Personality and Social Psychology Review
17(3) 273–292

DOI: 10.1177/1088868313495594
pspr.sagepub.com

Action, Outcome, and Value: A Dual-System Framework for Morality

Fiery Cushman[1]

Abstract

Dual-system approaches to psychology explain the fundamental properties of human judgment, decision making, and behavior across diverse domains. Yet, the appropriate characterization of each system is a source of debate. For instance, a large body of research on moral psychology makes use of the contrast between "emotional" and "rational/cognitive" processes, yet even the chief proponents of this division recognize its shortcomings. Largely independently, research in the computational neurosciences has identified a broad division between two algorithms for learning and choice derived from formal models of reinforcement learning. One assigns value to actions intrinsically based on past experience, while another derives representations of value from an internally represented causal model of the world. This division between action- and outcome-based value representation provides an ideal framework for a dual-system theory in the moral domain.

Keywords

dual-system theory, morality, emotion, reasoning, reinforcement learning

A central aim of the current research in moral psychology is to characterize a workable dual-system framework (Bartels, 2008; Cushman, Young, & Hauser, 2006; Greene, 2007; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001; Pizarro & Bloom, 2003; Young & Koenigs, 2007). Candidate systems are sometimes described in terms of intuition versus reasoning, automaticity versus control, and emotion versus cognition. While there is considerable agreement that a dual-system framework of some kind is necessary, agreement is far from universal; moreover, the contours, functions, and relative influence of each are disputed (Haidt, 2001; Huebner, Dwyer, & Hauser, 2009; Kvaran & Sanfey, 2010; Moll, De Oliveira-Souza, & Zahn, 2008; Paxton & Greene, 2010; Pizarro & Bloom, 2003).

Meanwhile, current research into the neurological basis of decision making has been transformed by the identification of two broad classes of algorithms that structure learning and behavioral choice. One algorithm encodes the value of actions by associating them with subsequent punishment and reward, leveraging prediction error and temporal difference learning to efficiently represent whether an action is rewarding without representing what makes it so. The other algorithm encodes the value of outcomes and selects actions based on their expected value, relying on a probabilistic causal model that relates actions, outcomes, and rewards. This broad distinction, first identified in the machine learning literature (Sutton & Barto, 1999), provides a valuable description of the two systems of learning and decision making in the human brain (Daw & Doya, 2006; Daw & Shohamy, 2008; Schultz, Dayan, & Montague, 1997).

How can we bridge the gap between these literatures, applying insights from computational neuroscience to the qualitative distinction between two systems widespread in moral judgment research? This requires formulating core concepts from each domain in common terms. The approach pursued here is to distinguish two systems of value representation—action- versus outcome-based—and to show that the connection between the computational models and psychological phenomena is more than a loose analogy or family resemblance. Rather, the models provide a detailed basis for understanding otherwise peculiar facets of moral judgment.

Dual-System Morality

The utility of a dual-system framework for understanding human judgment, reasoning, and behavior has been forcefully argued elsewhere (Epstein, 1994; Kahneman, 2011; Sloman, 1996; Stanovich & West, 2000). Some of the evidence that motivates a dual-system approach specifically for moral judgment is reviewed below, but equally emphasized are the shortcomings of the current characterizations of the systems (e.g., as emotional vs. rational). I argue for an alternative characterization of the systems that distinguishes

[1]Brown University, Providence, RI, USA

Corresponding Author:
Fiery Cushman, CLPS Department, Brown University, Box 1821, 190 Thayer St, Providence, RI 02912, USA.
Email: fiery_cushman@brown.edu

between valuing the intrinsic status of actions (e.g., "I must tell George the truth because lying is wrong") and valuing the expected consequences of actions (e.g., "if I deceive George it will ultimately cause him harm").

My discussion centers largely on the extensive literature motivated by moral dilemmas such as the trolley problem, but not because the ethics of improbable railway crises carry much intrinsic interest. Rather, cases such as the trolley problem appear to efficiently dissociate between processes of moral judgment that apply to a much broader set of empirical phenomena, and now constitute the *lingua franca* of dozens of studies. As such, they are a useful proving ground for the action/outcome framework.

The Trolley Problem and the Aversion to Harm

The trolley problem contrasts two cases—the "switch" case and the "push" case (Foot, 1967; Thomson, 1985). In the "switch" case, a runaway trolley threatens five workers ahead on the tracks, and it must be decided whether to flip a switch that diverts the trolley onto a side track where only one person is threatened. In the "push" case, five people are similarly threatened but there is no side track—instead, it must be decided whether to throw a person in front of the trolley to slow it down. A great majority of people say that one person should be sacrificed for the sake of five in the first case, whereas a small minority say that one person should be sacrificed in the second case (Greene et al., 2001; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Mikhail, 2000). Therein lies the problem: Given that in each case one person is sacrificed for the welfare of five, what accounts for this discrepancy?

An influential answer to this challenge rests on a dual-system theory that contrasts cognitive and emotional processing (Greene, 2007; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene et al., 2001). The controlled cognitive process is posited to underlie the welfare maximizing choice in both cases, diverting the train as well as pushing the man, because of the greater number of lives saved. The automatic emotional process is posited to underlie the aversion to doing harm in an up-close and personal manner, and thus to be engaged exclusively in the "push" case. Thus, the dual-system theory explains why the cases are judged differently (because the emotional system is more strongly engaged for the "push" case), and also why the push case seems to present a more difficult dilemma than the switch case (because both systems are engaged in the push case, but only the cognitive system is engaged in the switch case). In keeping with characteristic positions in the philosophical literature, the choice not to push is often referred to as "deontological," contrasting with the "utilitarian" choice to push.

Several sources of evidence support this division between the two processes. Utilitarian judgment is associated with activation in a network of brain regions that enable controlled, attentive processing, most notably the dorsolateral prefrontal cortex (Cushman, Murray, Gordon–McKeon, Wharton, & Greene, 2012; Greene et al., 2004). Under cognitive load, deontological judgment becomes more likely (Trémolière, De Neys, & Bonnefon, 2012) and utilitarian judgment is slowed (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008); moreover, the latter effect is exclusive to "personal" (i.e., push-type) cases. Meanwhile, deontological responding is reduced in individuals with damage to the ventromedial prefrontal cortex, a brain region thought to play a key role in integrating affect into decision making (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs et al., 2007; Moretto, Ladavas, Mattioli, & di Pellegrino, 2010); again, this effect is selective to personal cases. In addition, deontological responding is enhanced when serotonin levels are raised pharmacologically, consistent with the role of serotonin in aversive learning and inhibitory functions (Crockett, Clark, Hauser, & Robbins, 2010). This effect, too, is selective to personal cases. These and other findings (Conway & Gawronski, 2013; Mendez, Anderson, & Shapria, 2005; Moore, Clark, & Kane, 2008; Paxton, Ungar, & Greene, 2012; Suter & Hertwig, 2011; Valdesolo & DeSteno, 2006) suggest the influence of competing processes in personal cases.

While it is clear that some division between processes is necessary, it is equally clear that the simple division between "cognitive" and "emotional" processes is inadequate (Cushman, Young, & Greene, 2010; Huebner et al., 2009; Kvaran & Sanfey, 2010; Moll et al., 2008; Nucci & Gingo, 2010). First, both processes must involve affective content, in the sense that they do not merely process information but also yield competing motivations toward distinct behaviors. In particular, the system that supports utilitarian responding cannot merely represent the factual content, "5 lives is *more* than 1 life"; rather, it must carry the affective force of "choosing to save 5 lives is *better* than choosing to preserve 1." What is required, then, is not a theory distinguishing affective from nonaffective processing, but a theory distinguishing two processes both of which involve affective content. For instance, two qualitatively distinct kinds of affective response may be at play, or affect may be triggered by distinct factors in each process.

Second, both processes must involve cognition in the sense of information processing. In particular, the psychological mechanisms responsible for a deontological response must be triggered by some set of features that distinguish the push case from the switch case and, presumably, many other features as well. Two features that differ between these cases have been repeatedly demonstrated to trigger deontological response. The first is the manner of physical interaction between the agent and the victim (Cushman et al., 2006; Greene et al., 2009). When the agent directly transfers his or her bodily force onto the victim (as in the push case) this elicits reliably higher levels of moral condemnation than when no such transfer of "personal force" occurs (as in the switch case). The second is more subtle: the status of harm as a means to saving others versus a side-effect of saving others

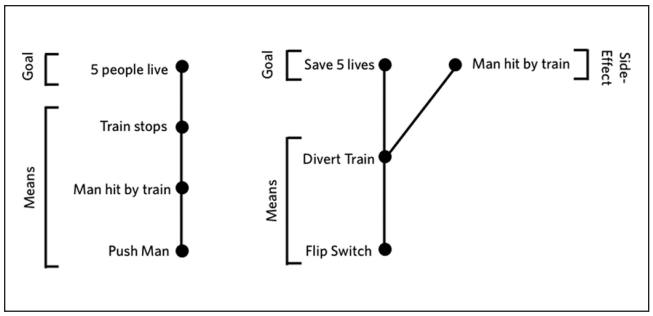

Figure 1. A schematic representation of the causal structure and goal structure of the trolley and footbridge cases, modeled on a schema developed by Goldman (1971) and applied to the moral domain by Mikhail (2000).

Note. In the footbridge case (left) the death of one person plays a causal role in saving five people, and therefore it constitutes a means that the agent employs toward the end of saving five. In the trolley case (right) the death of one person does not play a causal role in saving five people, and therefore it constitutes a side-effect of the agent's actions.

(Figure 1; Cushman et al., 2006; Foot, 1967; Greene et al., 2009; Mikhail, 2000; Royzman & Baron, 2002; Thomson, 1985). In the push case, the victim is used as a "trolley stopper"—a necessary instrument to accomplish the goal of saving five others. By contrast, in the switch case, the victim is merely an unlucky collateral damage. His death happens to be an unavoidable consequence of diverting the train, but saving the five people is not a consequence of his death. Harmful actions are judged more severely when used as a means to accomplish a goal than when brought about as a side-effect of accomplishing a goal. Because deontological judgment is sensitive to these factors and others (Royzman & Baron, 2002; Waldmann & Dieterich, 2007), its affective component must be triggered by computations performed over a detailed representation of the action in question, including its physical affordances, goal structure, and so on.

In summary, a simple division between affective and cognitive processes will not suffice. This point is appreciated by critics and supporters of dual-process theories in moral psychology. Notably, the most pointed critiques of the dual-process theory turn on precisely this point, arguing that affective and cognitive components must ultimately be integrated (Kvaran & Sanfey, 2010; Moll et al., 2008; L. P. Nucci & Gingo, 2010). In this sense the critiques do not necessarily dispute the existence of distinct processes that contribute to moral judgment, but rather to the characterization of the processes as emotional versus cognitive. Likewise, even proponents of a dual-process theory of moral psychology regard the cognition/emotion distinction as an imperfect placeholder. One approach has been to posit that qualitatively different kinds of affective content are involved in each system (Cushman & Greene, 2012; Cushman et al., 2010; Greene, 2007), but the distinction between affective kinds has not been precisely articulated, nor has it been firmly linked to research in the cognitive and affective neurosciences.

Action Versus Outcome in the Moral Domain

How else might the two processes be framed? One alternative is to distinguish between a process that assigns value directly to actions (e.g., hitting) and a process that chooses actions based on the value assigned to their expected outcomes (e.g., a broken nose). These processes would support characteristically deontological and utilitarian judgments, respectively. Critically, each process involves elements of cognition (the identification of relevant action and outcome properties) and affect (the value associated with actions and outcomes, respectively). Note that this approach does not depend on different kinds of affect (i.e., value representation). Rather, it depends on different structural targets for value representation.

Some evidence for a division between action- and outcome-based valuation in the moral domain comes from a study of people's aversion to pretend harmful actions, such as shooting a person with a fake gun or hitting a plastic baby doll against the table (Cushman, Gray, Gaffey, & Mendes, 2012; see also Hood, Donnelly, Leonards, & Bloom, 2010; King, Burton, Hicks, & Drigotas, 2007; Rozin, Millman, & Nemeroff, 1986). Performing such pretend actions reliably elicits peripheral vasoconstriction, a psychophysiological response associated with aversive reactivity (Gregg, James, Matyas, & Thorsteinsson, 1999; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007). This response is greater when performing pretend actions than when witnessing the identical actions or when performing kinetically matched non-harmful actions. Two features of these data indicate the operation of an action-based value representation. First, the aversive reaction occurs despite the absence of any expected harmful outcome. (If any expectation existed, presumably, participants would not pull the trigger!) Second, to the extent that a harmful outcome is imagined or associated with the action, it should be equally imagined or associated in the witness condition. However, peripheral vasoconstriction in the witness condition was no greater than in the no-harm control condition. The selective autonomic response to performing harm thus appears to be tied to features of the action, rather than to any real or imagined features of the outcome.

This is not meant to deny that harmful outcomes are aversive. To the contrary, an outcome-based representation presumably makes all the difference between pulling the trigger and not. One can only hope that if the participants were handed a real gun, and therefore expected a harmful outcome, they would spare the experimenter's life. Rather, the aversion to pretend harmful action observed in this study provides evidence for an additional mechanism—beyond the valuation of outcomes—that assigns value based on properties of the action under consideration.

A recent series of studies suggest that the personal aversion to performing harmful action constitutes an important basis for making moral judgments of third parties (R. Miller, Hannikainan, & Cushman, 2013). Participants first rated

"How upset would you feel" in several circumstances that dissociate aversive action properties from aversive outcome properties. For instance, "stabbing a fellow actor in the neck using a fake stage knife as part of play" assesses the personal aversion to an action divorced from any harmful outcome. Conversely, "seeing a football player break his leg during a game" assesses the personal aversion to harmful outcomes but without performing an action. Then, the participants judged the moral wrongness of others' behaviors in hypothetical moral dilemmas such as the trolley problem. The personal aversion to performing harmful action strongly predicted nonutilitarian moral judgment, even with a 2-year delay between the two tasks; in contrast, the personal aversion to harmful outcomes was at best weakly predictive and often failed to show any significant correlation. (It did, however, correlate well with a widely used measure of empathic concern.) These data suggest that the moral condemnation of harmful actions—even actions performed by third parties—is partly grounded in our personal aversion to performing such actions ourselves (reviewed in Miller & Cushman, in press).

Furthermore, evidence that action-based value representations guide moral judgment comes from the condemnation of "victimless" crimes, such as consensual sibling incest (Graham, Haidt, & Nosek, 2009; Haidt, Koller, & Dias, 1993). Although people often try to explain why such behaviors are morally wrong by appealing to potential harmful outcomes (Ditto & Liu, 2011; Haidt, 2001), they often maintain moral condemnation even in cases where they accept that no harm at all is caused. Presumably, then, these judgments depend on intrinsic properties of the action. In the case of consensual sibling incest, Lieberman and Lobel (2012) have used creative methods to convincingly demonstrate that the moral condemnation of a third-party incest is grounded in the personal aversion that people feel to engaging in incest themselves.

Another source of evidence favoring a distinction between outcome- and action-based value representations in the moral domain comes from one of the most widely studied features guiding patterns of judgment in hypothetical dilemmas: the distinction between active and passive harm (Baron & Ritov, 2009; Cushman et al., 2012; Cushman et al., 2006; DeScioli, Bruening, & Kurzban, 2011; DeScioli, Christner, & Kurzban, 2011; Spranca, Minsk, & Baron, 1991). People generally consider it morally worse to harm a person actively (e.g., by administering a fatal poison) than to passively allow a person to suffer harm (e.g., by withholding a life-saving antidote).[1] Some psychologists and philosophers have considered this pattern of judgment to be mysterious, as the victim ends up dead in either case. In other words, this pattern of moral judgment is difficult to accommodate on a theory of outcome-based value representations. On the other hand, positing a distinct action-based value representation provides an appealing solution: Action-based valuation systems require an action to be triggered.

In certain respects, the action/outcome framework complements the past dual-system approaches. For instance, previous research on action-based value representation associates this mode of decision making with habitual, automatic behaviors, while outcome-based value representation has been more often associated with effortful, controlled behaviors. (This research is reviewed below.) In this respect, the action/outcome framework is a natural complement to the automatic/controlled framework. The relationship between the action/outcome framework and the reason/emotion framework is more complicated. A virtue of the action/outcome framework is that it embraces the role of information processing and value representation in each of the two systems. In this respect, it denies the basic premise of the emotion/reason distinction.[2] Yet, it also places its emphasis on the structural role of value representations in decision making, and in this sense it shares with the reason/emotion framework a fundamental concern with the relationship between knowledge, value, and computation. We will return to consider the relationship between these approaches in greater detail, seeking a more satisfactory resolution between them.

Two Processes of Learning and Decision Making

The distinction between action- and outcome-based value representations finds a clear parallel in computational approaches to reinforcement learning. Two broad classes of learning algorithms were first identified by researchers working on machine learning (Sutton, 1988; Sutton & Barto, 1999), and evidence suggests that these algorithms roughly characterize distinct systems of human learning and decision making.

Reinforcement learning algorithms simultaneously solve two problems: learning and deciding. By guiding an agent's choices in an environment and then computing the reward value obtained they attempt to specify an optimal policy—that is, a set of choices that tend to maximize reward over the long run. Value representations therefore lie at the heart of a reinforcement learning algorithm, serving as a roadmap that allows the past experiences of reward to guide the future choices of action.

Consider, for instance, the reinforcement learning problem depicted in Figure 2. An agent (indicated by a smiley face) must navigate a grid by discrete steps. At each position on the grid, or "state," it can choose from up to four actions (move north, south, east, or west). It receives a large positive reward for arriving at a particular end state (indicated by a star), but small negative rewards in every other state to encourage efficient pursuit of the goal (i.e., aimless wandering gets costly). The agent is allowed to repeatedly explore this space. Over successive rounds of experience, a reinforcement learning algorithm must learn to guide action to attain optimal performance. It does this by constructing value representations.

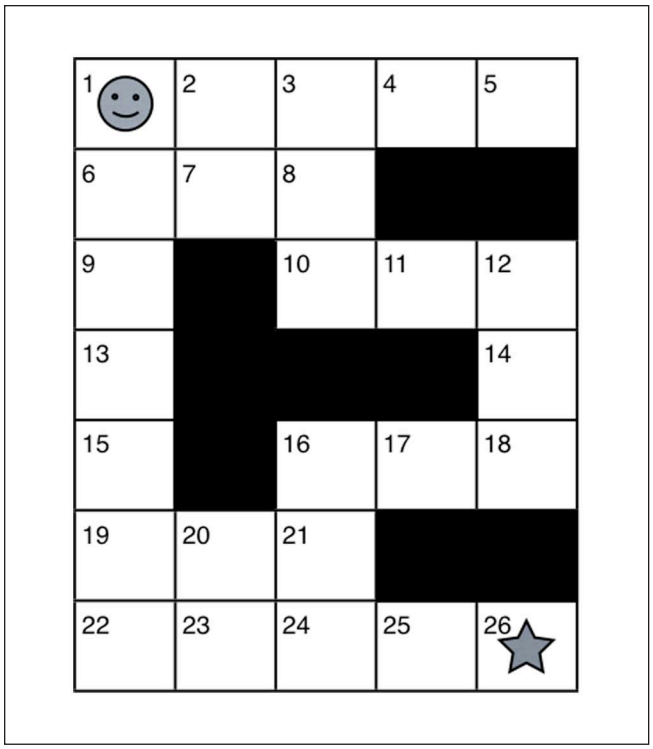

Figure 2. A simple reinforcement learning problem.
Note. The agent (indicated by a smiley face) can perform the actions of moving north, south, east, or west in each of the states denoted numerically. It gains value by reaching the goal state (26, indicated by a star) and loses value on each turn taken before reaching the goal state, thus motivating goal attainment within the minimum possible number of steps.

Model-Based Reinforcement Learning

One class of algorithms that solve the reinforcement learning problem, "model-based" algorithms, corresponds with what we ordinarily recognize as reasoning. In brief, it considers different courses of actions based on an internal representation of its environment, and then tends to choose the course of action that is expected to produce the best overall outcomes.

Model-based algorithms learn by building a causal model of the world they occupy (hence their name). Therefore, as the agent explores the grid world depicted in Figure 2, it constructs and stores an internal representation of the world. This internal representation includes information about all the states, the actions available in each state, the transitions to new states produced by selecting each action, and the reward associated with each state.

Thus, the agent can retrieve information of the form, "I am in State 1, and if I move south I will be in State 6 and will receive a reward of −1." It could retrieve similar information about State 6, and so on, allowing it to compute the accumulated value anticipated for a sequence of possible actions. Using this procedure, the agent simulates various paths that it could take and then selects the path that maximizes reward according to its internal representation. Put simply, it calculates expected values of different options. This is the element of a model-based algorithm that should sound familiar like everyday reasoning. If you have several options to choose from, you imagine the likely outcomes of each option and then select the option that seems likely to deliver the best outcome.

A model-based algorithm operating with a detailed and accurate model can make very good choices. Its choices can be farsighted, in the sense that they specify policies that require many actions to obtain a goal. Its choices can be flexible, in the sense that they can be recomputed at any time to reflect updates to the model. And its choices can be goal-oriented, in the sense that the agent can specify a particular desired outcome and then compute the necessary sequence of steps to attain the specified goal.

However, there is a high computational cost associated with all of these benefits. As the number of the available states and actions grows, so grows the space of the possible policies over which the model-based algorithm searches. This, too, is a familiar property of reasoning. When choosing between different mortgages (typically involving a small number of discrete, quantitative differences), the limited scope of relevant outcome measures allows reasoning to guide optimal choice. However, when choosing between different houses (typically involving a large number of interrelated, qualitative differences), the vast scope of relevant outcomes measure over which to search precludes the application of a fully "rational" approach to choice.

Model-Free Reinforcement Learning

The second class of reinforcement learning algorithm, "model-free," does not correspond much at all with our ordinary experience of reasoning; perhaps as a consequence, it is a relatively new to the literature (Sutton, 1988). A model-free learner does not carry a causal model of the world. Thus, if an agent chooses to turn left from State 1, it cannot predict what the next state will be, or what reward that subsequent state will bring. This means that it cannot make farsighted decisions by comparing expected outcomes for sequences of actions, as can the model-based learner.

Instead, the model-free learner builds sparse representations of the value of each action available in a particular state. Therefore, for instance, a model-free learner that navigates through the grid world depicted in Figure 2 might associate moving south from State 1 with a value of +1, and moving west from State 1 with a value of −1. In this case it would be more likely to choose to move south than to move west. Because the model-free learner only assesses the value of the actions that are immediately available, decision making is computationally cheap. Rather than performing searches over the potentially enormous space of future actions in all combinations, it simply queries the values of each action that is immediately available.

The difficulty, of course, is to specify a learning algorithm that leads these action-based value representations to produce optimal sequences of choices. It is especially hard for early value representations to appropriately guide the agent toward a much later reward. Two tricks allow model-free algorithms to accomplish this task. The first, prediction-error learning, allows an agent to maintain a representation of the value of an action based on the value of the rewards it obtains for performing that action on average. It does this by comparing its current prediction of the value of the action (e.g., 1.3) with the actual reward subsequently obtained (e.g., 2.0), and then adjusting its value representation of the action by some fraction of the error (e.g., from 1.3 to 1.5).

Prediction-error learning has several useful properties. For instance, it allows an agent to store a representation of the average value of a choice without remembering all the past outcomes of that choice.[3] It also emphasizes the most recent observations over very distant past observations, thus efficiently responding to a changing world (Courville, Daw, & Touretzky, 2006). In addition, it can help to apportion predictive power among candidate actions or cues. If a particular action reliably predicts reward, no prediction error is generated for the reward, and this prevents credit from being assigned to additional actions or cues that happen to be correlated with (but not causative of) the reward (Rescorla & Wagner, 1965).

These useful properties make prediction-error signals indispensable to model-free learning algorithms. For the same reasons, prediction-error signals can be a useful tool for model-based learning algorithms as well. A model-based algorithm needs to construct a causal model of the world—the state-to-state transitions brought about by actions and the reward values associated with each of those states—and prediction-error learning is an ideal solution to this problem. Thus, understanding prediction errors is necessary to understand how model-free algorithms accomplish learning, but it is not sufficient to understand how model-free and model-based mechanisms differ.

The second trick to model-free algorithms is called temporal difference reinforcement learning, TDRL, (Sutton, 1988). It solves the critically important temporal credit assignment problem: How to get a value representation for an early choice (e.g., the first move out of the starting position in grid world) to reflect the rewards obtained at a much later point in time (e.g., the last move, upon which the goal is attained). Somehow, value representations must be made to bridge temporal differences.

TDRL involves an adjustment to the prediction-error learning procedure: In essence, TDRL treats the value of an action as if it were itself a reward. For instance, if pressing a lever (an action) leads to eating food (a reward), TDRL begins to treat the action as if it were itself a reward—as if pressing the lever had an intrinsic value. This allows productive actions to string together: D is valuable because it leads to reward, for instance, but C is valuable because it leads to D (which has value), B because it leads to C, and A because it leads to B.

Consider the grid world depicted in Figure 2. The first time that the agent moves east from State 24, it obtains a reward. Thus, it assigns a positive value representation to moving east from State 24. Subsequently, suppose it happens to move east from State 23 into State 24. Although there is no reward directly available in State 24, there is an action available to which it has assigned value: moving east from State 24. Treating this value representation as if it were itself a reward, the agent associates value with moving east from State 23. This process can repeat itself indefinitely, eventually establishing a path of intrinsically rewarding actions that guides the agent efficiently through the maze. Critically, however, the agent has no knowledge that "east at 24" leads directly to reward, or that "east at 23" leads to 24. It performs these actions because of their intrinsic value representations (established by a history of reward) and not because of any association with a specific outcome. Put simply, a model-free algorithm knows that moving east at State 23 feels good, but it has no idea why.

Over time, a model-free algorithm will lead an organism to make adaptive choices, just like a model-based algorithm. Yet, the two algorithms differ in very fundamental respects. Unlike a model-based agent, a model-free agent cannot make flexible choices, in the sense that a local change to a specific value representation cannot immediately be used to adjust behaviors globally. This is impossible because the model-free learner has no representation of the causal structure that links one value representation to another. The links are forged only through trial-and-error learning, as an agent notes the value representations that follow on each of its behavioral choices. Moreover, a model-free agent cannot be goal-oriented, in the sense that the agent cannot select a particular goal to be obtained and then select the necessary sequence of actions to obtain that particular goal. Again, goal-oriented planning demands a model of the environment that a model-free learner simply lacks.

However, along with these costs comes one tremendous benefit: Model-free algorithms can be extremely computationally light. At no point does an agent's decision to act involve anything more than querying the value representation associated with each immediately available choice. This stands in contrast to the model-based algorithm, which demands a computationally intensive search over a potentially large space of available actions, states, and rewards over many successive choices.

Action- Versus Outcome-Based Value Representations

Can the distinction between model-based and model-free algorithms be equated with the distinction between action- and outcome-based value representations? This is not a common approach, and there are reasons to doubt the mapping.

In some sense, for both types of algorithm, value ultimately represents reward. Moreover, a reward function for either type of algorithm could be defined over outcomes (e.g., having food in one's belly) or over actions (e.g., shoveling food in one's mouth). After all, the reward function is a subjective mapping of events onto internal hedonic states, not an objective property of the world itself, such as biological fitness.

Yet, there is an important and very fundamental difference between the value representations involved in each system. Consider the agent depicted in Figure 2 who starts his journey across a grid world. Suppose this agent has substantial experience in the maze and is now using its value representations to guide choice. As noted above, a model-free agent will select its first action with something like the following in mind: "Here in State 1, among the actions available to me, moving south is associated with the highest value," or perhaps "moving south is my preferred choice of action." In any state, the choice among available actions depends solely on the representations tied directly to those specific actions (moving south), in that particular state (State 1)—there is no "look forward" to the anticipated outcomes of the actions. (Of course, past outcomes are causally responsible for establishing the action-based value representation; the point is that the current representation itself makes no reference to particular outcomes.)

In contrast, a model-based agent has the capacity to select its first action with something like the following in mind: "A sequence of actions beginning with moving south will maximize the total amount of reward obtained in the long run because this move will lead to," followed by a specification of the full sequence of actions and their individual outcomes. Unlike the model-free agent, it can specify both the precise sequence of actions and the precise rewards that are expected. In this sense, the value representation that it uses is linked not to the immediate choice of an action but to the expected outcomes.

The contrast between these algorithms is elegantly captured by the devaluation procedure, a well-studied behavioral paradigm (e.g., Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995). A rat is trained to press a lever to obtain a food reward. During training, the rat is kept on a restricted diet to motivate performance. But then a critical test is performed: The rat is taken out of the apparatus, fed until it shows no more interest in food, and then immediately returned to the apparatus. Under some conditions, it is observed to resume pushing the lever, even though it now has no desire for food. In other words, although the food reward has been "devalued" through satiation, the habitual behavior remains intact. This apparently irrational action is easily explained by a model-free mechanism. The rat has a positive value representation associated with the action of pressing the lever in the "state" of being in the apparatus. This value representation is tied directly to the performance of the action, without any model linking it to a particular outcome. The rat does not press the lever expecting food; rather, it simply rates lever pressing as the behavioral choice with the highest value. A model-based algorithm, in contrast, has the capacity to recognize that the specific outcome associated with pressing the lever is food. Thus, because the rat does not desire food, it would place little value on the action of pressing the lever. In fact, under some conditions, rats' behaviors are more consistent with this alternative possibility. This suggests that rats, like humans, have cognitive mechanisms of both types.

In summary, the functional role of value representation in a model-free system is to select actions without any knowledge of their actual consequences, whereas the functional role of value representation in a model-based system is to select actions precisely in virtue of their expected consequences. This is the sense in which modern theories of learning and decision making rest on a distinction between action- and outcome-based value representations.

Neural and Psychological Correlates

The distinction between model-based and model-free algorithms first identified in the machine learning literature now assumes a very large profile in neuroscience and psychology (reviewed in Daw & Shohamy, 2008; Dayan & Niv, 2008). This is largely due to the discovery of neural signatures of a model-free learning algorithm encoded by dopaminergic neurons in the midbrain and their targets and associated circuits in the basal ganglia (Houk, Adams, & Barto, 1995; Montague, Dayan, & Sejnowski, 1996; Schultz et al., 1997).

Early experiments indicate model-free learning in the dopamine reward system recorded from neurons in the midbrain of the rhesus macaque (Fiorillo, Tobler, & Schultz, 2003; Schultz et al., 1997; Schultz & Dickinson, 2000). The monkeys were given rewards in the form of juice and these rewards were predicted by preceding visual cues. Initially, dopaminergic neurons fired when the juice was obtained, but not when the cues were presented; however, consistent with the operation of the TDRL algorithm, these neurons eventually stopped firing when the juice was perfectly predicted by a preceding cue (indicating the operation of a prediction-error signal), and instead fired when the cue itself was presented (indicating the operation of temporal difference learning; see Figure 3). In other words, the predictive cue began to operate as if it had intrinsic reward value. Subsequent experiments demonstrated that this pattern of cell firing can "migrate" back further to a cue that predicts reward, and so on, as specified by TDRL (e.g., Seymour et al., 2004). Similar studies have been conducted in humans using functional magnetic resonance imaging (fMRI; for example, McClure, Berns, & Montague, 2003; J. P. O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003), and these consistently indicate analogous neural signals that are predicted by the operation of model-free algorithms for reward prediction.

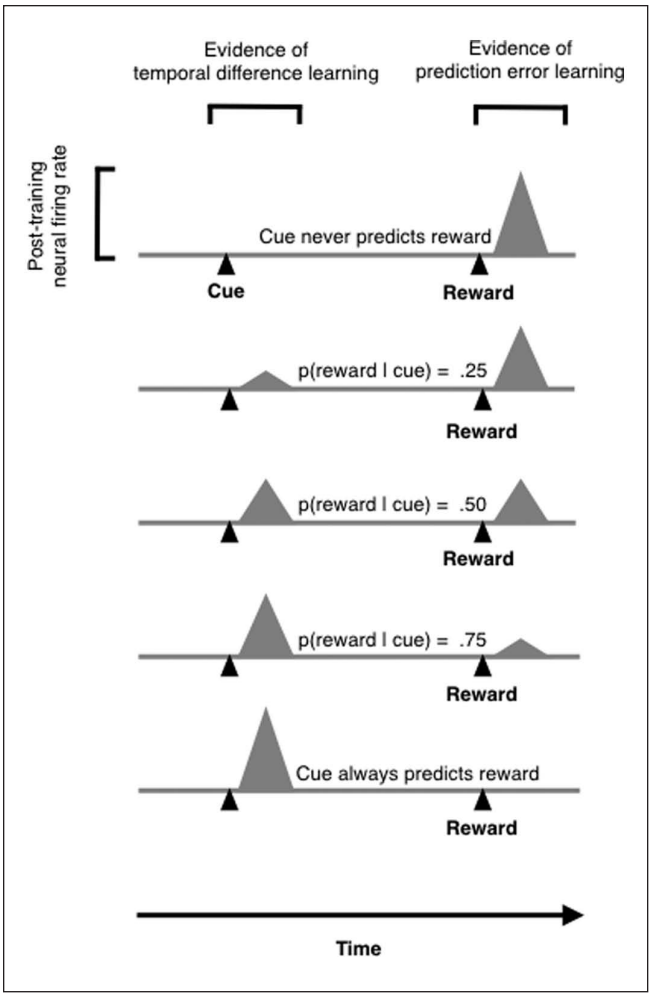

Figure 3. Idealized firing rates of midbrain dopamine neurons in response to reward and to cues that predict reward with differing probabilities.
Source. Adapted from Daw and Shohamy (2008).

Most of these studies cannot be characterized directly in terms of action-based versus outcome-based value representation because they involve classic rather than instrumental conditioning. In studies of classic conditioning, the experimental subject is not required to perform any action at all and the target of the value representation is instead a predictive cue. However, a number of other studies use instrumental rather than classic paradigms, which confirm the operation of action-based, model-free learning that relies on similar neural mechanisms (Bayer & Glimcher, 2005; Morris, Nevet, Arkadir, Vaadia, & Bergman, 2006; Roesch, Calu, & Schoenbaum, 2007). Moreover, recordings from macaque striatum identify neurons that encode the distinct value representations associated with individual actions (as opposed to their aggregate effect, observable in fMRI; Samejima, Ueda, Doya, & Kimura, 2005). Finally, optogenetic techniques have been used to directly invigorate dopaminergic pathways in the rodent brain following the performance of an action, which is observed to increase the subsequent performance of that action in the same state, just as would be expected if the dopaminergic signal were used to increase the value directly associated with that action by TDRL (Kravitz, Tye, & Kreitzer, 2012).

Although most studies of the neural basis of reinforcement learning have rewarded people with money or juice, equivalent patterns are obtained when using social rewards and punishments, including facial expressions of emotion (Lin, Adolphs, & Rangel, 2012), the comparison of the self to others (Fliessbach et al., 2007), social approval for charitable giving (Izuma, Saito, & Sadato, 2010), the reward of seeing antagonists or competitors suffer (Takahashi et al., 2009), and so on. Thus, there is every reason to expect these mechanisms of learning and decision making to operate in the social and moral domains.

A major focus of recent research has been to individuate the component processes in the human brain that accomplish learning and decision making by applying the broad framework specified by the model-based and model-free reinforcement learning mechanisms. Neural signatures of model-free mechanisms are most consistently associated with dopaminergic neurons in the midbrain and their targets in the basal ganglia (Houk et al., 1995; McClure et al., 2003; Montague et al., 1996; J. P. O'Doherty et al., 2003; Schultz et al., 1997). Meanwhile, neural signatures of model-based mechanisms are often associated with lateral prefrontal cortex (Glascher, Daw, Dayan, & O'Doherty, 2010; Li, Delgado, & Phelps, 2011).

Interactions Between Model-Based and Model-Free Systems

The emerging picture is not, however, as simple as "two independent systems: model-based and model-free." For one thing, each of the "systems" clearly comprises multiple dissociable components. For instance, independent regions of the striatum are responsible for classic and instrumental learning (J. O'Doherty et al., 2004; Yin, Knowlton, & Balleine, 2004; Yin, Ostlund, & Balleine, 2008). In addition, decision making in the model-free system appears to involve independent and opponent "go" and "no-go" processes (Cools, Robinson, & Sahakian, 2007; Crockett, Clark, & Robbins, 2009; Frank, Seeberger, & O'reilly, 2004; Guitart-Masip et al., 2011; Kravitz et al., 2012). More critical to the current discussion is the evidence that model-free and model-based systems interact closely. The successful operation of each system largely depends on interactions with the other (Dayan, 2012). Three specific examples help to illuminate this codependent relationship and also play an important role in extending the reinforcement learning framework to the moral domain.

First, representations of potential but unrealized rewards (sometimes called "fictive rewards") generated by a causal model can be used to establish the action-based value

representations used by the model-free system. For example, suppose an individual chooses action A and receives no reward but is then informed that B would have provided the reward had it been chosen. Striatal reward prediction-error signals track the values of A (the real reward) and B (the fictive reward) in these paradigms, and the magnitude of the neural response to each predicts subsequent choice behavior (Lohrenz, McCabe, Camerer, & Montague, 2007). These fictive reward signals must originate from outcome-based value representations derived counterfactually—that is, from a causal model—and yet they drive subsequent model-free response. Similarly, social learning depends on "observational" and "instructed" knowledge—that is, on information about rewards and punishment derived from watching other people or listening to their advice. Here, too, evidence suggests that cortical representations embedded within a casual model relating action, outcome, and reward are used to train action-based value representations (Biele, Rieskamp, Krugel, & Heekeren, 2011; Doll, Hutchison, & Frank, 2011; Doll, Jacobs, Sanfey, & Frank, 2009; Olsson, Nearing, & Phelps, 2007).

A second form of interaction between the two systems works in the opposite direction: Model-free value representations guide mechanisms central to model-based reasoning processes. For instance, model-free circuitry in basal ganglia appears to gate the flow of abstract representations into and out of working memory (Frank, Loughry, & O'Reilly, 2001; Gruber, Dayan, Gutkin, & Solla, 2006; O'Reilly & Frank, 2006). The maintenance of information and rules afforded by working memory systems is a hallmark of controlled cognition (Miller & Cohen, 2001). However, some processes must be responsible for recognizing which representations should be introduced into working memory, when they should be introduced, and when they should subsequently be removed. For instance, it is helpful to maintain the rule "suppress emotions" when playing poker with friends, but to remove it from working memory when the poker game is over and jokes are shared over beers. Evidence suggests that model-free mechanisms operating in basal ganglia perform this gating function by associating value with the action of introducing or removing information contents from working memory given a particular state (e.g., playing poker vs. sharing jokes; Cools, Sheridan, Jacobs, & D'Esposito, 2007; McNab & Klingberg, 2007). Thus, model-free learning mechanisms are not restricted to implementing low-level motor routines; to the contrary, they appear to play a pervasive role in regulating the use of arbitrarily complex cognitive abstractions. This concept, too, is critical to the moral domain: The "actions" that are valued by a model-free system include not just overt bodily movements but also the internal manipulation of concepts, rules, and abstract representations (Dayan, 2012).

Third, model-based and model-free mechanisms interact in planning and executing hierarchically organized behaviors. The hierarchical organization of information—and especially of behaviors into superordinate and subordinate routines or goals—is a recurrent and fundamental theme in psychology (Chomsky, 1957; Lashley, 1951; G. A. Miller, 1956) and machine learning (Barto & Mahadevan, 2003; Dietterich, 2000; Parr & Russell, 1998). Even a moderately complex task such as making a sandwich involves a nested sequence of hierarchically dependent goal-directed actions: putting cheese on the bread within making the sandwich, obtaining cheese within putting cheese on the bread, opening the refrigerator within obtaining cheese, and so forth. A growing family of hierarchical reinforcement learning models specify several related approaches to this problem (Barto & Mahadevan, 2003; Bornstein & Daw, 2011; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Dietterich, 2000; Frank & Badre, 2012; Parr & Russell, 1998; Ribas-Fernandes et al., 2011). In essence, each of them allows a subgoal to occupy the role of an "action" and a superordinate goal to occupy the role of a "state," and then leverages standard TDRL to select appropriate subgoal routines given a particular superordinate goal state. Thus, for instance, a model-free system could learn that in the state "goal: make sandwich," the action "select goal: obtain cheese" is associated with reward. Recent evidence suggests that is may be accomplished by looped corticostriatal circuits that build hierarchical levels of representation descending along a rostrocaudal gradient in prefrontal cortex (Badre & Frank, 2012; Frank & Badre, 2012).

In these cases, model-free mechanisms perform the function of helping a model-based planner to select appropriate subgoals without exhaustively searching the space of all available subgoals and computing their expected results. It accomplishes this by treating the superordinate goal as a "state" (analogous to Position 6 on grid world) and then treating the selection of a subordinate goal as an "action" (analogous to moving south). This trick—treating internal mental representations and states and actions—affords model-free mechanisms a potentially fundamental role in facilitating the complex thinking processes orchestrated by a model-based system (Bornstein & Daw, 2011; Dayan, 2012; Graybiel, 2008).

In each of these cases, model-based and model-free systems appear to interact. Research on these topics is still very much at an early stage, but a core theme is that model-free systems operate over arbitrarily abstract cognitive units, and thus facilitate model-based planning. Because model-free algorithms are well-equipped to learn and regulate low-level, habitual motor responses (e.g., typing words) and model-based algorithms are well-equipped to learn and regulate high-level, controlled planning (e.g., writing a manuscript), it can be tempting to map each category of algorithm exclusively to more "low-level" or "high-level" behaviors and representations. This mapping is flawed. Rather, mechanisms of cognitive control, goal-oriented planning, and model-based decision making operate in part by making abstract representations available for valuation and selection by model-free systems.

The lesson is not that model-free mechanisms are "smarter" than we thought. They remain just as dumb, executing actions in particular states given a past history of reward. Rather, the point is that these dumb mechanisms actually play an important role in facilitating smart behaviors. Thus, when we observe goal-oriented actions or highly abstract, productive thought processes, we have a good reason to believe that model-free mechanisms alone cannot be responsible, but that they play a contributory role.

Drawing on evidence that demonstrates a codependent relationship between automatic and controlled processes in the human brain, it has been argued that a dual-system framework for decision making generally—and moral decision making in particular—is both inaccurate and counterproductive (Kvaran & Sanfey, 2010; Moll et al., 2008; Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005; Nucci & Turiel, 1993). Often, this skepticism is rightly directed at the distinction between "cognitive" and "emotional" systems of decision making; as we have seen, this instantiation of a dual-system framework suffers from some conceptual and empirical difficulties. But does the evidence for integration between model-free and model-based mechanisms argue against any dual-system model of human decision making?

It is telling that the very research that best demonstrates these interactions between systems also demands the distinction between systems in the first place. Consider an analogy to politics. There is no sharp dividing line between Republican and Democrat. Multiple subdivisions exist within each group, and compromise, collaboration, and areas of ideological agreement exist between the two groups. So, on one hand, it would overstate the case to claim that the American political landscape is characterized by exactly two well-defined ideologies that interact exclusively in competition. But, on the other hand, without distinguishing between the two political parties it would be impossible to understand the American politics at all. So it goes with dual-system theories of learning and decision making. As illustrated by the case studies described above, only through a coarse division between two systems can the more fine-grained matters of subdivision and integration be understood.

Applying Reinforcement Learning Models to the Moral Domain

The distinction between model-based and model-free reinforcement learning provides a promising foundation on which to build a dual-process model of moral judgment. First, it accounts for the distinction between action- and outcome-based value representations. Second, it aligns this action/outcome distinction with mechanisms for automatic versus controlled processing. Third, it specifies precisely how cognitive and affective mechanisms contribute to both types of process.

Trolley Cases and "Personal" Harms

Can the distinction between model-free and model-based learning mechanisms account for the trolley problem? Specifically, can it explain why conflict between systems is engendered in the push case, but not in the switch case, typically leading people to judge the cases differently? It is easy to see how a model-based system could support the characteristically utilitarian response to both cases, favoring the lives of five individuals over the life of one individual. The relevant model of the trolley problem is small and easily explored: One branch involves five deaths, the other just a single death. Of course, it is not necessary that a model-based system disvalue death above all; it could be apathetic about others' lives, or could value nearby and faraway lives differently, or apply a steep temporal discounting function that places much value on the single life lost immediately and little value on five lives lost a few moments later. It could also value social reputation, abiding by the law, and so forth, more than lives. But if we assume that the lives matter the most, and each matters roughly equally compared with the others, a utilitarian resolution to the trolley problem looks likely.

In contrast, a model-free system does not have access to a representation of the likely outcome of each action as specified in the scenario. Instead, it depends on a value representation of the action itself, and this value representation reflects the past history of outcomes that have followed from that action. A model-free system might assign negative value to "pushing," for instance, because it typically lead to negative outcomes such as harm to the victim, punishment to the perpetrator, and so on. That is, most of the time that a person has personally pushed another (e.g., on the playground) or has witnessed one person push another (e.g., in a movie), this action lead to negative consequences. Meanwhile, a model-free system might not assign much value at all to flipping a switch because it does not typically lead to a negative outcome. In essence, the distinction between the push and switch variants of the trolley problem is proposed to be that pushing a person is a typical kind of moral violation, and thus carries an action-based negative value representation, while flipping a switch is an atypical kind of moral violation, and thus does not.

Past theories of the distinction between push and switch-flipping have emphasized physical contact between the perpetrator and the victim (Cushman et al., 2006) and the direct transfer of bodily force (Greene et al., 2009) as key variables. Insofar as direct physical contact and the transfer of bodily force are typical features of moral violations, these past proposals are easily reconciled with the current one. However, a key prediction of the current proposal is that typically harmful acts (e.g., pushing a person with your hands) will be considered morally worse than atypically harmful acts (e.g., pushing a person with your buttocks), even when the degree of physical contact and direct transfer of bodily force are equated. This is an important area for future research.

A similar analysis applies to experimental evidence that people are averse to performing pretend harmful actions, such as hitting a person's "leg" with a hammer when it is known that the "leg" is really a PVC pipe worn beneath the pants (Cushman et al., 2012). A model-free system will assess the action in terms of the value representation typically associated with hitting an apparent body part with a hard object, thus providing a natural explanation for the aversion to pretend harmful actions (or, for that matter, pretend disgusting ones; Rozin et al., 1986).

This analysis highlights an important feature of model-free reinforcement learning that is obscured by its characterization as "action-based": It represents the value of actions contingent on the context in which the action is performed. Thus, for instance, swinging a hammer at a body can carry a negative value representation, while swinging a hammer at a nail carries a positive value representation. In some sense the action is identical across the cases, but the context differs. Sensitivity to context (or "state") is critical for model-free reinforcement learning to work—the smiley learner on the grid depicted in Figure 2 clearly cannot learn "that east is good" as a general rule, but rather must learn that east is good in specific states. Applying this logic to the case of the aversion to harmful action, information about the target of a hammer (shin vs. nail) comprises an integral part of the value representation (shin/hammer = bad vs. nail/hammer = good).

Blair and colleagues have developed a similar theoretical model of the cognitive and neurobiological basis of psychopathy (Blair, 1995; Blair, 2007). They propose that distress cues—crying, expressions of pain, or fear, and so on—constitute an unconditioned aversive stimulus and motivate disengagement from aggressive actions. Over the course of development, normal individuals form a conditioned aversion to actions that typically lead to distress, such as hitting, insulting, deceit, and so on. Thus, normal individuals exhibit a basic, affective aversion to hitting, insulting, and deceit. Substantial evidence points to a pair of deficits in psychopaths that compromise this learning process: First, they fail to exhibit the ordinary aversive response to distress cues (Blair, Colledge, Murray, & Mitchell, 2001; Blair, Jones, Clark, & Smith, 1997) and second, they have difficulty learning to avoid actions that lead to aversive states (Blair et al., 2004; Blair, Morton, Leonard, & Blair, 2006)—a deficit that is not specific to the moral domain but rather affects their capacity for decision making generally.

Harm as Means Versus Side-Effect

A second dimension commonly implicated in moral dilemmas is the distinction between harm brought about as a means and harm brought about as a side-effect (Cushman & Young, 2011; Cushman et al., 2006; Hauser et al., 2007; Mikhail, 2000; Royzman & Baron, 2002). Philosophers refer to this moral distinction as the "doctrine of double effect" (Foot, 1967; Kamm, 1998; McIntyre, 2004; Thomson, 1985), and it appears to account for approximately half of the difference in judgment between the push and switch variants of the trolley problem (Greene et al., 2009). In the push variant, the victim is used as a "trolley stopper"—a tool and one that forms a necessary part of the plan for action to stop the train. Without using one person as a means of stopping the train, the other five are doomed. In the switch variant, however, the victim is merely a collateral damage—a side-effect of diverting the train but not a means of stopping it (Figure 1). This dimension of means versus side-effect can be isolated from motor properties of pushing a person versus flipping a switch, which continues to exert a substantial influence on moral judgment (Cushman et al., 2006; Greene et al., 2009; Hauser et al., 2007). This influence demands explanation.

Philosophers and psychologists have noted that harming as a means to an end demands that a person represent the subordinate goal of causing harm (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Cushman & Young, 2011; Cushman et al., 2006; Foot, 1967; Greene et al., 2009; Mikhail, 2000; Thomson, 1985). In other words, to accomplish the superordinate goal of stopping the train, the person must represent and pursue the subordinate goal of harming a person by putting them in the train's path. This is not the case for harming as a side-effect: The agent does not have a subordinate goal that the person on the side track be hit by the train. Consequently, a system that represents action plans in terms of their hierarchical structure must necessarily represent switch-flipping in terms of "harming a person" in the means case, whereas it can merely represent switch-flipping in terms of "diverting a train" in the side-effect case.

Yet, on its face, it's not clear why this feature should matter to moral judgment. What difference does it make whether you are using a person's death as a means or causing it as a side-effect? Surely it doesn't matter to the victim! In both the cases, you know you are killing a person and do so out of concern for others rather than malice.

One possible explanation for the means/side-effect distinction derives from the model of hierarchical reinforcement learning that we considered above. For instance, a model-free mechanism might assign positive value to the action "set subgoal: get milk" in the state "goal: make coffee." It does not represent how the subgoal supports the goal, or what the subgoal will accomplish, but rather values selecting this subgoal because doing so has been rewarding in past similar states. Turning to moral domain, consider cases where harm is used as a means to an end. This requires the cognitive action "select subgoal: harm a person." A model-free system will associate the execution of the subgoal with subsequent rewards or punishments. Generally, executing "select subgoal: harm a person" leads to aversive outcomes such as victim distress, reprimand, and so forth. Thus, a model-free system will tend to associate negative value with executing subgoals of the form "harm a person." By contrast, it will tend not to associate negative value with executing

subgoals of the form "divert a train" or, more abstractly, "divert a threat," "save several lives," and so on, because these subgoals are not typically associated with aversive outcomes.

In the "push" variant of the trolley problem, then, a model-based system constructs the subgoal "harm a person" and endorses it based on an overall representation of the expected value of executing it (a net gain of four lives). By contrast, a model-free system will evaluate the subgoal "harm a person" as negative and disprefer this option. In colloquial terms, the model-based system reasons "that harming a person achieves a necessary part of the plan to ultimately save five others," while the model-free system reasons that "when the subgoal 'harm a person' gets executed, things tend to turn out poorly."

Active Versus Passive Harm

A third dimension commonly implicated in moral dilemmas is the distinction between action and omission. Specifically, people tend to consider it morally worse to harm a person actively (e.g., by administering a fatal poison) than to passively allow them to die (e.g., by deliberately withholding a life-saving antidote; Baron & Ritov, 2004, 2009; Cushman et al., 2006; Spranca et al., 1991). Such an effect is explained by the fact, noted above, that model-free value representations preferentially encode the value associated with the available actions in a state (either promoting or inhibiting those actions) but not with the ever-available option to omit action.

Specifically, in the basal ganglia, behaviors are regulated by opponent "go" and "no-go" processes that promote and inhibit actions, respectively (Frank et al., 2004; Guitart-Masip et al., 2011; Hikida, Kimura, Wada, Funabiki, & Nakanishi, 2010). Thus, if an action is consistently punished in a particular state, the model-free value representation preferentially encodes "negative value for action" rather than "positive value for inaction." Conversely, if an action is consistently rewarded, the model-free value representation preferentially encodes "positive value for action," rather than "negative value for inaction." This property may account in part for the asymmetry in moral evaluation between actions and omissions: An action can carry a forceful negative valuation signal via a model-free system while an omission cannot. By contrast, the capacity to fully evaluate the likely consequences of an action afforded by a model-based system would allow the value of actions and omissions to be represented.

Recent evidence from functional neuroimaging demonstrates a positive correlation between activation in the frontoparietal control network—a set of brain regions implicated in cognitive control—and the condemnation of harmful omissions (Cushman et al., 2012). No such relationship is evident for the condemnation of harmful actions. This provides some evidence that relatively automatic processes are sufficient to condemn harmful actions, while controlled processes play a necessary role in fully condemning harmful omissions. This finding is consistent with the proposal that model-free systems provide a negative value representation for harmful actions, while model-based systems are required to represent the negative value associated with harmful omissions.

Observational Learning

If model-free reinforcement learning principles account for core elements of our aversion to harmful action, it is clear that they cannot rely exclusively on direct personal experience. For instance, people are averse to pushing a person off a footbridge and into the path of a train but few people have performed this action before. Perhaps the relevant representation is simply "pushing a person"—a behavior performed frequently in early childhood and extinguished as children experience the negative outcomes of punishment and the distress of their victims (Tremblay, 2000). But, consider the action of "shooting" an experimenter point-blank in the face using a weighty metal replica of a gun (Cushman et al., 2012). Qualitative observations suggest that participants found this to be among the most aversive actions performed in this study. Presumably, however, most participants in the study (principally Harvard undergraduates) had never personally shot a person.

Critically, several lines of evidence indicate that model-free value representations can be constructed on the basis of observational learning (e.g., seeing the consequences of gun violence in a movie) or instruction (e.g., being told what happens when somebody is shot point-blank in the face). In the domain of aversive learning, where, for instance, a neutral cue might be paired with a painful shock, social learning and direct experience are sufficient to produce a galvanic skin response (Olsson & Phelps, 2004) and activation in the amygdala (Hooker, Verosky, Miyakawa, Knight, & D'Esposito, 2008; Olsson et al., 2007; Olsson & Phelps, 2007) in response to the cue. In the domain of reward learning, several studies have shown striatal activations during observational experience that appear to encode prediction-error signals analogous to those obtained during direct experience (Bellebaum, Jokisch, Gizewski, Forsting, & Daum, 2012; Cooper, Dunne, Furey, & O'Doherty, 2012; Li et al., 2011); however, these appear to be attenuated in magnitude compared with direct experience. Thus, although the literature on observational reward learning is still relatively young, representations of others' actions and experiences appear to modulate reward learning in a model-free system.

Conclusion

Ordinary human moral judgments are rife with apparent inconsistency. We sometimes consider utilitarian harm wrong and other times do not. The difference often boils

down to seemingly irrelevant factors, such as sensorimotor properties such as a push.

We can understand these apparent inconsistencies, however, through the lens of model-free reinforcement learning and, specifically, through the concept of action-based value representation. Such value representations can be defined over relatively concrete features (e.g., pushing a body) and relatively abstract features, such as the selection of particular subgoal (e.g., harming a person). Critically, action-based value representations allow us to explain strong, systematic patterns of nonutilitarian choice in moral judgments and behaviors.

It is the role of action-based value representation that affords model-free algorithms a special explanatory role in the moral domain compared with alternative models of associative learning. Consider, for instance, the Rescorla-Wagner (1965) model. It is a model of associative learning and, in some respects, resembles TDRL. It is not equipped, however, to associate value with actions, or to pass value representations backward to successively earlier predictive actions or states via TDRL, or to use those value representations to guide future action. In other words, it can associate actions with outcomes but some other system must then be responsible for making decisions based on the predicted outcomes. In the context of the trolley problem, then, it would associate pushing with harm but would not directly represent the negative value of pushing. This leaves it poorly equipped to explain nonutilitarian patterns of moral judgment. After all, an association between pushing and harm is superfluous—the trolley problem specifies that harm will result from pushing and, moreover, that five times as much harm will result from not pushing. In other words, any outcome-based associative system will face a challenge explaining patterns of moral judgment that are not, themselves, outcome-based. Nonutilitarian moral judgment is best described as action-based judgment, and model-free algorithms are therefore especially useful because they posit action-based value representations.

Finally, a clear goal for any theory of moral judgment is to explain dilemmas: the conflict between distinct psychological systems in cases such as the push version of the trolley problem. While I have emphasized the role that model-free mechanisms may play in selecting nonutilitarian options, equally important is the natural explanation that model-based mechanisms provide for explaining utilitarian choice. It is the contrast between model-free and model-based systems—or between action- and outcome-based valuation—that can explain the conflict engendered by moral dilemmas.

Broader Applications to the Moral Domain

I have explored connections between reinforcement learning algorithms and moral judgments of trolley-type dilemmas in great detail, but this particular connection is of limited intrinsic interest. Ideally, it offers a case study of the utility of framing a dual-system theory of decision making in terms of action- and outcome-based value representations, highlighting the advantages of this approach over the more traditional distinction between emotion and reasoning. How well does the framework extend to further dimensions of moral domain?

Automaticity and Control

A cornerstone of current research in moral judgment is the distinction between automatic and controlled processes. The phenomenon of moral dumbfounding, discussed more fully below, reveals that people often arrive at moral judgments through automatic processes and then use controlled cognition to construct an explicit rationale (Haidt, 2001). The distinction between automatic and controlled processes is also fundamental to the research on trolley-type dilemmas. Evidence from functional neuroimaging (Cushman et al., 2012; Greene et al., 2004), cognitive load manipulations (Greene et al., 2008; Trémolière et al., 2012), timing manipulations (Suter & Hertwig, 2011), and priming (Valdesolo & DeSteno, 2006), all suggest that characteristically utilitarian (i.e., outcome-based) judgments rely relatively more on controlled processes, while characteristically deontological (i.e., action-based) judgments rely relatively more on automatic processes (but see Kahane et al., 2012).

We should therefore expect model-free mechanisms to be relatively more automatic and model-based mechanisms to be relatively more controlled, and indeed, this prediction is borne out. Most notably, the basal ganglia circuits characterized by model-free TDRL support habitual (i.e., automatic, or default) behavioral response (reviewed in Graybiel, 1998). A recent study used a behavioral paradigm specifically designed to dissociate model-based from model-free response and found a shift toward model-free response under cognitive load (Otto, Gershman, Markman, & Daw, 2013).

Rationalization

In addition to the trolley problem, recent studies of moral psychology have orbited around a second case with tremendous gravitational pull: that of Julie and Mark, adventurous siblings who try out sexual intercourse with each other for fun. They do it once, consensually, secretly, nonprocreatively, and quite passionately. The key finding is that many people consider it wrong but few can say precisely why, a phenomenon termed *moral dumbfounding* (Bjorklund, Haidt, & Murphy, 2000; Haidt & Hersh, 2001; Haidt et al., 1993). Dumbfounding has been demonstrated with a host of different cases such as eating a dead family pet or burning the national flag, and also in some versions of the trolley problem (Cushman et al., 2006).

Notably, the cases used to elicit dumbfounding typically involve ostensibly harmless actions such as incest. This is no

accident; when a harmful outcome occurs, people readily point toward the harm as the basis of their moral verdict. Studies on moral dumbfounding show that before people give up on explaining why sibling incest is wrong they often attempt to explain their judgment precisely in terms of harmful outcomes, invoking the potential for birth defects, regret, family shame, and so forth (Bjorklund et al., 2000; Haidt & Hersh, 2001). In fact, this reflects a general property of the relationship between moral judgment and moral justification (Ditto & Liu, 2011). Specifically, we often consider actions morally right or wrong for reasons other than the outcomes they produce, but then rationalize justifications for those moral judgments that appeal to outcomes.

A potential explanation for this puzzling mismatch is that judgments often depend on model-free mechanisms that directly value actions, while justifications are produced by mechanisms of controlled reasoning that operate in a model-based manner. Consequently, the process of justification may entail a search for outcome information that supports an action-based moral aversion.

Some caution is warranted in understanding the incest taboo, in particular, as a product of model-free learning, however. While it does appear to depend on action-based value representation, it probably does not derive model-free reinforcement learning—at least, not exclusively. Rather, evidence suggests that it depends, at least in part, on a biologically evolved mechanism for kin detection that is sensitive to early childhood cohabitation as well as shared maternity (Lieberman & Lobel, 2012; Lieberman, Tooby, & Cosmides, 2003; Lieberman, Tooby, & Cosmides, 2007). This highlights an important limitation of the current proposal: It attempts to account for some automatic, action-based value representations in the moral domain by appeal to model-free learning, but there are surely many action-based moral aversions that are not derived from model-free learning mechanisms.

Moral Philosophy

Greene (2007) has proposed that emotionally-grounded intuitions associated with nonutilitarian moral judgment play a critical role in explaining the origins of deontological philosophical theories—those concerned with rights, categorical obligations, and prohibitions, and associated most famously with the philosophical contributions of Kant. Meanwhile, he proposes that processes of cognitive control and reasoning associated with utilitarian moral judgment play a critical role in explaining the origins of consequentialist philosophical theories—those concerned with maximizing welfare and associated most famously with the philosophical contributions of Bentham and Mill.

The mapping of a dual-system psychological framework onto discrete philosophical positions has considerable appeal, yet the emotion/deontology and reasoning/utilitarianism linkages have been met with some skepticism, especially among philosophers (e.g., Kahane et al., 2012). Kant's deontological moral theory, for instance, places an extreme premium on the derivation of moral principles from processes of reasoning. Meanwhile, utilitarian moral theories are grounded in sentimental concern for others' welfare.

The alternative linking of deontological theories with action-based value representation and utilitarian moral theories with outcome-based value representation is truer to the core philosophical positions. Whether derived from emotional or rational processes, Kant's moral theory undeniably elevates the intrinsic moral quality of actions above the expected consequences of those actions. For instance, he advocated that the categorical prohibition against lying would make it immoral to deceive a murderer about the location of his would-be victim (Kant, 1785/1983). Likewise, whether derived from emotional or rational processes, consequentialist moral theories clearly operate by maximizing expected value over a causal model of the likely consequences of action.

Associating characteristically deontological judgments with model-free reinforcement learning helps to illustrate their functional rationale: They are an efficient compression of past experience into a simple representation of a policy that tends to maximize reward. Through this lens, model-free value representations might also be associated with rule utilitarianism (choosing the set of efficient and simple rules that tend to maximize welfare). This highlights the important sense in which deontological rules—and action-based value representations—can be normatively justified (see also Bennis, Medin, & Bartels, 2010).

Sacred Values

Intense focus on a single case of nonutilitarian choice—the trolley problem—has the ironic effect of obscuring just how deeply nonutilitarian choice pervades moral decision making (Baron, 1994). Moral norms frequently place categorical prohibitions on action, specifically proscribing the possibility of engaging in utility-maximizing tradeoffs. Such norms have been called "sacred" or "protected" values (Baron & Spranca, 1997; Fiske & Tetlock, 2000; Ritov & Baron, 1999; Tetlock, 2003). For instance, most people consider it categorically wrong to buy or sell organs on an open market, although this may facilitate the efficient distribution of resources. Political liberals sometimes regard certain forms of environment destruction or the disparate treatment of racial groups to be categorically prohibited no matter what the net benefits, while political conservatives sometimes exhibit equivalent patterns of judgment regarding abortion and homosexuality. As George W. Bush wrote in an op-ed column in the *New York Times* justifying his prohibition of embryonic stem cell research, "There is at least one bright line: We do not end some lives for the medical benefit of others."

Sacred values are complex and multifaceted phenomena, implicating concepts of purity and profanity, routines of cleansing, and a division between distinct relational

schemata (Fiske, 2004; Fiske & Tetlock, 2000). It would be an error to suggest that they can be fully captured within a simple computational framework. Nevertheless, a core feature of sacred/protected values is the taboo against tradeoffs—in other words, a categorical prohibition of some category of action that maintains insensitivity to the ultimate outcome of that action (Baron & Ritov, 2009; Bartels & Medin, 2007). As such, a key contributor to sacred/protected values as a psychological kind may be action-based value representation. A particularly intriguing connection may be drawn between the proscription of impure thoughts (Tetlock, 2003) and the role of model-free systems in gating information held in working memory (O'Reilly & Frank, 2006).

More broadly, the landscape of moral obligation and prohibition is abundantly populated with rules defined over action types. Sexual taboos against incest, masturbation, sodomy, intercourse during menstruation, and so forth are predicated on the type of action performed rather than the expected outcome of that action. The same is true of many taboos concerning consumption and cleanliness (rules defining what is kosher, halal, etc.). Rituals (taking communion, doing a team cheer before a game, etc.) and rules of etiquette (eating with the fork in the right hand, not cursing, etc.) are also commonly defined over actions rather than outcomes, and it has been suggested that these may rely on model-free value representation (Graybiel, 2008).

Norms of charitable giving present a more ambiguous case. Is a 10% tithe to the church conceptualized principally as an instrumental action aimed at the betterment of the institution, or more simply as a mandatory norm of action? What about the 2-year commitment to missionary work in the Mormon faith, or secular norms such as bringing a bottle of wine to a dinner party, or purchasing a birthday gift? Such behaviors may feel compulsory because of our interest in the welfare of nonbelievers, dinner hosts, and birthday boys; alternatively, they may feel compulsory because processes of social learning assign value directly to the actions themselves. Consistent with this possibility, research finds that among more religious individuals, compassion plays less of a role in explaining variance in prosocial behavior (Saslow et al., 2012); this may be because rituals, rules, and social pressures of religions establish value in prosocial actions intrinsically rather than deriving value from prosocial outcomes.

Each of the cases described above involves an apparent commitment to explicit norms, or deontological rules. One interpretation of these rules is that they are nearly always posthoc rationalizations constructed in response to affectively laden intuitions (Cushman & Greene, 2011; Greene, 2007; Haidt, 2001) or, in the terms pursued here, the value assigned intrinsically to certain actions. An alternative possibility is also quite consistent with the reinforcement learning perspective but accords rules in a causal role in moral judgments. As noted above, many current models accord model-free mechanisms in basal ganglia a role in selecting the contents of working memory—that is, a cognitive action rather than a motor action. Thus, the selection and execution of explicitly represented rules may depend on model-free value representations, explaining the association of canonically "controlled" mechanisms with characteristically deontological moral judgment.

Conclusion

Dual-system theories are widely used in the moral domain, yet there is pervasive disagreement about the nature of the two systems. I have noted a few of the most common contrasts offered in the decision-making literature: automatic versus controlled processes, intuitive versus rational processes, and emotional versus cognitive processes. Dissatisfied with these terms, many have taken to referring simply to System 1 and System 2 (Stanovich & West, 2000); this opaque nomenclature is a faithful reflection of murky theoretical waters. Some form of dual-system theory is indispensable for explaining core features of moral judgments. These include the nature of moral dilemmas, as well as the relationship between moral judgment and justification. The automatic/controlled and intuitive/rational distinctions are appropriate at a descriptive level but lack explanatory precision. By analogy, a subway operates automatically while a bicycle requires manual operation; yet, these superficial descriptions fail to explain the actual mechanics of either mode of transportation. Meanwhile, the emotion/cognition distinction can be misleading because both systems involve information processing and value representation.

A more precise characterization of the two systems within the moral domain distinguishes between mechanisms of value representation: one that assigns value directly to actions, and another that selects actions based on the value assigned to their likely outcomes. This same distinction captures an essential difference between the two families of reinforcement learning algorithm. The basic principles of these reinforcement learning algorithms—and specific details of their neural implementation—provide an explanation for several otherwise puzzling phenomena of moral judgments, and of human judgment and decision making more broadly. Their explanatory power becomes especially broad when conceiving of certain internal mental representations as constituting "states" and certain processes of thought as constituting "actions."

At the same time, clear limitations of this approach must be emphasized. The role of learning—and more specifically, of TDRL—is clearly limited to a subset of moral norms. For instance, evidence suggests that the aversion to sibling incest is grounded in an action-based aversion (Bjorklund et al., 2000; Haidt & Hersh, 2001; Haidt et al., 1993), and that this aversion has an innate basis (Lieberman et al., 2003). Thus, it is unlikely that TDRL contributes strongly to the aversion to incest. Innate and learned action-based aversions may or may not draw on similar psychological mechanisms; at present there is little relevant evidence (but see Delgado, Jou, & Phelps, 2011). In any

event, the mapping between action-based value representation and model-free TDRL is not perfect.

A second limitation of this approach concerns the opposite mapping between cognitive control, model-based reasoning, and the valuation of outcomes. I have emphasized the role of cognitive control in overriding model-free habits (E. K. Miller & Cohen, 2001); however, controlled cognition can be used much generally to impose rule-like structure on decision processes (Sloman, 1996). Humans have the capacity to specify action-based rules; above, I noted categorical prohibitions against lying, cursing, homosexual sex, the consumption of pork, and so on. These kinds of rules play a common and crucial role in human moral judgments and behaviors, as evidenced by their ubiquity in law, religion, custom, and etiquette. The cultural transmission of these rules depends, at least partially, on their explicit representation and communication using controlled cognitive processes; presumably, the application of the rules in specific circumstances often does as well. Thus, action-based rules present an apparent case of action-based valuation operating via characteristically model-based mechanisms as follows: namely, working memory and executive function. This apparent conflict can be reconciled in at least two ways. First, as with implementation intentions, the power of the action-based rules may lie precisely in their ability to be assimilated into an automatic system supported by model-free value representations. Second, as noted above, the engagement of rule-based processing may be due to model-free mechanisms valuing the cognitive "action" of implementing the rule.

There is an inherent tension in any attempt to account for complex psychological phenomena in terms of relatively simple and computationally precise mechanisms. The more a specific mechanism is isolated and understood the less the general phenomenon is explained; yet, unless specific mechanisms are isolated and understood, the general phenomenon cannot be explained at all. This tension bedevils research in moral judgment, which necessarily draws upon multiple levels of analysis: evolutionary, cultural, historical, social, mechanistic, and neural among them. Yet, philosophers (Hume, 1739/1978; Kant, 1785/1959) and psychologists (Greene, 2007; Haidt, 2001) alike have often concluded that moral psychology cannot be understood with a broad division between the two systems of decision making, and the attraction of the emotion/reason division has repeatedly proved irresistible. Current research in computational neuroscience can improve on this old division by formalizing the two targets of value representation and the systems that support them: action and outcome.

Acknowledgment

Thanks to Nathanial Daw, Michael Frank, Tamar Gendler, Sam Gershman, Joshua Greene, Ryan Miller, Steven Sloman, and Liane Young for helpful discussion and feedback on this work.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by a grant from the Office of Naval Research, N00014-13-1-0281.

Notes

1. There are, however, a number of contexts in which this distinction between active and passive harm is not observed (see, for example, Mandel & Vartanian, 2008; Patt & Zeckhauser, 2000).
2. It is important to recognize that emotion, affect, and value representations are not identical psychological constructs. One dimension of the current proposal is to propose that the motivational response identified as "emotional" in prior literature is in fact better characterized as a form of negative value representation.
3. For example, imagine that a particular choice leads to a reward of 0 half the time, and a reward of 1 half the time. An agent could store the full past history of reward values (0, 1, 0, 0, 1, 1, 1, etc.) but this would require a large amount of memory and computation. Or, it could just remember the very last reward obtained, but this will lead to overestimate the value half the time (at 1) and to underestimate the value half the time (at 0). Using a prediction-error mechanism, the agent will tend to have a value representation around 0.5 at any given time (because of the combined effects of small adjustments upward or downward), and yet does not have to remember the full history of the past choices.

References

Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in Cortio-Striatal circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*, 527-536.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral & Brain Sciences*, *17*, 1-10.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*, 74-85.

Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. Medin (Eds.), *Moral judgment and decision making* (Vol. 50). San Diego, CA: Academic Press.

Baron, J., & Spranca, M. (1997). Protected values. *Virology*, *70*(1), 1-16.

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381-417.

Bartels, D., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices? *Psychological Science*, *18*, 24-28.

Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, *13*, 341-379.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 129-141.

Bellebaum, C., Jokisch, D., Gizewski, E., Forsting, M., & Daum, I. (2012). The neural coding of expected and unexpected monetary performance outcomes: Dissociations between active and observational learning. *Behavioural Brain Research*, *227*, 241-251.

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, *5*, 187-202.

Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS Biology*, *9*(6), e1001089.

Bjorklund, F., Haidt, J., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Lund Psychological Reports*, *2*, 1-23.

Blair, K., Morton, J., Leonard, A., & Blair, R. J. R. (2006). Impaired decision-making on the basis of both reward and punishment information in individuals with psychopathy. *Personality and Individual Differences*, *41*, 155-165.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, *57*, 1-29.

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*, 387-392.

Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology*, *29*, 491-498.

Blair, R. J. R., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, *34*, 192-198.

Blair, R. J. R., Mitchell, D., Leonard, A., Budhani, S., Peschardt, K., & Newman, C. (2004). Passive avoidance learning in individuals with psychopathy: Modulation by reward but not by punishment. *Personality and Individual Differences*, *37*, 1179-1192.

Borg, J. S., Hynes, C., Van Horn, J., Grafton, S. T., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, *18*, 803-837.

Bornstein, A. M., & Daw, N. D. (2011). Multiplicity of control in the basal ganglia: Computational roles of striatal subregions. *Current Opinion in Neurobiology*, *21*, 374-380.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652.

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

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive Affective Neuroscience*, *2*, 84-92.

Conway, P., & Gawronski, B. (2013). Deontological versus utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235.

Cools, R., Robinson, O. J., & Sahakian, B. (2007). Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. *Neuropsychopharmacology*, *33*, 2291-2299.

Cools, R., Sheridan, M., Jacobs, E., & D'Esposito, M. (2007). Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory. *Journal of Neuroscience*, *27*, 5506-5514.

Cooper, J. C., Dunne, S., Furey, T., & O'Doherty, J. P. (2012). Human dorsal striatum encodes prediction errors during observational learning of instrumental actions. *Journal of Cognitive Neuroscience*, *24*, 106-118.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*, 294-300.

Crockett, M. J., Clark, L., Hauser, M., & Robbins, T. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*, 17433-17438.

Crockett, M. J., Clark, L., & Robbins, T. W. (2009). Reconciling the role of serotonin in behavioral inhibition and aversion: Acute tryptophan depletion abolishes punishment-induced inhibition in humans. *Journal of Neuroscience*, *29*, 11993-11999.

Cushman, F. A., Gray, K., Gaffey, A., & Mendes, W. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*, 2-7.

Cushman, F. A., & Greene, J. D. (2011). "The philosopher in the theater" in social psychology of morality: The origins of good and evil. In M. Mikulincer, & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 33-50). Washington, DC: APA Press.

Cushman, F. A., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, *7*, 269-279.

Cushman, F. A., 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.

Cushman, F. A., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*, 1052-1075.

Cushman, F. A., Young, L., & Greene, J. D. (2010). Multi-system moral psychology. In J. M. Doris, & T. M. P. R. Group (Eds.), *The Oxford handbook of moral psychology* (pp. 47-71). New York, NY: Oxford University Press.

Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*, 1082-1089.

Daw, N., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*, 199-204.

Daw, N., & Shohamy, D. (2008). The cognitive neuroscience of motivation and learning. *Social Cognition*, *26*, 593-620.

Dayan, P. (2012). How to set the switches on this thing. *Current Opinion in Neurobiology*, *22*, 1068-1074.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*, 185-196.

Delgado, M. R., Jou, R. L., & Phelps, E. A. (2011). Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. *Frontiers in Neuroscience*, *5*, 71.

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution & Human Behavior*, *32*, 204-215.

DeScioli, P., Christner, J., & Kurzban, R. (2011). The omission strategy. *Psychological Science*, *22*, 442-446.

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Learning & Behavior*, *23*, 197-206.

Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, *13*, 227-303.

Ditto, P. H., & Liu, B. (2011). Deontological dissonance and the consequentialist crutch. In M. Mikulincer, & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 51-70). Washington, DC: APA Press.

Doll, B. B., Hutchison, K. E., & Frank, M. J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *Journal of Neuroscience*, *31*, 6188-6198.

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research*, *1299*, 74-94.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709-724.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*, 1898-1902.

Fiske, A. P. (2004). Four modes of constituting relationships. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 57-142). Mahwah, NJ: Lawrence Erlbaum.

Fiske, A. P., & Tetlock, P. E. (2000). Taboo trade-offs: Constitutive prerequisites for political and social life. In S. A. Renshon, & J. Duckitt (Eds.), *Political psychology: Cultural and crosscultural foundations* (pp. 47-65). London, England: Macmillan.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., & Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, *318*, 1305-1308.

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

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in cortio-striatal circuits 1: Computational analysis. *Cerebral Cortex*, *22*, 509-526.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*, 137-160.

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, *306*, 1940-1943.

Glascher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*, 585-595.

Goldman, A. (1971). The individuation of action. *Journal of Philosophy*, *68*, 761-774.

Graham, J., Haidt, J., & Nosek, B. (2009). Liberals and conservatives use different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029-1046.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*, 119-136.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, *31*, 359-387.

Greene, J. D. (2007). *The secret joke of Kant's soul*. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (Vol. 3, pp. 35-80). Cambridge, MA: MIT Press.

Greene, J. D., Cushman, F. A., 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.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144-1154.

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*, 389-400.

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.

Gregg, M. E., James, J. E., Matyas, T. A., & Thorsteinsson, E. B. (1999). Hemodynamic profile of stress-induced anticipation and recovery. *International Journal of Psychophysiology*, *34*, 147-162.

Gruber, A. J., Dayan, P., Gutkin, B. S., & Solla, S. A. (2006). Dopamine modulation in the basal ganglia locks the gate to working memory. *Journal of Computational Neuroscience*, *20*, 153-166.

Guitart-Masip, M., Fuentemilla, L., Bach, D. R., Huys, Q. J. M., Dayan, P., Dolan, R. J., & Duzel, E. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *Journal of Neuroscience*, *31*, 7867-7875.

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

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, *31*, 191-221.

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

Hauser, M. D., Cushman, F. A., Young, L., Jin, R., & Mikhail, J. M. (2007). A dissociation between moral judgment and justification. *Mind & Language*, *22*, 1-21.

Hikida, T., Kimura, K., Wada, N., Funabiki, K., & Nakanishi, S. (2010). Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron*, *66*, 896-907.

Hood, B. M., Donnelly, K., Leonards, U., & Bloom, P. (2010). Implicit voodoo: Electrodermal activity reveals a susceptibility to sympathetic magic. *Journal of Cognition and Culture*, *10*, 391-399.

Hooker, C. I., Verosky, S. C., Miyakawa, A., Knight, R. T., & D'Esposito, M. (2008). The influence of personality on neural mechanisms of observational fear and reward learning. *Neuropsychologia*, *46*, 2709-2724.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict

reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249-270). Cambridge, MA: MIT Press.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*, 1-6.

Hume, D. (1978). A treatise of human nature. In L. Selby-Bigge, & P. H. Nidditch (Eds.). (Original work published 1739)

Izuma, K., Saito, D. N., & Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of Cognitive Neuroscience*, *22*, 621-631.

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*, 393-402.

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

Kamm, F. M. (1998). *Morality, mortality: Death and whom to save from it*. New York, NY: Oxford University Press.

Kant, I. (1959). *Foundations of the metaphysics of morals* (L.W. Beck, Trans.). New York, NY: Macmillan. (Original work published 1785)

Kant, I. (1983). *On a supposed right to lie because of philanthropic concerns*. Indianapolis, IN: Hackett. (Original work published 1785)

King, L. A., Burton, C. M., Hicks, J. A., & Drigotas, S. M. (2007). Ghosts, UFOs, and magic: Positive affect and the experiential system. *Journal of Personality and Social Psychology*, *92*(5), 905-919.

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.

Kravitz, A. V., Tye, L. D., & Kreitzer, A. C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature Neuroscience*, *15*, 816-818.

Kvaran, T., & Sanfey, A. G. (2010). Toward an integrated neuroscience of morality: The contribution of neuroeconomics to moral cognition. *Topics in Cognitive Science*, *2*, 579-595.

Lashley, K. S. (1951). *The problem of serial order in behavior*. New York, NY: John Wiley.

Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences*, *108*, 55-60.

Lieberman, D., & Lobel, T. (2012). Kinship on the Kibbutz: Coresidence duration predicts altruism, personal sexual aversions and moral attitudes among communally reared peers. *Evolution & Human Behavior*, *33*, 26-34

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

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, *445*, 727-731.

Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, *7*, 274-281.

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences*, *104*, 9493-9498.

Mandel, D. R., & Vartanian, O. (2008). Taboo or tragic: Effect of tradeoff type on moral choice, conflict, and confidence. *Mind & Society*, *7*, 215-226.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*, 339-346.

McIntyre, A. (2004). The double life of double effect. *Theoretical medicine and bioethics*, *25*, 61-74.

McNab, F., & Klingberg, T. (2007). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*, 103-107.

Mendes, W., Blascovich, J., Hunter, S., Lickel, B., & Jost, J. (2007). Threatened by the unexpected: Physiological responses during social interactions with expectancy-violating partners. *Journal of Personality and Social Psychology*, *92*, 698-716.

Mendez, M. F., Anderson, E., & Shapria, J. S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*, 193-197.

Mikhail, J. M. (2000). *Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in "A theory of justice"* (Doctoral dissertation). Cornell University, Ithaca, NY.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Miller, R., & Cushman, F. A. (in press). Tough for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social & Personality Psychology Compass*.

Miller, R., Hannikainan, I., & Cushman, F. (2013). *Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm*. Manuscript submitted for publication.

Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition. *Annals of the New York Academy of Sciences*, *1124*, 161-180.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*, 799-809.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936-1947.

Moore, A., Clark, B., & Kane, M. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*, 549-557.

Moretto, G., Ladavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*, 1888-1899.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience*, *9*, 1057-1063.

Nucci, L., & Turiel, E. (1993). God's word, religious rules, and their relation to Christian and Jewish children's concepts of morality. *Child Development*, *64*, 1475-1491.

Nucci, L. P., & Gingo, M. (2010). The development of moral reasoning. In U. Goswami (Ed.), *The Wiley-Blackwell handbook*

of childhood cognitive development (pp. 420-445). Malden, MA: Wiley-Blackwell.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience*, *2*, 3-11.

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science*, *15*, 822-828.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience*, *10*, 1095-1102.

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement learning systems by taxing the central executive. *Psychological Science*. Advance online publication.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*, 452-454.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*, 329-337.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*, 283-328.

Parr, R., & Russell, S. (1998). Reinforcement learning with hierarchies of machines. In M. Jordan, M. Kearns, & S. Solla (Eds.), *Advances in neural information processing systems* (Vol. 10, pp. 1043-1049). Cambridge, MA: MIT Press

Patt, A., & Zeckhauser, R. (2000). Action bias and environmental decisions. *Journal of Risk and Uncertainty*, *21*, 45-72.

Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, *2*, 511-527.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163-177.

Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychological Review*, *110*, 193-196; discussion, 197-198.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York, NY: Appleton-Century-Crofts.

Ribas-Fernandes, J. Ú. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., & Botvinick, M. M. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, *71*, 370-379.

Ritov, I. I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Process*, *79*, 79-94.

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, *10*, 1615-1624.

Royzman, E., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*, 165-184.

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, *50*, 703-712.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*, 1337-1340.

Saslow, L. R., Willer, R., Feinberg, M., Piff, P. K., Clark, K., Keltner, D., & Saturn, S. R. (2012). My brother's keeper? Compassion predicts generosity more among less religious individuals. *Social Psychological & Personality Science*. Advance online publication.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593-1599.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473-500.

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*, 664-667.

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*, 76-105.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, *23*, 645-665.

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*, 454-458.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9-44.

Sutton, R. S., & Barto, A. (1999). Reinforcement learning. *Journal of Cognitive Neuroscience*, *11*, 126-134.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., & Okubo, Y. (2009). When your gain is my pain and your pain is my gain: Neural correlates of envy and schadenfreude. *Science*, *323*, 937-939.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Science*, *7*, 320-324.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*, 1395-1415.

Tremblay, R. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development*, *24*, 129-141.

Trémolière, B., De Neys, W., & Bonnefon, J.-F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*, 379-384.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*, 476-477.

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.

Yin, H. H., Knowlton, B., & Balleine, B. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*, 181-189.

Yin, H. H., Ostlund, S. B., & Balleine, B. W. (2008). Reward-guided learning beyond dopamine in the nucleus accumbens: The integrative functions of cortico-basal ganglia networks. *European Journal of Neuroscience*, *28*, 1437-1448.

Young, L., & Koenigs, M. (2007). Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin*, *84*, 69-79.