyright 1999 by the American Psychological Association. Reprinted by permission.

In addition to the dramatic shift in means, people also showed a shift in the *correlations* among their ratings of the individual issues. On the pretest these correlations tended to be near zero, but on the posttest *all* correlations were highly reliable. For example, a person who decided in favor of Quest now tended to give correlated ratings for free speech (less positive) and Smith's motive (more malicious), even though their pretest ratings on these issues had not been not correlated. Coherence, initially lacking, was created in the very process of decision making.

Importantly, the bulk of the coherence shift had already taken place by the interim test, prior to the participant committing to a firm decision. This finding supports the hypothesis that (contrary

to the view of Festinger, 1957) coherence-based reasoning is prior to and determinant of the eventual decision, rather than some sort of post hoc rationalization. Indeed, subsequent work showed that a coherence shift is obtained even if no decision is ever called for. Any task that elicits comprehension of the scenario (e.g., expecting to communicate the information to others, or simply memorizing the case) is sufficient to generate a coherence shift (Simon et al., 2001; see also Kruglanski & Shteynberg, 2012).

Table 1 summarizes some of the key empirical phenomena associated with coherence shifts, based on the large body of relevant studies. Manipulating one important factor influencing a decision (e.g., motive) can alter not only the decision, but also evaluations of all the other relevant factors. Further, it seems that people have little metacognitive awareness of their own coherence shifts. When later asked to recall their entering opinions (i.e., their own pretest ratings), they tend to report a blend of their entering and final opinions. An attitude altered by coherence-based reasoning during one decision can "prime" a congruent decision on a subsequent problem that is given immediately afterward. After longer delays people's attitudes tend to regress toward their entering position. For a moral issue, this transience implies that one might confidently base a decision in part on a moral stance (e.g., toward free speech) that could differ in the context of some other decision problem encountered in a different time and context.

Simon et al. (2015; see also Simon, Snow, & Read, 2004) performed an extensive set of experiments investigating decision making with other complex legal cases. Their findings showed that coherence shifts can involve not only "cold" cognition, but also "hot" emotion (Thagard, 2006; Phenomenon 7 in Table 1). A large body of research has established that emotions are sensitive to and rely on cognitive appraisals of situations (e.g., Ellsworth, 2013; Ellsworth & Scherer, 2003). Perceiving an actor as the cause of harm triggers anger (perhaps guilt or shame, if the culprit is oneself); perceiving someone as a victim triggers sympathy or compassion. Evoking anger tends to increase attributions of blame (e.g., Goldberg, Lerner, & Tetlock, 1999); evoking sympathy inhibits punitive tendencies (e.g., Rudolph et al., 2004).

Simon et al. (2015) performed several experiments using a "cheating" scenario, in which participants had to decide whether a student named Debbie had committed academic misconduct. A pretest and matched posttest were administered to assess participants' attitudes on a variety of factors relevant to the decision. Figure 4 shows a coherence network (using essentially the same constraint satisfaction algorithm as that employed in the Co3 simulation depicted in Figure 1) based on these factors. The decision nodes are labeled "Cheated" and "Did Not Cheat." In addition to several factual arguments for Debbie's guilt or innocence (e.g., "InnFact1"), other factors measured emotional reactions to Debbie (e.g., "Anger" and "Sympathy") and overall emotional valence (both negative and positive).

Simon et al. (2015) found that the emotional factors strongly interacted with the more cognitive factors within the constraint

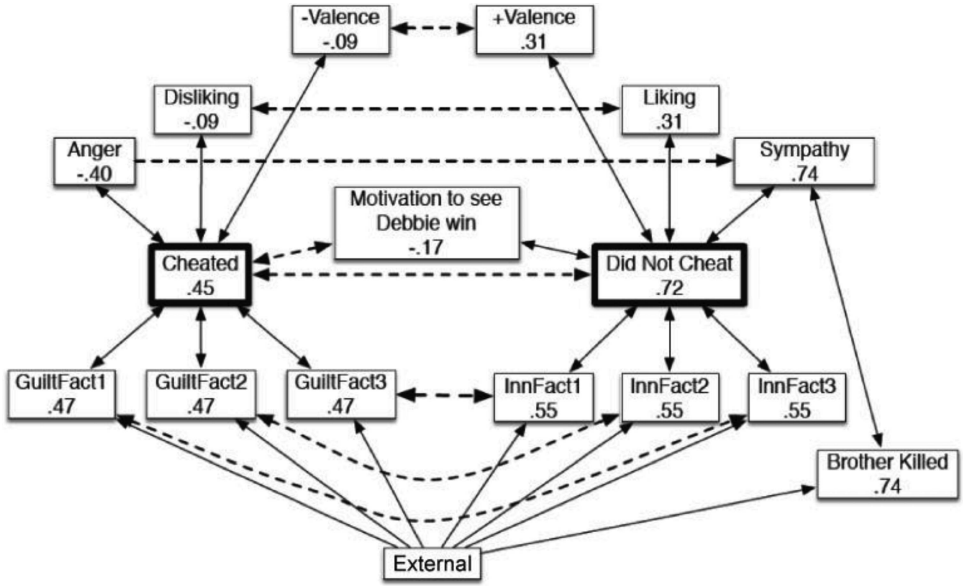

Figure 4. Network of a coherence simulation for "sympathy condition" with the "Cheating" case (Simon et al., 2015, Study 3). Each node represents a different concept in participants' model of the task and corresponds to a measured dependent variable, with the exception of the External node and the Brother Killed node (which corresponds to the manipulation). The External node provides starting activations for the facts and the manipulation. The number in the node represents its activation once the network has "settled" or reached maximal coherence. Excitatory links are represented by solid lines and inhibitory links by dashed lines. All excitatory weights are .10 and all inhibitory links are −.18. Double-headed arrows represent bidirectional connections with bidirectional spread of activations; single-headed arrows represent spread of activation only in the direction of the arrowhead. From "The Coherence Effect: Blending Hot and Cold Cognitions," by D. Simon, D. M. Stenstrom and S. J. Read, 2015, *Journal of Personality and Social Psychology, 109,* p. 384. Copyright 1999 by the American Psychological Association. Reprinted by permission.

network. Furthermore, these influences were bidirectional. In separate experiments the investigators showed that manipulating Debbie's perceived guilt altered anger and sympathy responses, and that manipulating anger or sympathy impacted the verdict and other cognitive factors. The network in Figure 4 represents a condition from the latter experiment, where sympathy for Debbie has been increased by the information that her brother was recently killed. This information had no logical bearing on whether or not she had cheated on an exam, but it served to increase sympathy for Debbie (and reduce anger), leading to an increase in "innocent" verdicts and reduced agreement with the cognitive arguments supporting her guilt. This finding illustrates a general principle of deontological coherence: cognitive and emotional factors interact within a single integrated decision process. In contrast to some dual-process theories of moral judgment (Greene et al., 2001), deontological coherence denies that "hot" emotion and "cold" cognition necessarily map onto distinct decision processes. Rather, emotion and cognition both provide inputs to a unified coherence-based decision process.

As a further illustration, we will sketch how the famous "Heinz" dilemma (Kohlberg, 1981) could be represented in terms of deontological coherence. The essence of the situation is that a poor man's wife is dying of a rare disease. A druggist has a drug that could cure her, but is demanding an exorbitant sum that impoverished Heinz cannot pay. His dilemma is whether or not to steal the drug to save his wife's life. Like Kohlberg, we are not concerned here with the actual decision, but rather with the factors that enter into the decision process.

Figure 5 illustrates the basic structure of the constraint network for the Heinz problem, couched in terms of the core concepts of deontological coherence. (Weight values, which would determine the eventual decision, are left unspecified.) Schemas for permissions and obligations will apply to identify decision-relevant rights and duties; causal knowledge will help to infer intentions and expected consequences. The essence of the dilemma is the conflict between two incompatible *duties*: obedience to the law (hence not stealing) versus loyalty to one's spouse (and therefore stealing the medicine). Figure 5 includes nodes for the two conflicting duties, each respectively supported by their *ground*. Grounds may themselves vary in terms of the support they provide to their associated rules (consistent with a graded interpretation of Turiel's, 1983, categorical distinction between moral rules and conventional norms). The EXTERNAL node at the left of the figure operates like the similar nodes shown in Figures 1 and 4, feeding excitatory activation to each ground in proportion to background attitudes about the importance of each set of values.

Corresponding to each duty is an *action option* that would be consistent with it (excitatory connection) and another than would violate it (inhibitory connection). For example, the duty of obeying the law supports not stealing (and inhibits stealing). The action options in turn are connected to their apparent consequences (stealing means jail for Heinz and medicine for his wife; not stealing means no medicine and hence death for Heinz' wife). The "default" outcome stated in the story (the wife dies for lack of the drug) is supported by the EXTERNAL node. Finally, other nodes represent various emotions associated with the possible actions and outcomes: Heinz's fear of jail and/or his wife's death, sadness should she die, and possible guilt triggered by whichever duty is violated by his decision.

Depending on the critical details (weights on links), the network in Figure 5 could potentially settle into various asymptotic states that would favor one or the other action option. To provide a detailed theory, the framework of deontological coherence would need to be augmented by an account of how morally relevant factors and weights are acquired (as we discuss further below). But the framework does explain how moral judgments interact with shifts in attitudes, beliefs and emotions. Importantly, there is no guarantee that any solution will entirely resolve the conflicts embodied in the constraint network. Heinz may decide that on balance he needs to steal the medicine, but still feel lingering guilt at violating the law (in addition to fearing punishment). Deontological coherence thus supports the intuition that the decision maker may decide that the "morally right" action in a specific context requires doing a moral wrong (Nagel, 1972). And indeed, people typically report feelings of anger and guilt even when they approve of taking action in sacrificial dilemmas (Horne & Powell, 2013, 2016), and view agents who make sacrificial choices as having deficiencies in moral character, even when they view their

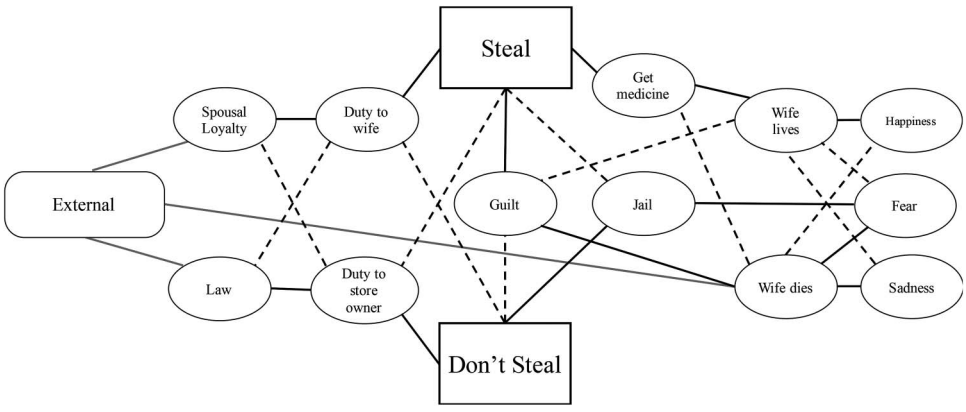

Figure 5. A coherence network for the Heinz dilemma (Kohlberg, 1981). Duties and their grounds, consequences, and emotions are integrated by constraint-based reasoning to evaluate options for action. Excitatory connections are indicated by solid lines, inhibitory connections by dashed lines.

actions as morally correct (Uhlmann, Zhu, & Tannenbaum, 2013). More obviously, in some situations a decision maker may elect to do a morally wrong action (favored perhaps by self-interest) despite being aware that it is immoral. Coherence-based reasoning does what it can to make everything "fit together," but moral values, although flexible, are not infinitely elastic. In general (for better or worse), moral values that are less firm are more flexible (Skitka, Bauman, & Sargis, 2005). Indeed, manipulations that enhance the credibility of moral relativism (vs. realism) have been shown to increase cheating behavior (Rai & Holyoak, 2013; see also Young & Durwin, 2013), presumably because they reduce people's certainty about their own moral values.

Evidence for Coherence Shifts in Moral Judgment

The framework of deontological coherence assumes that many types of factors play a role in determining people's moral judgments. Moral rules, anticipated consequences, evaluations of agents' intentions and their ability to act freely (e.g., Darley & Shultz, 1990), together with emotional responses (e.g., Greene & Haidt, 2002; Haidt, 2001; Wheatley & Haidt, 2005), all feed into moral judgments. Deontological coherence implies that all such factors enter into bidirectional interactions with people's moral judgments.

Moral judgments affect endorsement of moral rules. In cases where an agent's eventual moral judgment countermands a relevant moral rule, deontological coherence predicts that the weight ascribed to that rule should be reduced. Horne, Powell, and Hummel (2015) examined the effect of participants' moral judgments on their endorsement of general moral principles. Participants in the experimental condition were asked to make a judgment about the "transplant" dilemma (Thomson, 1985): whether it would be acceptable for a doctor to sacrifice one patient in order to harvest his organs and save the lives of five other patients awaiting transplants. Participants almost universally agreed that this action was unacceptable. They subsequently were asked to rate their agreement with a utilitarian principle: "In the context of life or death situations, always take whatever means necessary to save the most lives." Whereas participants in a control condition tended to strongly agree with the utilitarian principle, agreement was significantly reduced for those who had just considered the transplant dilemma. Strikingly, this effect was not only observed when participants were asked the question immediately after making their moral judgments, but also after a delay of six hours. It appears that encountering a clear counterexample to a putative moral principle may lead to a more long-lasting change in reflective equilibrium (Rawls, 1971), effectively down-weighting the principle.

Coherence-based belief shifts may explain why participants' judgments of moral situations are often affected by their prior moral judgments. For example, making judgments about the footbridge dilemma affects participants' judgments in the standard trolley dilemma (Liao et al., 2012; Lombrozo, 2009; Schwitzgebel & Cushman, 2012; Sinnott-Armstrong, 2008; Wiegmann et al., 2008; Wiegmann & Okan, 2012). Like the transplant dilemma (though perhaps to a lesser degree), considering the unpalatable footbridge dilemma seems to undermine utilitarian beliefs, reducing the acceptability of taking a sacrificial action (Horne, Powell, & Spino, 2013).

Moral judgments affect evaluations of consequences. It seems natural that the consequences of an action should affect people's moral judgments about that action. More interestingly, people's moral judgments have also been shown to affect their evaluations of the potential or actual consequences of actions. Liu and Ditto (2013) asked participants to consider short deontological arguments favoring or opposing capital punishment, thereby altering their attitudes about the morality of the death penalty. Following this shift in moral attitudes, participants demonstrated a parallel shift in their beliefs about the efficacy of capital punishment as a deterrent for violent crime, and in the overall positive and negative societal benefits of capital punishment. Similarly, people perceive an inverse correlation between benefits and risk. For example, a manipulation that increased the perceived benefits of nuclear power reduced its perceived risk (Finucane et al., 2000).

In a similar vein, Ames and Fiske (2013, 2015) found that agents' *intentions* when committing harmful acts affected people's evaluations of the *consequences* of those actions. In one study, participants were asked to quickly tally a series of financial damages (expressed in dollar amounts) that were produced by an agent's intentional or else unintentional actions. When the damages were produced unintentionally, participants were quite accurate. However, when they were told to that the harms were committed intentionally, they greatly overestimated the total sum of the damages. These researchers concluded that agents' intentional actions warranted greater blame for the outcomes of those actions, which in turn shifted people's estimates of the damages caused by these actions. As predicted under the deontological coherence framework, these data suggest that the links between evaluations of outcomes or consequences and moral judgments are bidirectional.

Moral judgments affect evaluations of secondary moral factors. Moral judgments have been shown to shift evaluations about a host of other, secondary factors relevant to the moral judgments. For example, the "side-effect effect" is the tendency for moral judgments to influence people's interpretations of an agent's intentions (Knobe, 2003). An essential aspect of the phenomenon is an asymmetry between harmful and helpful actions: people ascribe intent to a corporate executive who institutes a program that harms the environment, but not to one who institutes a program that helps it, even if both deny having considered environmental effects in their decision-making process. Consistent with the deontological aspects of our framework, Leslie, Knobe, and Cohen (2006; also Knobe, 2010) argued this asymmetry arises because people perceive the executive to be under a stronger obligation to avoid causing harm than to actively cause good. Researchers using similar paradigms have found that people are more likely to say that agents whose actions led to harm acted with foreknowledge (Beebe & Buckwalter, 2010) and freely chose their actions (Phillips & Knobe, 2009; Young & Phillips, 2011).

Just as they affect participants' evaluations of the mental states of agents, moral judgments also affect evaluations of cause-effect relationships. Judging that an action was wrong makes people more likely to judge that it played a causal role in bringing about bad outcomes (Alicke, 2000; Cushman, Knobe, & Sinnott-Armstrong, 2008; Hitchcock & Knobe, 2009; Knobe & Fraser, 2008). Similarly, making a moral judgment influences participants' counterfactual predictions regarding what would have occurred if the action had not been taken (McCloy & Byrne, 2000;

N'gbala & Branscombe, 1997; Roese, 1997). In addition, Chituc et al. (2016) recently showed that people's moral judgments influence their assessments of agents' obligations—sometimes leading them to claim that agents "ought" to have done something that was not possible for them to do.

Similar coherence shifts have often been interpreted as instances of "motivated" reasoning (Kunda, 1990; for a review see Ditto et al., 2009). Motivated reasoning occurs when people's self-interested biases potentially undermine their moral judgments. (Recall that the coherence network for the Heinz dilemma included such nonmoral factors as "fear of jail.") Motivated reasoning is simply a variety of coherence-based reasoning, though one with particularly important implications for the evaluation of people's moral judgments.

Altogether, these findings are well-explained as coherence shifts produced by bidirectional links between moral judgments and intentions, mental state attributions of knowledge, judgments of free action, and causal attributions.

Coherence shifts and moral absolutism (or stubbornness). In some situations, moral values seem to resist trade-offs of the sort expected under moderate forms of deontology (see "Moderate Deontology" above). Trémolière and Bonnefon (2012) found that although most people were willing to sacrifice one person when doing so would save a much greater number of people, a significant proportion of participants refused to sacrifice even when doing so would save 5000 people (see also Greene et al., 2008; Nichols & Mallon, 2006). Research on "protected" or "sacred" values" (Fiske & Tetlock, 1997; Tetlock, 2002) suggests that some moral values (e.g., preventing extinction of species) are largely insensitive to consequentialist trade-offs (Baron & Ritov, 2004, 2009). Baron and Spranca (1997) reported that protected values display "quantity insensitivity"—people seemed indifferent to the consequences produced by actions violating these values (but see Bartels & Medin, 2007, for evidence of some sensitivity to trade-offs).

Just as probabilities may take on extreme values (0 or 1) indicative of certainty, deontological coherence allows the possibility of extremely high weights on some duties. Experiments by Baron and Leshner (2000) provide a test of two important predictions of this account. Their study examined judgments requiring trade-offs between two protected values (identified for each participant on a pretest). According to deontological coherence, this is simply a case of conflict between two highly weighted constraints. When participants were asked to choose between policies that would cut government funding for protection of one or another sacred value, a majority made a choice, indicating an ability to resolve conflicts between protected values. In addition, deontological coherence implies that the weights placed on protected values, though high, may nonetheless be malleable. Consistent with this prediction, Baron and Leshner found that people were less likely to endorse values as protected when they were asked to think of an instance in which the protected value could be sacrificed. Although protected values may have very high weights, they are not immune from coherence shifts.

Interplay between emotional responses and moral judgments. As we noted earlier, moral judgments can influence emotional evaluations of actors, and conversely, emotional evaluations of actors can influence moral judgments about them (Simon et al., 2015). Other studies have shown that negative emotional responses can lead to negative moral appraisals (e.g., Greene et al., 2001; Haidt, 2001), and that emotions may play some causal role in determining moral judgments (e.g., Wheatley & Haidt, 2005). The framework of deontological coherence implies that emotion (like other types of morally relevant factors, such as motive; e.g., Rai & Holyoak, 2014) will enter into bidirectional interactions with moral judgments. More generally, research indicates that cognition and emotion are closely coupled at both the behavioral and the neural level, rather than operating as independent systems (e.g., Ellsworth, 2013; Pessoa & Pereira, 2013).

Further Theoretical Implications

Dual-Process Accounts of Moral Judgment

In recent decades, numerous models of memory, learning, reasoning, and decision making have invoked a binary distinction between two types of cognitive processes. Some of the more notable early examples (none dealing with moral judgments) were attributable to Reber (1993), Evans and Over (1996), Stanovich (1999), and Kahneman and Frederick (2002). Kahneman's (2011) distinction between "thinking, fast and slow" (in his book of that title) further popularized the dual-process approach. Although the two types of models have received many different names, the currently most popular are the generic "System 1" versus "System 2." Crudely put, System-1 processes tend to be fast, easy and intuitive, whereas System-2 processes tend to be slow, difficult, and reflective.

As described earlier, Greene et al. (2001; Greene, 2008) were the first to propose a dual-process model of moral judgment, which has been extremely influential. At its most basic, Greene's (2008) dual-process theory makes three claims: (a) moral judgments are based on two cleanly separable types of psychological and neural processes; (b) effortful deliberative processes map onto utilitarianism, which is rational; and (c) automatic affective processes map onto deontology, which reflects an incoherent set of emotion-driven biases. For instance, Shenhav and Greene (2014) provided evidence that moral judgments in "personal" moral dilemmas (e.g., the footbridge version) are mediated by activation in the amygdala, which is subsequently integrated with utility evaluations supported by the VMPFC to produce "all things considered" moral judgments.

It should be clear that our framework of deontological coherence denies each of these claims. (a) In the deontological coherence framework, the inputs to moral judgment processes are intertwined rather than dissociated. (b) Utilitarianism is not interpreted as a normative standard for morality; rather, both utilitarianism and deontology provide concepts relevant to a descriptive theory of moral judgment. (c) Deontological moral judgments and beliefs are the products of systematic rule-based reasoning, rather than simple emotion-driven processes.

Greene's dual-process theory differs from most domain-general dual-process accounts by focusing on the distinction between affective and cognitive processes and by positing associations between cognitive processes and particular moral judgments. In contrast, a number of other researchers have developed dual-process theories of moral judgments that invoke the more typical domain-general distinction between System 1 and System 2, but without linking these systems to particular patterns of moral judg-

ments (e.g., Baron, Gurcay, Moore, & Starcke, 2012; Kahane et al., 2012; Koop, 2013; Trémolière & Bonnefon, 2012). Other proposals have differed from Greene's dual-process theory in their characterizations of the cognitive processes tied to different types of judgments (Crockett, 2013; Cushman, 2013). It might be claimed (with a tinge of irony) that the field of moral psychology finds itself in the odd position of having achieved a fair degree of consensus that moral judgments are based on dual processes, without having established what those processes might be.

Against this backdrop, the framework of deontological coherence raises a radical alternative: that people have a single unified system for making moral judgments. This system is certainly complex, involving the integration of different types of information, which doubtless is retrieved or calculated at different rates (so that decisions may be influenced by cognitive load or speed pressure). Within this single decision process, representations and subprocesses can vary in ways that yield radically different moral judgments. However, coherence-based reasoning offers an account of how different cognitive representations can be affected by a unified decision-making process: a constraint satisfaction process operating via bidirectional connections to integrate moral values and their grounds, subjective utilities of consequences, emotions, and action options. The outcome takes the form of a de novo "gestalt" in which this entire configuration undergoes changes to maximize coherence with an emerging decision. Although affective reactions, utility calculations, deontological rules, and so forth may seem radically different and may indeed have quite different mental representations, coherence-based reasoning allows for all of these elements to interact with one another in what can meaningfully be considered a single decision-making process.

Of course, whether a set of cognitive operations constitutes a single process or multiple ones almost always depends on the theorist's conceptual vantage point (see Evans, 2009). The framework of deontological coherence seeks to move that vantage point a few steps back, describing the process of moral judgment from a relatively global perspective. In more detail it is undoubtedly possible to identify subprocesses that contribute to constraint-based reasoning about moral issues. Indeed, neuroscientific evidence suggests that a variety of different brain areas support various subprocesses involved in moral judgment making. The inferior frontal gyrus may be important for top-down inhibitory control (Cho et al., 2010), required to dampen the impact of salient but nonmoral factors, such as the agent's self-interest. A recent review of neuroscientific evidence (Greene, 2014) suggests that the amygdala tracks potential harms (including moral harms incurred by violation of a duty), and that the ventral striatum is sensitive to the magnitude of expected consequences for a group (much as it tracks subjective utilities of monetary gambles for an individual; Tom, Fox, Trepel, & Poldrack, 2007). Perhaps most importantly, the VMPFC has been identified as a crucial area involved in integration of information during moral judgment (Shenhav & Greene, 2014), raising the possibility (albeit speculatively) that this area is a potential substrate for the integrative mechanisms required to carry out constraint-based reasoning. The deontological coherence framework does not deny the existence or importance of these subprocesses, but views them and their interactions as elements of a larger process of coherence-based reasoning.

Differences Between Moral Judgments and Moral Justifications

Haidt's "social intuitionist" model (2001) also draws a binary distinction, not between cognitive processes engaging in moral judgments, but rather between moral judgment and moral justifications. According to Haidt, there is generally only one type of cognitive process involved in moral judgments: moral judgments are produced by intuitive processes, and much of the cognitive and reflective thinking that would seem to suggest the involvement of more reflective processes in moral judgment is actually being employed in the secondary task of justifying moral judgments. Moreover, though they often occur together, these two processes can and should be dissociated. To provide evidence for this claim, Haidt points to research showing that people sometimes struggle to explain the reasoning behind their judgments, a phenomenon termed "moral dumbfounding." For example, Haidt and Hersh (2001) recounted an interview with one participant who argued that cleaning a toilet with an American flag was morally wrong because, "It might clog up the drain." Haidt argues that people's confidence in their moral judgments, even when their justifications are clearly deficient, implies that moral judgments are made intuitively and independently from their attempts to justify these judgments.[3]

The framework of deontological coherence might be reconciled with Haidt's theory by viewing coherence-based reasoning and its resulting coherence shifts as an aspect of moral justification rather than judgment. This conceptualization would accord well with early coherence theories, notably Festinger's (1957) theory of cognitive dissonance, in which coherence-based reasoning was assumed to justify or rationalize behaviors after-the-fact. However, the more recent research on coherence-based reasoning reviewed earlier (e.g., Holyoak & Simon, 1999) suggests that coherence shifts often occur "online" as part of the decision process. In addition, there appears to be evidence for coherence shifts during moral judgment tasks that do not call for justifications. For example, it is unclear how Haidt's account would explain why agents' moral or nonmoral intentions influence assessments of damages during a task calling only for estimates of those damages (Ames & Fiske, 2013, 2015). It is also unclear why coherence shifts would affect subsequent moral judgments (as suggested by ordering effects on moral judgments; e.g., Horne et al., 2013; Horne et al., 2015; Liao et al., 2012; Lombrozo, 2009; Schwitzgebel & Cushman, 2012; Sinnott-Armstrong, 2008; Wiegmann et al., 2008; Wiegmann & Okan, 2012) if these coherence shifts are driven by justificatory processes separate from the intuitive judgment processes driving subsequent moral judgments. More generally, coherence-based reasoning suggests a blurring of the distinction between judgment and justification. Rather, these two goals are viewed as intertwined: people strive to increase or maintain coherence (an inherently justificatory concept) as they make judgments.

[3] Also note another potential interpretation of this anecdote in terms of coherence: the participant is so confident the action is wrong that a need arises to identify a coherent set of negative outcomes that it will cause.

Metacognitive Awareness and the Theory of Universal Moral Grammar

Moral dumbfounding reveals a lack of metacognitive awareness of the elements of moral judgment processes. This apparent dissociation has led some researchers to draw an analogy to competent language users' unawareness of many grammatical rules, proposing that moral judgments are accomplished by an innate cognitive module termed a Universal Moral Grammar (UMG; Mikhail, 2011). Like the framework of deontological coherence, UMG theorists emphasize the role of sophisticated deontological moral rules in moral judgment, and the role of experimental studies in decoding those rules. Beyond these general points of agreement, however, the two viewpoints could not be more different. UMG theorists argue that moral judgments depend on a highly specialized and modular grammar, whereas the deontological coherence framework posits that moral judgments are produced by coherence-based reasoning that integrates a variety of domain-general inputs and outputs.

In addition to other objections raised against UMG theories (e.g., Prinz, 2008), UMG is at odds with two major types of evidence we have reviewed in support of the deontological coherence framework. First, UMG has no way of accounting for coherence shifts in moral judgments, especially those involving nonmoral factors such as causality judgments. Bidirectional influences among many different moral and nonmoral factors is the last thing that should be expected if moral judgments are accomplished by a modular system. Second, people show a lack of metacognitive awareness in many types of nonmoral judgments, including problem solving and reasoning (e.g., Nisbett & Wilson, 1977), as well as coherence-based decision making in general (Holyoak & Simon, 1999). Thus although the metacognitive unawareness evidenced by moral dumbfounding is consistent with the modularity of moral judgment, it is by no means uniquely predicted by it.

Toward a Theory of Moral Judgment

Throughout this paper we have referred to deontological coherence as a "framework" for understanding moral judgment, rather than a "theory." Our choice of term is meant to acknowledge that the present proposal has yet to be developed into a well-specified and testable theory. Many important gaps are evident. In particular, the framework of deontological coherence has relatively little to say about what types of moral judgments people are likely to make and when. Instead, the framework is primarily focused on the reasoning process by which moral judgments are reached, and of important second-order consequences of that reasoning process (e.g., coherence shifts, translations between rights and duties). We have argued that different patterns of moral judgments within and across individuals may not be a product of strictly dichotomous processes underlying moral judgments, but rather may arise from the complex systems of moral beliefs and values those individuals hold, which can differ in many ways.

Table 2 provides a sample of possible variations (within normal and/or clinical populations) in moral representations and processes that would be expected to lead to different outcomes. The most obvious (and most relevant to cultural differences in moral judgments) is that people may have different moral values (and different grounds for them, and different assessments of moral harm caused by violating duties). One man may value a perceived duty to keep his immediate family members safe from physical harm, unconditionally. Another (in a different culture) may value a perceived duty to kill his sister because she was raped (and hence made impure, contaminating the entire family; Fiske & Rai, 2014). Both men may reach their respective judgments by seeking deontological coherence.

Future work should aim to integrate hypotheses concerning the content of moral factors into processing models of moral judgment. This may require research examining judgments of moral situations in which the elements are well-understood. Often, moral psychologists have studied laypeople's reactions to the same cases that have drawn the attention of ethicists, exemplified by the contrasting intuitions about the "standard trolley" and "footbridge" dilemmas. These dilemmas have interested ethicists, at least in part, because intuitions about these cases seemed somewhat mysterious. However, mystery is rarely a positive quality for the materials of psychological experiments. Integrating the content of moral beliefs into processing models of moral judgment will require research directed at moral situations in which it is possible to clearly identify the individual factors that drive judgments.

Human morality, and all the social and cognitive complexities that have attached themselves to it (e.g., religion, the supernatural, rituals, and law), likely depend upon what is special about human thinking: the ability to think about higher-order relations (Penn, Holyoak, & Povinelli, 2008). An adequate account of the role of relational reasoning in moral judgment will require models that are more sophisticated than those previously applied to constraint-based reasoning (e.g., Spellman et al., 1993). In particular, computational approaches to analogical and other forms of relational reasoning may be relevant. To think about moral (or nonmoral) higher-order relations requires neural mechanisms for coping with the formal complexity of variable binding (i.e., keeping track of "who does what to whom"; Halford, Wilson, & Phillips, 1998, 2010; Halford et al., 2014; Hummel & Holyoak, 1997, 2003), and for learning relations from nonrelational inputs (e.g., Doumas, Hummel, & Sandhofer, 2008; Lu, Chen, & Holyoak, 2012). We predict that whenever moral judgments require integrating multiple relations (e.g., in cases such as the Heinz dilemma in which conflicting rights and duties involving multiple people need to be considered), performance (e.g., as measured by solution time) will be affected by the number and complexity of relations involved. Moreover, individuals' ability to navigate this complexity will be correlated with measures of working memory and fluid intelli-

Table 2
Some Sources of Variation in Moral Judgments

1. Differences in moral values
2. Differences in grounds
3. Differences in schemas based on rights and duties
4. Differences in concern for types of moral patients
5. Different utility functions for consequences (including direct and/or moral harms)
6. Differences in emotional responses (e.g., lack of empathy/compassion)
7. Differences in cognitive capacity (e.g., ability to integrate multiple factors)
8. Differences in cognitive control (e.g., ability to inhibit certain types of factors, such as self-interest or desire for retribution)

gence, as observed in other relational reasoning tasks (e.g., Andrews, Birney, & Halford, 2006; Vendetti, Wu, & Holyoak, 2014).

To accommodate human working memory limitations, some computational models of relational reasoning (e.g., Hummel & Holyoak, 2003) operate in a sequential fashion. Applied to moral judgments, such models predict that manipulations of attention to aspects of a moral situation will yield systematic differences in judgments (with those aspects attended to earliest and most frequently having the greatest impact). The idea that reasoning depends on the integration of multiple constraints was first proposed in the field of analogical reasoning (Holyoak, 1985; Holyoak & Thagard, 1989). Key tests of this proposal (Spellman & Holyoak, 1996) involved setting up mappings that would be ambiguous if only one type of constraint were considered, and observing whether a second constraint sufficed to yield a unique solution (thereby establishing that people consider the two types of constraints together). Similar tests could potentially be performed to determine whether multiple types of constraints interact to determine moral judgments, as our framework predicts. It will be particularly important to perform such manipulations using participants who hold strong moral values concerning the relevant topics (to ensure that coherence effects truly operate on moral values, and not solely on nonmoral factors that might lead to similar judgment patterns).

Work on analogical reasoning has additional implications for moral reasoning. In particular, moral rules and schemas may be acquired in a manner similar to other relational concepts. The key to successful learning is to establish the range of situations to which a moral schema applies. A single example is often insufficient to yield spontaneous generalization to novel cases (Gick & Holyoak, 1980, 1983). Instead, the most potent mechanism for learning relational concepts is comparison of multiple cases (Gick & Holyoak, 1980; for recent reviews see Gentner, 2010; Holyoak, 2012). Such comparisons support acquisition of relationally defined categories (e.g., Corral & Jones, 2014; Doumas & Hummel, 2013; Goldwater & Gentner, 2015). Comparison of multiple cases can foster learning and generalization of complex negotiation strategies (Loewenstein, Thompson, & Gentner, 1999), and recent evidence indicates that similar learning experiences can facilitate acquisition of a schema for an ethical principle, such as avoiding conflicts of interest (Kim & Loewenstein, 2015). This work suggests that apparent moral lapses may sometimes result from limited understanding of the relevant ethical principles, which can be remedied by appropriate learning experiences.

If our view is correct, then a complete theory of human moral reasoning will require understanding how relational thinking is implemented in the brain (e.g., Knowlton et al., 2012; Waltz et al., 1999). As mentioned earlier, a fronto-parietal network is engaged in tasks that require working memory resources and cognitive control (Duncan, 2010), with the left RLPFC playing a particularly important role when relational reasoning is triggered (Hobeika et al., 2016). We would predict that these brain areas will be active whenever moral judgments require integrating multiple relations.

But critically, our framework implies that the neural bases for moral judgments will vary with learning and expertise in the relevant domains. Neuroimaging studies of relational learning have revealed that the RLPFC is highly active early in relational learning, but that its involvement decreases as more examples are processed (Davis, Goldwater, & Giron, 2016; Tenison, Fincham, & Anderson, 2016). We therefore predict that acquiring novel moral schemas will initially require heavy use of the fronto-parietal network, and specifically RLPFC to abstract relational commonalities across situations to which the same moral schema applies. After sufficient learning, application of a learned moral rule may no longer require the RLPFC. However, reflective modification of moral rules to fit new contexts, or integration of multiple rules, may again activate the RLPFC. This account of the neural substrate for moral rules provides a novel explanation for why some "non-utilitarian" judgments are made quickly and easily (Greene et al., 2001, 2008). An overlearned moral principle such as "do not kill people" can yield fast judgments not because it depends on an evolutionarily ancient cognitive process (as postulated by some dual-system theories), but simply because most normal people have achieved expertise in its application.

In summary, we believe that the framework of deontological coherence, when combined with work on the psychology and the computational cognitive neuroscience of relational thinking, may provide a starting point for future theory development. In addition to implications for theory, one practical aim of work on moral psychology is to help people understand each other's different, sometimes incompatible, moral codes. By understanding *how* people make moral judgments we may hope to discover better ways to resolve misunderstandings and reconcile disagreements. A mutual appreciation of the perceived moral bases for opposing views will not solve all the world's problems, but perhaps might buy us a bit of quiet time to talk them over. As a descriptive framework, deontology supports this goal by defining morality in terms of basic concepts that ordinary people—even young children—actually grasp. The notions of rights and duties thus provides a common conceptual vocabulary—a meeting point—that potentially enables people who profoundly disagree with one another to "see where the other is coming from."

References

Alicke, M. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126,* 556–574. http://dx.doi.org/10.1037/0033-2909.126.4.556

Almond, B. (1991). Rights. In P. Singer (Ed.), *A companion to ethics* (pp. 259–269). Oxford, UK: Basil Blackwell.

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science, 24,* 1755–1762. http://dx.doi.org/10.1177/0956797613480507

Ames, D. L., & Fiske, S. T. (2015). Perceived intent motivates people to magnify observed harms. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 112,* 3599–3605. http://dx.doi.org/10.1073/pnas.1501592112

Andrews, G., Birney, D. P., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & Cognition, 34,* 1325–1340. http://dx.doi.org/10.3758/BF03193275

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics, 118,* 73–106. http://dx.doi.org/10.1162/00335530360535153

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development, 2004,* 37–49. http://dx.doi.org/10.1002/cd.96

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences, 17,* 1–10. http://dx.doi.org/10.1017/S0140525X0003301X

Baron, J., Gürçay, B., Moore, A. B., & Starcke, K. (2012). Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese, 189*(1), 107–117.

Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology: Applied, 6,* 183–194. http://dx.doi.org/10.1037/1076-898X.6.3.183

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94,* 74–85. http://dx.doi.org/10.1016/j.obhdp.2004.03.003

Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. *Psychology of Learning and Motivation, 50,* 133–167. http://dx.doi.org/10.1016/S0079-7421(08)00404-0

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70,* 1–16. http://dx.doi.org/10.1006/obhd.1997.2690

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108,* 381–417. http://dx.doi.org/10.1016/j.cognition.2008.03.001

Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices? *Psychological Science, 18,* 24–28. http://dx.doi.org/10.1111/j.1467-9280.2007.01843.x

Beebe, J., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language, 25,* 474–498. http://dx.doi.org/10.1111/j.1468-0017.2010.01398.x

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation with moral rules. *Perspectives on Psychological Science, 5,* 187–202. http://dx.doi.org/10.1177/1745691610362354

Bentham, J. (2009). *An introduction to the principles of morals and legislation* (2nd ed.). Mineola, NY: Dover. (Original work published 1823)

Berger, P. (1967). *The sacred canopy: Elements of a sociological theory of religion*. Garden City, NY: Doubleday.

Bleske-Rechek, A., Nelson, L. A., Baker, J. P., & Brandt, S. J. (2010). Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a romantic partner. *Journal of Social, Evolutionary, and Cultural Psychology, 4,* 115–127. http://dx.doi.org/10.1037/h0099295

Borg, S. J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience, 18,* 803–817. http://dx.doi.org/10.1162/jocn.2006.18.5.803

Broeders, R., van den Bos, K., Müller, P. A., & Ham, J. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology, 47,* 923–934. http://dx.doi.org/10.1016/j.jesp.2011.03.018

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist, 64,* 1–11. http://dx.doi.org/10.1037/a0010932

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

Chaxel, A.-S. (2015). The impact of a relational mindset on information distortion. *Journal of Experimental Social Psychology, 60,* 1–7. http://dx.doi.org/10.1016/j.jesp.2015.04.007

Chao, S.-J., & Cheng, P. W. (2000). The emergence of inferential rules: The use of pragmatic reasoning schemas by preschoolers. *Cognitive Development, 15,* 39–62. http://dx.doi.org/10.1016/S0885-2014(00)00018-6

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416. http://dx.doi.org/10.1016/0010-0285(85)90014-3

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition, 33,* 285–313. http://dx.doi.org/10.1016/0010-0277(89)90031-0

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328. http://dx.doi.org/10.1016/0010-0285(86)90002-2

Chi, M. T. H., & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 371–399). New York, NY: Cambridge University Press.

Chituc, V., Henne, P., Sinnott-Armstrong, W., & De Brigard, F. (2016). Blame, not ability, impacts moral "ought" judgments for impossible actions: Toward an empirical refutation of "ought" implies "can." *Cognition, 150,* 20–25. http://dx.doi.org/10.1016/j.cognition.2016.01.013

Cho, S., Moody, T. D., Fernandino, F., Mumford, J. A., Poldrack, R. A., Cannon, T. D., . . . Holyoak, K. J. (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex, 20,* 524–533. http://dx.doi.org/10.1093/cercor/bhp121

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology, 73,* 481–494. http://dx.doi.org/10.1037/0022-3514.73.3.481

Ciaramelli, E., Muccioli, M., Làdavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2*(2), 84–92. http://dx.doi.org/10.1093/scan/nsm001

Cikara, M., Farnsworth, R., Harris, L. T., & Fiske, S. T. (2010). On the wrong side of the trolley track: Neural correlates of relative social valuation. *Social Cognitive and Affective Neuroscience, 5,* 404–413. http://dx.doi.org/10.1093/scan/nsq011

Clark, C. J., Chen, E., & Ditto, P. H. (2015). Moral coherence processes: Constructing culpability and consequences. *Current Opinion in Psychology, 132,* 280–300.

Corral, D., & Jones, M. (2014). The effects of higher-order structure on relational learning. *Cognition, 132,* 280–300. http://dx.doi.org/10.1016/j.cognition.2014.04.007

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences, 17,* 363–366. http://dx.doi.org/10.1016/j.tics.2013.06.005

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107,* 17433–17438. http://dx.doi.org/10.1073/pnas.1009396107

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 27–51). Cham, Switzerland: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-19671-8_2

Cushman, F. (2013). Action, outcome and value: A dual-system framework for morality. *Personality and Social Psychology Review, 17,* 273–292. http://dx.doi.org/10.1177/1088868313495594

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion, 12,* 2–7. http://dx.doi.org/10.1037/a0025071

Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition, 108,* 281–289. http://dx.doi.org/10.1016/j.cognition.2008.02.005

Cushman, F., Murray, D., Gordon-Mckeon, S., Wharton, S., & Greene, J. D. (2012). Judgment before principle: Engagement of the frontoparietal control network in condemning harms of omission. *Social Cognitive and Affective Neuroscience, 7,* 888–895. http://dx.doi.org/10.1093/scan/nsr072

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment. *Psychological Science, 17,* 1082–1089. http://dx.doi.org/10.1111/j.1467-9280.2006.01834.x

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2015). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review, 123,* 2–22. http://dx.doi.org/10.1037/a0039802

Dancy, J. (1991). intuitionism. In P. Singer (Ed.), *A companion to ethics* (pp. 411–420). Oxford, UK: Basil Blackwell.

D'Andrade, R. (1982, April). Reason versus logic. Paper presented at the Symposium on the Ecology of Cognition: Biological, cultural, and historical perspectives. Greensboro, NC.

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

Darwall, S. (Ed.), (2002). *Deontology*. Hoboken, N. J.: Wiley-Blackwell.

Davis, N. (1991). Contemporary deontology. In P. Singer (Ed.), *A companion to ethics* (pp. 205–218). Oxford, UK: Basil Blackwell.

Davis, T., Goldwater, M., & Giron, J. (2016). From concrete examples to abstract relations: The rostrolateral prefrontal cortex integrates novel examples into relational categories. *Cerebral Cortex*. [Advance online publication]. http://dx.doi.org/10.1093/cercor/bhw099

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

DeScioli, P., & Kurzban, R. (2012). A solution to the mysteries of morality. *Psychological Bulletin, 139,* 477–496. http://dx.doi.org/10.1037/a0029065

DiSessa, A. A. (1982). Unlearning Aristotelean physics: A study of knowledge-based learning. *Cognitive Science, 6,* 37–75. http://dx.doi.org/10.1207/s15516709cog0601_2

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 50, pp. 307–338). New York, NY: Academic Press. http://dx.doi.org/10.1016/S0079-7421(08)00410-6

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Noûs (Detroit, Mich.), 32,* 504–530. http://dx.doi.org/10.1111/0029-4624.00136

Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PLoS ONE, 8,* e63889. http://dx.doi.org/10.1371/journal.pone.0063889

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115,* 1–43. http://dx.doi.org/10.1037/0033-295X.115.1.1

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behavior. *Trends in Cognitive Sciences, 14,* 172–179. http://dx.doi.org/10.1016/j.tics.2010.01.004

Eggleston, B., & Miller, D. (Eds.). (2014). *The Cambridge companion to utilitarianism*. Cambridge, UK: Cambridge University Press.

Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review, 5,* 125–131. http://dx.doi.org/10.1177/1754073912463617

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). New York, NY: Oxford University Press.

Evans, J. St. B. T. (2009). How many dual-process theories do we need: One, two or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199230167.003.0002

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13,* 1–17. http://dx.doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781316104668

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transgressions that transgress the spheres of justice. *Political Psychology, 18,* 255–297. http://dx.doi.org/10.1111/0162-895X.00058

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5–15.

Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology, 50,* 1–26. http://dx.doi.org/10.1037/a0037077

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34,* 752–775. http://dx.doi.org/10.1111/j.1551-6709.2010.01114.x

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355. http://dx.doi.org/10.1016/0010-0285(80)90013-4

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38. http://dx.doi.org/10.1016/0010-0285(83)90002-6

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127–171. http://dx.doi.org/10.1016/0010-0277(92)90060-U

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology, 29,* 781–795. http://dx.doi.org/10.1002/(SICI)1099-0992(199908/09)29:5/6<781::AID-EJSP960>3.0.CO;2-3

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition, 137,* 137–153. http://dx.doi.org/10.1016/j.cognition.2014.12.001

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96,* 1029–1046. http://dx.doi.org/10.1037/a0015141

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

Greene, J. D. (2014). The cognitive neuroscience of moral judgment and decision making. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences V* (pp. 1014–1023). Cambridge, MA: MIT Press.

Greene, J. D., Cushman, F., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111,* 364–371. http://dx.doi.org/10.1016/j.cognition.2009.02.001

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6,* 517–523. http://dx.doi.org/10.1016/S1364-6613(02)02011-9

Greene, J. D., Morelli, S., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107,* 1144–1154. http://dx.doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389–400. http://dx.doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral

judgment. *Science, 293,* 2105–2108. http://dx.doi.org/10.1126/science.1062872

Griffin, D. W., Gonzalez, R., Koehler, D. J., & Gilovich, T. (2012). Judgmental heuristics: A historical overview. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 322–345). Oxford, UK: Oxford University Press.

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

Haidt, J., & Baron, J. (1996). Social roles and the moral judgment of acts and omissions. *European Journal of Social Psychology, 26,* 201–218. http://dx.doi.org/10.1002/(SICI)1099-0992(199603)26:2<201::AID-EJSP745>3.0.CO;2-J

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

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

Haidt, J., & Kesebir, M. A. (2010). Morality. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th edition) (pp. 797–832). New York, NY: Wiley. http://dx.doi.org/10.1002/9780470561119.socpsy002022

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

Halford, G. S., Wilson, W. H., Andrews, G., & Phillips, S. (2014). *Categorizing cognition: Conceptual coherence in the foundations of psychology*. Cambridge, MA: MIT Press.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803–831. http://dx.doi.org/10.1017/S0140525X98001769

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14,* 497–505. http://dx.doi.org/10.1016/j.tics.2010.08.005

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

Hauser, M., Cushman, F., Young, L., Jin, R. K., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22,* 1–21. http://dx.doi.org/10.1111/j.1468-0017.2006.00297.x

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology, 21,* 107–112. http://dx.doi.org/10.1080/00223980.1946.9917275

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley. http://dx.doi.org/10.1037/10628-000

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy, 106,* 587–612. http://dx.doi.org/10.5840/jphil200910611 28

Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R., & Volle, E. (2016). General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies. *Human Brain Mapping, 37,* 1953–1969. http://dx.doi.org/10.1002/hbm.23149

Hoffman, E., McCabe, K., Shachat, J., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7,* 346–380. http://dx.doi.org/10.1006/game.1994.1056

Hohfeld, W. N. (1919). Some fundamental legal conceptions as applied in judicial reasoning. In W. W. Cook (Ed.), *Some fundamental legal conceptions as applied in judicial reasoning and other legal essays, by Wesley Newcomb Hohfeld* (pp. 23–64). New Haven, CT: Yale University Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59–87). New York, NY: Academic Press.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 234–259). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199734689.001.0001

Holyoak, K. J., & Cheng, P. W. (1995a). Pragmatic reasoning with a point of view. *Thinking & Reasoning, 1,* 289–313. http://dx.doi.org/10.1080/13546789508251504

Holyoak, K. J., & Cheng, P. W. (1995b). Pragmatic reasoning about human voluntary action: Evidence from Wason's selection task. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 67–89). Hove, East Sussex, UK: Erlbaum.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128,* 3–31. http://dx.doi.org/10.1037/0096-3445.128.1.3

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355. http://dx.doi.org/10.1207/s15516709cog1303_1

Horne, Z., & Powell, D. (2013). More than a feeling: When emotions don't predict moral judgments. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PLoS ONE, 11,* e0154780. http://dx.doi.org/10.1371/journal.pone.0154780

Horne, Z., Powell, D., & Hummel, J. (2015). A single counterexample leads to moral belief revision. *Cognitive Science, 39,* 1950–1964. http://dx.doi.org/10.1111/cogs.12223

Horne, Z., Powell, D., Hummel, J. E., & Holyoak, K. J. (2015). Countering antivaccination attitudes. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 112,* 10321–10324. http://dx.doi.org/10.1073/pnas.1504019112

Horne, Z., Powell, D., & Spino, J. (2013). Belief updating in moral dilemmas. *Review of Philosophy and Psychology, 4,* 705–714. http://dx.doi.org/10.1007/s13164-013-0159-y

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 427–466. http://dx.doi.org/10.1037/0033-295X.104.3.427

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264. http://dx.doi.org/10.1037/0033-295X.110.2.220

Jern, A., Chang, K. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review, 121,* 206–224. http://dx.doi.org/10.1037/a0035941

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience, 7*(4), 393–402. http://dx.doi.org/10.1093/scan/nsr005

Kagan, S. (1989). *The limits of morality*. Oxford, UK: Clarendon Press.

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

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kamm, F. M. (1994). Action, omission, and the stringency of duties. *University of Pennsylvania Law Review, 142,* 1493–1512. http://dx.doi.org/10.2307/3312460

Kanngiesser, P., & Hood, B. M. (2014). Young children's understanding of ownership rights for newly made objects. *Cognitive Development, 29,* 30–40. http://dx.doi.org/10.1016/j.cogdev.2013.09.003

Kant, I. (1953). *Groundwork of the metaphysics of morals*. Translated as *The moral law* by H. J. Paton. London, UK: Hutchinson. (Original work published 1785)

Kim, J., & Loewenstein, J. (2015, August 11). Learning about ethics: Analogical encoding increases moral awareness. Paper presented at the Academy of Management 2015 Annual Meeting. Vancouver, B. C., Canada.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63,* 190–194. http://dx.doi.org/10.1093/analys/63.3.190

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33,* 315–329. http://dx.doi.org/10.1017/S0140525X10000907

Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 441–444). Cambridge, MA: MIT Press.

Knowlton, B. J., Morrison, R. G., Hummel, J. E., & Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends in Cognitive Sciences, 16,* 373–381. http://dx.doi.org/10.1016/j.tics.2012.06.002

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446,* 908–911. http://dx.doi.org/10.1038/nature05631

Kohlberg, L. (1963). The development of children's orientations toward a moral order. *Vita Humana, 6,* 11–33.

Kohlberg, L. (1981). *Essays on moral development, Vol. I: The philosophy of moral development*. San Francisco, CA: Harper & Row.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making, 8*(5), 527–539.

Korchmaros, J. D., & Kenny, D. A. (2001). Emotional closeness as a mediator of the effect of genetic relatedness on altruism. *Psychological Science, 12,* 262–265. http://dx.doi.org/10.1111/1467-9280.00348

Kruger, D. J. (2003). Evolution and altruism: Combining psychological mediators with naturally selected tendencies. *Evolution and Human Behavior, 24,* 118–125. http://dx.doi.org/10.1016/S1090-5138(02)00156-3

Kruglanski, A. W., & Shteynberg, G. (2012). Cognitive consistency as means to an end: How subjective logic affords knowledge. In B. Gawronski & F. Strack (Eds.), *Cognkitive consistency: A fundamental principle in social cognition* (pp. 245–264). New York, NY: Guilford Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498. http://dx.doi.org/10.1037/0033-2909.108.3.480

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103,* 284–308. http://dx.doi.org/10.1037/0033-295X.103.2.284

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior, 33,* 323–333. http://dx.doi.org/10.1016/j.evolhumbehav.2011.11.002

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect theory of mind and moral judgment. *Psychological Science, 17,* 421–427. http://dx.doi.org/10.1111/j.1467-9280.2006.01722.x

Lewin, K. (1938). *The conceptual representation and the measurement of psychological forces*. Durham, NC: Duke University Press. http://dx.doi.org/10.1037/13613-000

Liao, S. M., Wiegmann, A., Alexander, J., & Vong, G. (2012). Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology, 25,* 661–671. http://dx.doi.org/10.1080/09515089.2011.627536

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511618031

Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science, 4,* 316–323. http://dx.doi.org/10.1177/1948550612456045

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review, 6,* 586–597. http://dx.doi.org/10.3758/BF03212967

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science, 33,* 273–286. http://dx.doi.org/10.1111/j.1551-6709.2009.01013.x

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review, 119,* 617–648. http://dx.doi.org/10.1037/a0028719

Machery, E., & Mallon, M. (2010). The evolution of morality. In J. M. Doris, & the Moral Psychology Research Group (Ed.), *The moral psychology handbook* (pp. 3–46). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199582143.003.0002

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, MA: MIT Press.

McCloskey, H. J. (1957). An examination of restricted utilitarianism. *The Philosophical Review, 66,* 466–485. http://dx.doi.org/10.2307/2182745

McCloy, R., & Byrne, R. (2000). Counterfactual thinking about controllable events. *Memory & Cognition, 28,* 1071–1078. http://dx.doi.org/10.3758/BF03209355

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

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511780578

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology, 67,* 371–378. http://dx.doi.org/10.1037/h0040525

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18,* 57–76. http://dx.doi.org/10.1177/001872676501800105

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York, NY: Harper & Row.

Mill, J. S. (2004). *Utilitarianism and other essays*. London, UK: Penguin Books. (Original work published 1861)

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19,* 549–557. http://dx.doi.org/10.1111/j.1467-9280.2008.02122.x

Moore, M. (2008). Patrolling the borders of consequentialist justifications: The scope of agent-relative restrictions. *Law and Philosophy, 27,* 35–96. http://dx.doi.org/10.1007/s10982-007-9011-9

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108,* 2688–2692. http://dx.doi.org/10.1073/pnas.1011734108

Moretto, G., Làdavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial

prefrontal damage. *Journal of Cognitive Neuroscience, 22*(8), 1888–1899. http://dx.doi.org/10.1162/jocn.2009.21367

Nagel, T. (1972). War and massacre. *Philosophy & Public Affairs, 1,* 123–144.

Nagel, T. (1986). *The view from nowhere*. New York, NY: Oxford University Press.

Neyer, F. J., & Lang, F. R. (2003). Blood is thicker than water: Kinship orientation across adulthood. *Journal of Personality and Social Psychology, 84,* 310–321. http://dx.doi.org/10.1037/0022-3514.84.2.310

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100,* 530–542. http://dx.doi.org/10.1016/j.cognition.2005.07.005

Nietzsche, F. (1996). *On the genealogy of morals* (D. Smith, Trans.). Oxford, UK: Oxford World's Classics. (Original work published 1887)

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238,* 625–631. http://dx.doi.org/10.1126/science.3672116

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259. http://dx.doi.org/10.1037/0033-295X.84.3.231

Noles, N. S., Keil, F. C., Bloom, P., & Gelman, S. A. (2012). Children's and adults' intuitions about who can own things. *Journal of Cognition and Culture, 12,* 265–286. http://dx.doi.org/10.1163/15685373-12342076

N′gbala, A., & Branscombe, N. (1997). When does action elicit more regret than inaction and is counterfactual mutation the mediator of this effect? *Journal of Experimental Social Psychology, 33,* 324–343. http://dx.doi.org/10.1006/jesp.1996.1322

O'Neill, O. (1991). Kantian ethics. In P. Singer (Ed.), *A companion to ethics* (pp. 175–185). Oxford, UK: Basil Blackwell.

O'Neill, P., & Petrinovich, L. (1998). A preliminary cross-study of moral intuitions. *Evolution and Human Behavior, 19,* 349–367. http://dx.doi.org/10.1016/S1090-5138(98)00030-0

Pence, G. (1991). Virtue theory. In P. Singer (Ed.), *A companion to ethics* (pp. 249–258). Oxford, UK: Basil Blackwell.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–130. http://dx.doi.org/10.1017/S0140525X08003543

Pessoa, L., & Pereira, M. G. (2013). Cognition–emotion interactions: A review of the functional magnetic resonance imaging literature. In M. D. Robinson, E. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 55–68). New York, NY: Guilford Press.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology, 64,* 467–478. http://dx.doi.org/10.1037/0022-3514.64.3.467

Pettit, P. (1991). Consequentialism. In P. Singer (Ed.), *A companion to ethics* (pp. 230–240). Oxford, UK: Basil Blackwell.

Phillips, J., & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry, 20,* 30–36.

Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about promises and warnings: Darwinian algorithms, mental models, relevance judgments or pragmatic schemas? *Quarterly Journal of Experimental Psychology, 44A,* 402–421.

Portmore, D. W. (2011). *Commonsense consequentialism: Wherein morality meets rationality*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199794539.001.0001

Prehn, K., Korczykowski, M., Rao, H., Fang, Z., Detre, J. A., & Robertson, D. C. (2015). Neural correlates of post-conventional moral reasoning: A voxel-based morphometry study. *PLoS ONE, 10,* e0122914. http://dx.doi.org/10.1371/journal.pone.0122914

Prinz, J. J. (2008). Resisting the linguistic analogy: A commentary on Hauser, Young, and Cushman. In *Moral psychology, Vol. 2*: *The cognitive science of morality: Intuition and diversity* (pp. 1–11). Cambridge, MA: MIT Press.

Quinn, W. S. (1989). Actions, intentions, and consequences: The doctrine of doing and allowing. *The Philosophical Review, 98,* 287–312. http://dx.doi.org/10.2307/2185021

Rachels, J. (1975). Active and passive euthanasia. *The New England Journal of Medicine, 292,* 78–80. http://dx.doi.org/10.1056/NEJM197501092920206

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118,* 57–75. http://dx.doi.org/10.1037/a0021867

Rai, T. S., & Holyoak, K. J. (2010). Moral principles or consumer preferences? Alternative framings of the trolley problem. *Cognitive Science, 34,* 311–321. http://dx.doi.org/10.1111/j.1551-6709.2009.01088.x

Rai, T. S., & Holyoak, K. J. (2013). Exposure to moral relativism compromises moral behavior. *Journal of Experimental Social Psychology, 49,* 995–1001. http://dx.doi.org/10.1016/j.jesp.2013.06.008

Rai, T. S., & Holyoak, K. J. (2014). Rational hypocrisy: A Bayesian analysis based on informal argumentation and slippery slopes. *Cognitive Science, 38,* 1456–1467. http://dx.doi.org/10.1111/cogs.12120

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

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York, NY: Oxford University Press.

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 109,* 14824–14829. http://dx.doi.org/10.1073/pnas.1203179109

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology, 25,* 1731–1735. http://dx.doi.org/10.1016/j.cub.2015.05.014

Ritov, I., & Baron, J. (1994). Judgments of compensation for misfortune: The role of expectation. *European Journal of Social Psychology, 24,* 525–539. http://dx.doi.org/10.1002/ejsp.2420240502

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121,* 133–148. http://dx.doi.org/10.1037/0033-2909.121.1.133

Ross, W. D. (1930). *The right and the good*. Oxford, UK: Clarendon Press.

Rudolph, U., Roesch, S. C., Greitemeyer, T., & Weiner, B. (2004). A metaanalytic review of help giving and aggression from an attributional perspective: Contributions to a general theory of motivation. *Cognition and Emotion, 18,* 815–848. http://dx.doi.org/10.1080/02699930341000248

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Russo, J. E. (2014). The predecisional distortion of information. In E. A. Wilhelms & V. F. Reyna (Eds.), *Neuroeconomics, judgment, and decision making* (pp. 91–110). New York, NY: Psychology Press.

Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology: General, 137,* 456–470. http://dx.doi.org/10.1037/a0012786

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research, 35,* 438–452. http://dx.doi.org/10.2307/3152163

Scheffler, S. (1982). *The rejection of consequentialism*. Oxford, UK: Clarendon Press.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and nonphilosophers. *Mind & Language, 27,* 135–153. http://dx.doi.org/10.1111/j.1468-0017.2012.01438.x

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The*

Journal of Neuroscience, 34, 4741–4749. http://dx.doi.org/10.1523/JNEUROSCI.3390-13.2014

Shultz, T. R., Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility, and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement, 13*(3), 238–253. http://dx.doi.org/10.1037/h0081183

Shultz, T. R., Wright, K., Schleifer, M., & Url, S. (1986). Assignment of moral responsibility and punishment. *Child Development, 57*(1), 177–184

Sidgwick, H. (1981). *The methods of ethics* (7th ed.). Indianapolis, IN: Hackett Publishing. (Original work published 1907)

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *The University of Chicago Law Review, 71,* 511–586.

Simon, D. (2012). *In doubt: The psychology of the criminal justice process*. Cambridge, MA: Harvard University Press. http://dx.doi.org/10.4159/harvard.9780674065116

Simon, D., & Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review, 6,* 283–294. http://dx.doi.org/10.1207/S15327957PSPR0604_03

Simon, D., Krawczyk, D. C., Bleicher, A., & Holyoak, K. J. (2008). The transience of constructed preferences. *Journal of Behavioral Decision Making, 21,* 1–14. http://dx.doi.org/10.1002/bdm.575

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science, 15,* 331–336. http://dx.doi.org/10.1111/j.0956-7976.2004.00678.x

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1250–1260. http://dx.doi.org/10.1037/0278-7393.27.5.1250

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology, 86,* 814–837.

Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending hot and cold cognitions. *Journal of Personality and Social Psychology, 109,* 369–394. http://dx.doi.org/10.1037/pspa0000029

Singer, P. (1979). *Practical ethics*. Cambridge, UK: Cambridge University Press.

Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics, 9,* 331–352. http://dx.doi.org/10.1007/s10892-005-3508-y

Sinnott-Armstrong, W. (2008). Framing moral intuitions. In W. Sinnott-Armstrong (Ed.), *Moral psychology Vol. 2*: *The cognitive science of morality intuition and diversity* (pp. 47–76). Cambridge, MA: MIT Press.

Skitka, L., Bauman, C., & Sargis, E. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88,* 895–917. http://dx.doi.org/10.1037/0022-3514.88.6.895

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science, 16,* 1–40. http://dx.doi.org/10.1207/s15516709cog1601_1

Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62,* 913–933. http://dx.doi.org/10.1037/0022-3514.62.6.913

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology, 31,* 307–346. http://dx.doi.org/10.1006/cogp.1996.0019

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A coherence model of cognitive consistency. *Journal of Social Issues, 49,* 147–165. http://dx.doi.org/10.1111/j.1540-4560.1993.tb01185.x

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27,* 76–105. http://dx.doi.org/10.1016/0022-1031(91)90011-T

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stewart-Williams, S. (2007). Altruism among kin vs. non-kin: Effects of cost of help and reciprocal exchange. *Evolution and Human Behavior, 28,* 193–198. http://dx.doi.org/10.1016/j.evolhumbehav.2007.01.002

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences, 28,* 531–541. http://dx.doi.org/10.1017/S0140525X05000099

Swann, W. B., Gómez, A., Dovidio, J. F., Hart, S., & Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science, 21,* 1176–1183. http://dx.doi.org/10.1177/0956797610376656

Tenison, C., Fincham, J. M., & Anderson, J. R. (2016). Phases of learning: How skill acquisition impacts cognitive processing. *Cognitive Psychology, 87,* 1–28. http://dx.doi.org/10.1016/j.cogpsych.2016.03.001

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109,* 451–471. http://dx.doi.org/10.1037/0033-295X.109.3.451

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–467. http://dx.doi.org/10.1017/S0140525X00057046

Thagard, P. (2000). *Coherence in thoughts and action*. Cambridge, MA: MIT Press.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.

Thagard, P., & Nerb, J. (2002). Emotional gestalts: Appraisal, change and the dynamics of affect. *Personality and Social Psychology Review, 6,* 274–282. http://dx.doi.org/10.1207/S15327957PSPR0604_02

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59,* 204–217. http://dx.doi.org/10.5840/monist197659224

Thurstone, L. (1927). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology, 21,* 384–400. http://dx.doi.org/10.1037/h0065439

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neuroscience of loss aversion in decision-making under risk. *Science, 315,* 515–518. http://dx.doi.org/10.1126/science.1134239

Treadway, M. T., Buckholtz, J. W., Martin, J. W., Jan, K., Asplund, C. L., Ginther, M. R., . . . Marois, R. (2014). Corticolimbic gating of emotion-driven punishment. *Nature Neuroscience, 17,* 1270–1275. http://dx.doi.org/10.1038/nn.3781

Trémolière, B., & Bonnefon, J.-F. (2014). Efficient kill-save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin, 40,* 923–930. http://dx.doi.org/10.1177/0146167214530436

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making, 4,* 476–491.

Uhlmann, E. L., Zhu, L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition, 126,* 326–334. http://dx.doi.org/10.1016/j.cognition.2012.10.005

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science, 25,* 1–6.

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

Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 364–389). Oxford, UK: Oxford University Press.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., . . . Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10,* 119–125. http://dx.doi.org/10.1111/1467-9280.00118

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (Vol. I). Harmondsworth, UK: Penguin.

Webster, G. D. (2003). Prosocial behavior in families: Moderators of resource sharing. *Journal of Experimental Social Psychology, 39,* 644–652. http://dx.doi.org/10.1016/S0022-1031(03)00055-6

Wertheimer, M. (1923/1967). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt theory* (pp. 71–88). New York, NY: Humanities Press.

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

Wiegmann, A., Nagel, J., & Mangold, S. (2008). Order effects in moral judgment. *Philosophical Psychology, 25,* 2111–2116.

Wiegmann, A., & Okan, Y. (2012). Order effects in moral judgment: Searching for an explanation. *Proceedings of the thirty-fourth annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128. http://www.sciencedirect.com/science/article/pii/0010027783900045. http://dx.doi.org/10.1016/0010-0277(83)90004-5

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 104,* 8235–8240. http://dx.doi.org/10.1073/pnas.0701408104

Young, L., & Durwin, A. J. (2013). Moral realism as moral motivation: The impact of meta-ethics on everyday decision-making. *Journal of Experimental Social Psychology, 49,* 302–306. http://dx.doi.org/10.1016/j.jesp.2012.11.013

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

Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47,* 2065–2072. http://dx.doi.org/10.1016/j.neuropsychologia.2009.03.020

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120,* 202–214. http://dx.doi.org/10.1016/j.cognition.2011.04.005

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology, 24,* 358–365. http://dx.doi.org/10.1037/0012-1649.24.3.358

Zamir, E., & Medina, B. (2010). *Law, economics, and morality*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195372168.001.0001

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development, 67,* 2478–2492. http://dx.doi.org/10.2307/1131635

Received February 17, 2016
Revision received July 7, 2016
Accepted July 26, 2016 ■