86 The Cognitive Neuroscience of Moral Judgment and Decision Making

JOSHUA D. GREENE

ABSTRACT   This article reviews recent advances in the cognitive neuroscience of moral judgment and behavior. This field is conceived, not as the study of a distinct set of neural functions, but as an attempt to understand how the brain's core neural systems coordinate to solve problems that we define, for nonneuroscientific reasons, as "moral." These systems enable the representation of value, cognitive control, the imagination of distal events, and the representation of mental states. Research examines the brains of morally pathological individuals, the responses of healthy brains to prototypically immoral actions, and the brain's responses to more complex moral problems such as philosophical and economic dilemmas.

Cognitive neuroscience aims to understand the mind in physical terms. This endeavor assumes that the mind *can* be understood in physical terms, and, insofar as it is successful, validates that assumption. Against this philosophical backdrop, the cognitive neuroscience of moral judgment takes on special significance. Moral judgment is, for many, the quintessential operation of the mind beyond the body, the earthly signature of the soul (Greene, 2011). (In many religious traditions it is, after all, the quality of a soul's moral judgment that determines where it ends up.) Thus, the prospect of understanding moral judgment in physical terms is especially alluring, or unsettling, depending on your point of view. In this brief review I provide a progress report on our attempts to understand how the human brain makes moral judgments and decisions.

The paradox of the "moral brain"

The fundamental problem with the "moral brain" is that it threatens to take over the entire brain, and thus cease to be a meaningful neuroscientific topic. This is not because morality is meaningless, but rather because neuroscience is centrally concerned with physical mechanisms, and it's increasingly clear that morality has few, if any, neural mechanisms of its own (Greene & Haidt, 2002; Parkinson et al., 2011; Young & Dungan, 2012).

By way of analogy, consider the concept of a *vehicle*. Motorcycles and sailboats are vehicles. Lawnmowers and kites are not. But, mechanically speaking, motorcycles have more in common with (gas-powered) lawnmowers than with sailboats, and sailboats operate more like kites than motorcycles. This doesn't mean that the concept of a vehicle is meaningless. Rather, the world's vehicles are united, not by their internal mechanisms, but at a more abstract, functional level. So, too, with morality. More specifically, I (Greene, 2013), like many others (Darwin, 1871/2004; Frank, 1988; Gintis, Bowles, Boyd, & Fehr, 2005; Haidt, 2012), believe that morality is a suite of cognitive mechanisms that enable otherwise selfish individuals to reap the benefits of cooperation. That is, we have psychological features that are straightforwardly moral (such as empathy, righteous indignation, and an aversion to harming innocent people) and others that are not (such as gossip, embarrassment, vengefulness, and ingroup favoritism) because they enable us to achieve goals that we can't achieve through collective selfishness. I won't defend this controversial thesis here. Instead, my point is that *if* this unified theory of morality is correct, it doesn't bode well for a unified theory of moral neuroscience. What's more, as we'll see, the data increasingly bear out this skepticism. In the early days of moral neuroscience, it was thought, perhaps not unreasonably, that one might isolate the distinctive neural mechanisms of moral thought (Moll, Eslinger, & Oliveira-Souza, 2001) and that the human brain might house a dedicated "moral organ" (Hauser, 2006). These views, however, are no longer tenable. It's now clear that the "moral brain" is, more or less, the whole brain, applying its computational powers to problems that we, on nonneuroscientific grounds, identify as "moral."

Understanding this is, itself, a kind of progress, but it leaves the cognitive neuroscience of morality—and the author of a chapter that would summarize it—in an awkward position. To truly understand the neuroscience of morality, we must understand the many neural systems that shape moral thinking, none of which, so far, appears to be specifically moral. These include systems that enable the representation of value and that motivate its pursuit (Knutson, Taylor, Kaufman,

Peterson & Glover, 2005; Pessoa, 2010; Rangel, Camerer, & Montague, 2008; Schultz, Dayan, & Montague, 1997), systems that orchestrate thought and action in accordance with internal goals (Miller & Cohen, 2001), systems that enable the imagination of complex distal events (Buckner, Andrews-Hanna, & Schacter, 2008; Raichle et al., 2001), and systems that enable the representation of people's hidden mental states (Frith & Frith, 2006; Mitchell, 2009), among others. In short, if you want to understand the neuroscience of morality, you might start by working your way through this weighty volume.

Of course, some neuroscientific topics bear more directly on morality than others, as indicated by my nonrandom list of relevant neural systems. This suggests that the present task isn't hopeless, that we can make some useful generalizations about the cognitive neuroscience of morality, even while acknowledging that the moral brain is not a distinct entity. This field, properly understood, will not isolate and describe the mechanisms essential for morality while the rest of cognitive neuroscience goes about its business. Instead, it provides a set of useful *entry points* into the broader problems of complex cognition and decision making (cf. Buckholtz & Meyer-Lindenberg, 2012, for a parallel view of psychopathology). More specifically, we can study the brains of people who reliably commit basic moral transgressions, the reactions of healthy brains to such transgressions, and the ways in which our brains handle more complex moral problems. Along the way we'll encounter some recurring themes that point the way toward a more encompassing account of moral, and nonmoral, cognition.

Bad brains

In the 1990s, Damasio and colleagues published a series of path-breaking studies of decision making in patients with damage to ventromedial prefrontal cortex (VMPFC), one of the regions damaged in the famous case of Phineas Gage (Damasio, 1994). VMPFC patients were mysterious because their real-life decision making was clearly impaired, but their deficits typically evaded detection using standard neurological measures of executive function (Saver & Damasio, 1991) and moral reasoning (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999). Using a game designed to simulate real-world risky decision making (the Iowa Gambling Task), Bechara, Tranel, Damasio, and Damasio (1996) documented these behavioral deficits and demonstrated, using autonomic measures, that these deficits are emotional. It seems that such patients make poor

decisions because they are unable to generate the feelings that guide adaptive decision making in healthy individuals. These early studies, while identifying a key biological substrate for moral choice, also underscore the critical role of learning in moral development. Late-onset VMPFC damage typically results in poor decision making and a deterioration of "moral character" (Damasio, 1994), but children with early-onset VMPFC damage are likely to develop into "sociopathic" adults who, in addition to being reckless and irresponsible, are duplicitous, aggressive, and strikingly lacking in empathy (Anderson et al., 1999; Grattan & Eslinger, 1992).

Studies of psychopaths and other individuals with antisocial personality disorder (APD) underscore the importance of emotion in moral decision making. APD is a catch-all diagnosis for individuals whose behavior is unusually antisocial. Psychopathy, in contrast, is a more specific, somewhat heritable disorder (Viding, Blair, Moffitt, & Plomin, 2005) whereby individuals exhibit a pathological degree of callousness, lack of empathy or emotional depth, and lack of genuine remorse for their antisocial actions (Hare, 1991). Psychopaths tend to engage in instrumental aggression, while other individuals with APD are characterized by reactive aggression (Blair, 2001).

Psychopathy is characterized by profound but selective emotional deficits. Psychopaths exhibit normal electrodermal responses to threat cues (e.g., a picture of shark's open mouth), but reduced responses to distress cues (e.g., a picture of a crying child; Blair, Jones, Clark, & Smith, 1997). In a classic study, Blair (1995) provided evidence that psychopaths fail to distinguish between rules that authorities cannot legitimately change ("moral" rules, e.g., a classroom rule against hitting) from rules that authorities can legitimately change ("conventional" rules, e.g., a rule prohibiting talking out of turn). According to Blair, psychopaths see all rules as *mere* rules because they lack the emotional responses that lead ordinary people to imbue moral rules with genuine, authority-independent moral legitimacy. While this is consistent with what is generally known about psychopathic psychology, a more recent study challenges the original finding that psychopaths do not draw the moral/conventional distinction (Aharoni, Sinnott-Armstrong, & Kiehl, 2012).

Studies of psychopathy and APD implicate a wide range of brain regions including the insula, posterior cingulate cortex, parahippocampal gyrus, and superior temporal gyrus (Kiehl, 2006; Raine & Yang, 2006). However, as emphasized by Blair (2007), two interconnected structures take center stage: the amygdala and

the VMPFC. These regions, along with subregions of subgenual anterior cingulate cortex and lateral prefrontal cortex, form a network that is essential for generating and regulating responses to salient stimuli (Pessoa, 2010). Blair (2007) has proposed that psychopathy arises primarily from amygdala dysfunction, which is crucial for stimulus-reinforcement learning (Davis & Whalen, 2001) and thus for normal moral socialization (Oxford, Cavell, & Hughs, 2003). In psychopaths (or individuals with psychopathic traits) the amygdala exhibits weaker responses to fearful faces (Marsh et al., 2008), to emotional words (Kiehl et al., 2001), to pictures indicating moral violations (Harenski, Harenski, Shane, & Kiehl, 2010; Harenski, Kim, & Hamann, 2009), and to dilemmas involving harmful actions (Glenn, Raine, & Schug, 2009). As noted above, the amygdala operates in tight conjunction with the VMPFC, and, consistent with this, psychopathic individuals also exhibit reduced VMPFC responses to morally salient stimuli (Harenski et al., 2010). Beyond the amygdala-VMPFC circuit, psychopaths also exhibit hypoactivity in the default mode network (DMN; Buckner et al., 2008; Raichle et al., 2001) during moral judgment (Pujol et al., 2012), consistent with this network's heightened response to emotionally engaging moral dilemmas in healthy people (Greene et al., 2001). (Note that some of the participants in this study failed to meet standard criteria for psychopathy. See Schaich Borg & Sinnott-Armstrong, 2013.)

Psychopaths, in addition to their weak affective responses to harm, are known for their impulsive behavior (Hare, 1991). The VMPFC serves as part of the frontostriatal pathway, responsible for representing the values of outcomes and actions based on past experience (Knutson et al., 2005; Rangel et al., 2008). Individuals with psychopathic traits (specifically, impulsive antisocial behavior) exhibit heightened responses to reward within this system (Buckholtz et al., 2010) along with increased striatal volume (Glenn, Raine, Yaralian, & Yang, 2010). Finally, their emotional deficits may sometimes cause them to rely more heavily on explicit reasoning, dependent on the frontoparietal control network (Glenn, Raine, Schug, Young & Hauser, 2009; Koenigs, Kruepke, Zeier, & Newman, 2012). Thus, while the origins of psychopathy may lie in one or more discrete neural abnormalities, their influence is felt throughout the brain.

Good brains

Studies of healthy individuals responding to moral transgressions are generally consistent with studies of psychopaths and others with APD. They, too, highlight the importance of the amygdala and VMPFC (Blair, 2007; Decety & Porges, 2011; Heekeren et al., 2005; Moll et al., 2002; Schaich Borg et al., 2006;) and confirm the importance of these structures in moral development (Decety, Michalska, & Kinzler, 2012). For reasons explained below, studies of moral judgment employing text-based narrative stimuli tend to implicate the entire DMN. Several studies highlight the importance of the insula in representing the aversiveness of moral transgressions (Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Decety, Michalska, & Kinzler, 2012; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Schaich Borg, Lieberman, & Kiehl, 2008; Schaich Borg, Sinnott-Armstrong, Calhoun, & Kiehl, 2011). Others indicate that the representation of moral value, like other forms of value, depends on the brain's domain-general valuation mechanisms enabled by the frontostriatal pathway (Decety & Porges, 2011; Moll et al., 2006; Shenhav & Greene, 2010).

One of the most basic distinctions in moral evaluation is between intentional and accidental harm. (As Oliver Wendell Holmes Jr. famously observed, even a dog knows the difference between being tripped over and being kicked.) Young, Saxe, and colleagues have conducted a series of studies examining how the brain represents and applies this distinction in the context of moral judgment. Their work highlights the importance of the temporoparietal junction (TPJ) along with other DMN regions, which are widely implicated in ToM mentalizing (Frith & Frith, 2006; Mitchell, 2009). The TPJ is especially sensitive to attempted harms (Koster-Hale, Saxe, Dungan, & Young, 2013; Young, Cushman, Hauser, & Saxe, 2007), which are wrong because of the agent's mental state, not the action's outcome. Disrupting TPJ activity results in a child-like (Piaget, 1965), "no harm no foul" pattern of judgment in which attempted harms are judged less harshly (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010). We see the same pattern in patients with VMPFC damage (Young, Bechara et al., 2010) and split-brain patients (Miller et al., 2010), indicating that the use of mental-state information in moral judgment depends, at least in part, on translating this information into an affective signal and on the integration of information across the cerebral hemispheres. Individuals with high-functioning autism exhibit a complementary pattern, "if harm, then foul," judging accidental harms unusually harshly (Moran et al., 2011). Accidental harms appear to set up a tension between outcome-based and intention-based harm. Consistent with this, such harms preferentially engage the frontoparietal control network (Miller & Cohen, 2001).

Puzzled brains

We've considered the two most straightforward entry points into moral neuroscience: the unhealthy brains of people who act badly and the healthy brain's responses to protoypically bad acts. A third approach begins with moral dilemmas. Moral dilemmas are useful, not because they reflect everyday moral experience, but because dilemmas, by their nature, pit competing processes against one another. They are high-contrast stimuli, analogous to the flashing checkerboards of vision scientists, and thus especially useful for revealing cognitive structure (Cushman & Greene, 2012).

The research described above emphasizes the role of emotion in moral judgment (Haidt, 2001), while traditional theories of moral development emphasize the role of controlled cognition (Kohlberg, 1969; Turiel, 2006). I and others have developed a dual-process (Chaiken & Trope, 1999; Kahneman, 2003) theory of moral judgment that synthesizes these perspectives (Greene et al., 2001; Greene, 2007, 2013). According to this theory, both intuitive emotional responses and more controlled cognitive responses play crucial and, in some cases, competing roles. More specifically, this theory associates controlled cognition with utilitarian (or consequentialist) moral judgment aimed at promoting the "greater good" (Mill, 1861/1998) while associating automatic emotional responses with competing deontological judgments that are naturally justified in terms of rights or duties (Kant, 1785/1959).

We developed this theory in response to a long-standing philosophical puzzle known as the Trolley Problem (Foot, 1978; Thomson, 1985). In one version, which I'll call the *switch* case, one can save five people who are mortally threatened by a runaway trolley by hitting a switch. This will turn the trolley onto a side track, where it will run over and kill only one person instead. Here, most people approve of diverting the trolley (Petrinovich, O'Neill, & Jorgensen, 1993), a characteristically utilitarian judgment favoring the greater good. In the contrasting *footbridge* dilemma, a runaway trolley once again threatens five people. The only way to save the five is to push a large person off a footbridge and into the trolley's path, stopping the trolley but killing the person pushed. (Yes, this will work, and, no, you can't stop the trolley yourself.) Here, most people say that it's wrong to trade one life for five, consistent with the deontological perspective favoring the rights of the individual over the greater good. The question: why do people typically say "yes" to hitting the switch, but "no" to pushing?

We hypothesized that this pattern of judgment reflects the outputs of distinct and (in some cases) competing neural systems (Greene et al., 2001). The more "personal"[1] harmful action in the *footbridge* case, pushing the man off the footbridge, triggers a relatively strong negative emotional response, while the relatively impersonal harmful action in the *switch* case does not. This predicts increased activity in emotion-related brain regions in response to "personal" dilemmas, such as the *footbridge* case, as compared to "impersonal" dilemmas, such as the *switch* case.

This emotional response can explain why people say "no" to pushing the man off the footbridge. But why do people say "yes" to hitting the switch? The answer seems obvious enough: hitting the switch saves more lives. We hypothesized that this utilitarian response depends on explicit cost-benefit reasoning enabled by the frontoparietal control network (Miller & Cohen, 2001), including the DLPFC. Thus, we predicted increased DLPFC activity in response to "impersonal" dilemmas, such as the *switch* case, in which this controlled response tends to dominate. Likewise, we predicted increased DLPFC activity when people override a negative emotional response in making a utilitarian judgment, as when people say "yes" to the *footbridge* dilemma.

We first tested this theory using functional MRI (fMRI; Greene et al., 2001), contrasting a (rather heterogeneous) set of "personal" dilemmas with a set of (even more heterogeneous) "impersonal" dilemmas. (More recent studies have been better controlled, focusing on differing responses to "high-conflict" dilemmas such as the *footbridge* case.) We found that the "personal" dilemmas elicited increased activity in what is now known as the DMN (Buckner et al., 2008; Raichle et al., 2001), including large portions of medial prefrontal cortex, medial parietal cortex, and the TPJ, all of which had been previously associated with emotion (e.g., Maddock, 1999). In contrast, the "impersonal" dilemmas elicited relatively greater activity in the frontoparietal control network. Also as predicted, our second fMRI experiment (Greene et al., 2004) found increased DLPFC activity for utilitarian judgment and increased amygdala activity for "personal" dilemmas. These results provided initial support for the dual-process theory, which has been both supported and refined by subsequent research using a broad range of methods.

---

[1]The personal/impersonal distinction (Greene et al., 2001) has been revised (Greene et al., 2009) since it was originally introduced. For present purposes, one can think of "personal" harms as ones in which the agent actively and intentionally harms the victim using the direct force of his or her muscles.

In retrospect, the DMN's response to "personal" dilemmas is best interpreted as *related* to increased emotional engagement, but not as its proper neural substrate. The DMN is active when people are doing nothing in particular (hence "default") and is most reliably engaged by attention to nonpresent events, as in remembering the past, imaging the future, thinking about contents of other minds, and imaging hypothetical possibilities (Buckner et al., 2008; DeBrigard, Addis, Ford, Schacter, & Giovanello, 2013). Thus, if "personal" dilemmas preferentially engage the DMN, it's probably not because DMN activity reflects emotional engagement per se. Rather, it's because "personal" dilemmas make for especially gripping mental television, which may be both a cause and a consequence of their emotional salience. Consistent with this hypothesis, Amit and Greene (2012) found that individuals with more visual cognitive styles tend to make fewer utilitarian judgments in response to high-conflict personal dilemmas and that disrupting visual imagery while contemplating these dilemmas increases utilitarian judgment.

More direct evidence for the dual-process theory comes from studies of patients with emotion-related deficits. Mendez, Anderson, and Shapira (2005) found that patients with frontotemporal dementia, who are known for their "emotional blunting," were disproportionately likely to approve of the utilitarian action in the *footbridge* dilemma. Likewise, patients with VMPFC lesions make up to five times as many utilitarian judgments in response to standard high-conflict dilemmas (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs et al., 2007) and in response to dilemmas pitting familial duty against the greater good (e.g., your sister vs. five strangers; Thomas, Croft, & Tranel, 2011). VMPFC patients also exhibit correspondingly weak physiological responses when making such judgments (Moretto, Ladàvas, Mattioli, & di Pellegrino, 2010), and healthy people who are more physiologically reactive are less utilitarian (Cushman, Murray, Gordon-Mckeon, Wharton, & Greene, 2012). Low-anxiety psychopaths (Koenigs et al., 2012) and people with high levels of testosterone (Carney & Mason, 2010), which is associated with a higher tolerance for stress, tend to make more utilitarian judgments, as do people with alexithymia (Koven, 2011), a condition that reduces awareness of one's own emotional states. Here, the VMPFC seems to respond specifically to harmful behavior that is active and also intentional, rather than merely foreseen (Schaich Borg et al., 2006).

Other studies highlight the role of the amygdala. As noted above, individuals with psychopathic traits exhibit reduced amygdala responses to personal moral dilemmas (Glenn, Raine, & Schug, 2009). In healthy people, amygdala activity tracks self-reported emotional responses to harmful transgressions and predicts deontological judgments in response to them (Shenhav & Greene, 2013). Studies employing pharmacological interventions paint a consistent picture. Citalopram—a selective serotonin-reuptake inhibitor (SSRI) that, in the short-term, increases emotional reactivity through its influence on the amygdala and VMPFC, among other regions—increases deontological judgment (Crockett, Clark, Hauser, & Robbins, 2010). By contrast, lorazepam, an anti-anxiety drug, has the opposite effect (Perkins et al., 2012). Consistent with the effects of citalopram, variation in the serotonin transporter (*5-HTTLPR*) genotype (S alleles) predicts deontological judgment, but in response dilemmas in which the harm is a foreseen side effect (Marsh et al., 2011).

Most of the evidence linking controlled cognition to utilitarian judgment comes from behavioral studies beyond the scope of this chapter (e.g., Greene et al., 2008; Paxton, Ungar, & Greene, 2012). However, a few neuroscientific studies, in addition to those described above (Greene et al., 2001, 2004), provide further evidence. Sarlo et al. (2012) examined the temporal dynamics of moral judgment using EEG and found a pattern consistent with the results of Greene et al. (2001, 2004). Here, *footbridge*-like dilemmas produced a stronger early neural response (P260) in regions consistent with VMPFC activity, while *switch*-like dilemmas elicited more utilitarian responses and a more pronounced later component consistent with the engagement of the frontoparietal control network. Also consistent with this, activity in the frontoparietal control network is associated with rejecting the deontological distinction between harmful acts and harmful omissions (Cushman et al., 2012). (See also Schaich Borg et al., 2006.) Likewise, VMPFC patients who tend to give more utilitarian responses are thought to do so because their capacity for explicit, cost-benefit reasoning remains intact (Koenigs et al., 2007).

A recent study (Shenhav & Greene, 2013) helps differentiate the functions of the amygdala and VMPFC in moral judgment. As noted above, amygdala signal tracks with self-reports of negative emotional responses to harmful actions and predicts deontological condemnation of those actions. The VMPFC, however, does not. Instead, the VMPFC is most active when people have to make "all things considered" judgments, as compared to simply reporting on emotional reactions or utilitarian considerations. This suggests that the amygdala generates an initial negative response to personally harmful actions (consistent with Glenn, Raine, & Schug, 2009), while the VMPFC weighs that signal against a

competing signal reflecting the utilitarian advantages of committing the harmful act. This is consistent with an evolving understanding of the VMPFC as a domain-general integrator of decision weights (Rangel & Hare, 2010). However, this leaves us with a puzzle: if the VMPFC is acting as a neutral broker among competing decision weights, then why does VMPFC damage so reliably increase utilitarian judgment? Our hypothesis is that the frontoparietal control network's explicit utilitarian reasoning can influence behavior independent of the VMPFC, while the amygdala's competing deontological signal requires the VMPFC's integration, at least when competing utilitarian considerations are in play. Thus, if this is correct, VMPFC damage favors utilitarian judgment, not by damaging a region with inherent deontological tendencies, but by damaging a pathway that is necessary for deontological judgment, but not utilitarian judgment, to prevail.

This integrative role for the VMPFC is consistent with its role in integrating other kinds of morally relevant information. Shenhav and Greene (2010) examined people's responses to dilemmas in which failing to save one person can allow one to save a group of others. We varied the size of the group and the probability of saving them. We found that neural sensitivity to the magnitude of the outcome (group size) in the ventral striatum predicts behavioral sensitivity to this variable, and we observed a parallel effect for outcome probability in the insula. The VMPFC, by contrast, responded to the interaction of these two variables, reflecting the probability-discounted magnitude of the moral consequences. In other words, the VMPFC represents "expected moral value," just as it represents expected value in self-interested economic decision making (Knutson et al., 2005). Thus, once again, we see a domain-general system—here, the frontostriatal pathway—operating in the context of moral judgment. This system evolved in mammals to value goods that tend to exhibit diminishing marginal returns. This may explain our puzzling (and highly consequential) tendency to regard the saving of human lives as exhibiting diminishing marginal returns, as if the hundredth life saved is somehow worth less than the first.

In an important theoretical development, Cushman (2013) and Crockett (2013) have proposed that the dissociation between deontological and utilitarian/ consequentialist judgment reflects a more general dissociation between model-free and model-based learning systems (Daw & Doya, 2006). Model-free learning mechanisms assign values to actions intrinsically based on past experience, while model-based learning mechanisms attach values to actions based on internal models of causal relations in the world. Thus, an action may seem intrinsically wrong because past experience has associated actions of that type (e.g., pushing people) with negative consequences (e.g., social disapproval), and yet the same action may seem right because it will, according to one's world-model, produce optimal consequences (saving five lives instead of one). Thus, the fundamental tension in normative ethics, reflected in the competing philosophies of Kant and Mill, may find its origins in a competition between distinct, domain-general mechanisms for assigning values to actions.

Cooperative brains

Research on altruism and cooperation does not always fall under the heading of "morality," but it could not be more central to our understanding of the moral brain. The most basic question about the cognitive neuroscience of altruism and cooperation is this: what neural processes enable and motivate people to be "nice"— that is, to pay costs to benefit others?

Consistent with our evolving story, the value of helping others, both in unidirectional altruism and bidirectional cooperation, is represented in the frontostriatal pathway. Activity in this pathway tracks the value of charitable contributions (Moll et al., 2006; Hare et al., 2010), sharing resources with other individuals (Zaki & Mitchell, 2011), and cooperation (Rilling et al., 2007), maximizing benefits delivered by a distribution of resource (i.e., "efficiency"), and optimizing the subjective tradeoff between efficiency and equality (Hsu, Anen, & Quartz, 2008). Likewise, this pathway tracks the value of punishing individuals who are insufficiently "nice" (de Quervain et al., 2004; Singer et al., 2006). As above, the DMN has a hand in altruism as well. TPJ volume (Morishima, Schunk, Bruhin, Ruff, & Fehr, 2012) and medial PFC activity (Rilling et al., 2007; Waytz, Zaki, & Mitchell, 2012) both predict altruistic behavior.

Thus, the brain uses its endogenous carrots—reward signals—to motivate cooperative behavior. It also uses its sticks—negative affective responses to uncooperative behavior. Activity in the insula, known for its role in the representation of somatic states and the awareness of feelings (Craig, 2009), scales with the magnitude of the unfairness in unfair Ultimatum Game offers (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), predicts aversion to inequality in the distribution of resources (Hsu et al., 2008), and predicts egalitarian behavior and attitudes (Dawes et al., 2012). The insula and the amygdala both respond to the punishment of well-behaved people (Singer, Kiebel, Winston, Dolan, & Frith, 2004).

The dual-process tension between automatic and controlled processes is observed in a range of morally laden economic choices. Accepting unfair Ultimatum Game offers, despite their distastefulness, is associated with increased activity in the frontoparietal control network (Sanfey et al., 2003; Tabibnia, Satpute, & Lieberman, 2008). Perhaps surprisingly, VMPFC damage leads to increased *rejection* of unfair offers (Koenigs & Tranel, 2007). (Consistent with this, psychopaths do the same; Koenigs, Kruepke, & Newman, 2010.) This may be because the VMPFC integrates signals responding both to unfairness and material gain (which compete in the Ultimatum Game) and because, in the absence of such signals, one applies a reciprocity rule. In a study of dishonesty, Greene and Paxton (2009) gave people repeated opportunities to gain money by lying about their accuracy in predicting the outcomes of coin-flips. Consistently honest subjects appeared to be "gracefully" honest, exhibiting no additional engagement of the frontoparietal control network in forgoing dishonest gains. By contrast, subjects who behaved dishonestly (as indicated by improbably high self-reported accuracy) exhibited increased control-related activity, both when lying and when refraining from lying. A follow-up study (Abe & Greene, 2013) traces these behavioral differences to response characteristics of the frontostriatal pathway. Baumgartner et al. (2009) describe a similar dual-process dynamic, in which breaking promises involves increased engagement of the amygdala and the frontoparietal control network. (For a behavioral approach to dual-process cooperation, also see Rand, Greene, & Nowak, 2012.)

Cooperation depends on trust, which in turn requires evaluating individuals (Delgado, Frank, & Phelps, 2005; Singer et al., 2004) and groups (Phelps et al., 2000) as potential cooperation partners. Oxytocin, a neuropeptide known for its role in social attachment and affiliation in mammals (Insel & Young, 2001) appears to be important for both kinds of decisions. Intranasal administration of oxytocin increases investment in a "trust game" (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005), but also biases judgment and behavior toward ingroup members and against outgroup members (de Dreu et al., 2010; de Dreu, Greer, Van Kleef, Shalvi, & Handgraaf, 2011). Likewise, genetic variants associated with oxytocin are associated with increased prosocial behavior, particularly when the world is seen as threatening (Poulin, Holman, & Buffone, 2012).

From an evolutionary perspective, the double-edged sword of human morality comes as no surprise. Morality evolved, not as device for universal cooperation, but as a competitive weapon, as a system for turning Me into Us, which in turn enables Us to outcompete Them.

Morality's dark, tribalistic side is powerful, but there's no reason why it must prevail. The flexible thinking enabled by our enlarged prefrontal cortices may enable us to retain the best of our moral impulses while transcending their inherent limitations (Greene, 2013; Pinker, 2011).

Looking back, and ahead

How does the moral brain work? Answer: exactly the way you'd expect it to work if you understand (1) which cognitive functions morality requires and (2) which cognitive functions are performed by the brain's core neural systems. On the one hand, this means that morality has no proprietary neural territory of its own. On the other hand, it means that the cognitive neuroscience of morality, beginning with the entry points described above, can teach important lessons about how the brain's core neural systems interact to solve complex problems.

From its inception, cognitive neuroscience has focused on structure-function relationships. We have a general understanding of what various neural structures do, but when it comes to complex cognition, we're mostly blind to the specific information content shuttled about the brain. We know, for example, that the thought of pushing someone off a footbridge pushes our emotional buttons, but we know almost nothing about how we think such thoughts in the first place. However, with the advent of multivariate analysis methods (Kriegeskorte, Goebel, & Bandettini, 2006; Norman, Polyn, Detre, & Haxby, 2006), we may finally be ready to understand how the brain encodes and manipulates the *contents* of thoughts. When we finally do, we will learn a lot more about morality—and everything else.

REFERENCES

ABE, N., & GREENE, J. D. (2013). Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. (Under review.)

AHARONI, E., SINNOTT-ARMSTRONG, W., & KIEHL, K. A. (2012). Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *J Abnorm Psychol*, *121*(2), 484.

AMIT, E., & GREENE, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychol Sci*, *23*(8), 861–868.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat Neurosci*, *2*, 1032–1037.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron*, *64*(5), 756–770.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb Cortex*, *6*, 215–225.

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

Blair, R. J. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J Neurol Neurosurg Psychiatry*, *71*, 727–731.

Blair, R. J. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends Cogn Sci*, *11*, 387–392.

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

Buckholtz, J. W., & Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: Toward a transdiagnostic model of risk for mental illness. *Neuron*, *74*(6), 990–1004.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., & Zald, D. H. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nat Neurosci*, *13*(4), 419–421.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Ann NY Acad Sci*, *1124*(1), 1–38.

Carney, D. R., & Mason, M. F. (2010). Decision making and testosterone: When the ends justify the means. *J Exp Soc Psychol*, *46*(4), 668–671.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York, NY: Guilford.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Soc Cogn Affect Neurosci*, *2*, 84–92.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci*, *10*(1), 59–70.

Crockett, M. J. (2013). Models of morality. *Trends Cogn Sci*, *17*(8), 363–366.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc Natl Acad Sci USA*, *107*(40), 17433–17438.

Cushman, F. (2013). Action, outcome and value: A dual-system framework for morality. *Pers Soc Psychol Rev*, *17*(3), 273–292.

Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Soc Neurosci*, *7*(3), 269–279.

Cushman, F., Murray, D., Gordon-McKeon, S., Wharton, S., & Greene, J. D. (2012). Judgment before principle: Engagement of the frontoparietal control network in condemning harms of omission. *Soc Cogn Affect Neurosci*, *7*(8), 888–895.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: G. P. Putnam.

Darwin, C. (1871/2004). *The descent of man*. New York, NY: Penguin.

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Mol Psychiatry*, *6*, 13–34.

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Curr Opin Neurobiol*, *16*(2), 199–204.

Dawes, C. T., Loewen, P. J., Schreiber, D., Simmons, A. N., Flagan, T., McElreath, R., & Paulus, M. P. (2012). Neural basis of egalitarian behavior. *Proc Natl Acad Sci USA*, *109*(17), 6479–6483.

De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia*, *51*(12), 2401–2414.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., van Kleef, G. A., Baas, M., & Feith, S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*(5984), 1408–1411.

De Dreu, C. K., Greer, L. L., van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. (2011). Oxytocin promotes human ethnocentrism. *Proc Natl Acad Sci USA*, *108*(4), 1262–1266.

De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.

Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cereb Cortex*, *22*(1), 209–220.

Decety, J., & Porges, E. C. (2011). Imagining being the agent of actions that carry different moral consequences: An fMRI study. *Neuropsychologia*, *49*(11), 2994–3001.

Delgado, M. R., Frank, R., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci*, *8*, 1611–1618.

Foot, P. (1978). The problem of abortion and the doctrine of double effect. In P. Foot, Virtues and vices, and other essays in moral philosophy (pp. 19–32). Oxford, UK: Blackwell.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York, NY: Norton.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Gintis, H. E., Bowles, S. E., Boyd, R. E., & Fehr, E. E. (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. Cambridge, MA: MIT Press.

Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Mol Psychiatry*, *14*(1), 5.

Glenn, A. L., Raine, A., Schug, R. A., Young, L., & Hauser, M. (2009). Increased DLPFC activity during moral decision-making in psychopathy. *Mol Psychiatry*, *14*(10), 909–911.

Glenn, A. L., Raine, A., Yaralian, P. S., & Yang, Y. (2010). Increased volume of the striatum in psychopathic individuals. *Biol Psychiatry*, *67*(1), 52–58.

GRATTAN, L. M., & ESLINGER, P. J. (1992). Long-term psychological consequences of childhood frontal lobe lesion in patient DT. *Brain Cogn, 20*, 185–195.

GREENE, J. (2011). Social neuroscience and the soul's last stand. In A. Todorov, S. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind.* New York, NY: Oxford University Press.

GREENE, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them.* New York, NY: Penguin.

GREENE, J., & HAIDT, J. (2002). How (and where) does moral judgment work? *Trends Cogn Sci, 6*, 517–523.

GREENE, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology.* Vol. 3: *The neuroscience of morality: Emotion, disease, and development* (pp. 35–79). 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*(3), 364–371.

GREENE, J. D., MORELLI, S., LOWENBERG, K., NYSTROM, L., & 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., & PAXTON, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proc Natl Acad Sci USA, 106*(30), 12506–12511.

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.

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

HAIDT, J. (2012). *The righteous mind: Why good people are divided by politics and religion.* New York, NY: Random House.

HARE, R. D. (1991). *The Hare psychopathy checklist.* Revised. Toronto: Multi-Health Systems.

HARE, T. A., CAMERER, C. F., KNOEPFLE, D. T., O'DOHERTY, J. P., & RANGEL, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J Neurosci, 30*(2), 583–590.

HARENSKI, C. L., HARENSKI, K. A., SHANE, M. S., & KIEHL, K. A. (2010). Aberrant neural processing of moral violations in criminal psychopaths. *J Abnorm Psychol, 119*(4), 863.

HARENSKI, C. L., KIM, S. H., & HAMANN, S. (2009). Neuroticism and psychopathy predict brain activation during moral and nonmoral emotion regulation. *Cogn Aff Behav Neurosci, 9*(1), 1–15.

HAUSER, M. (2006). The liver and the moral organ. *Soc Cogn Affect Neurosci, 1*, 214–220.

HEEKEREN, H. R., WARTENBURGER, I., SCHMIDT, H., PREHN, K., SCHWINTOWSKI, H. P., & VILLRINGER, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage, 24*, 887–897.

HSU, M., ANEN, C., & QUARTZ, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science, 320*, 1092–1095.

INSEL, T. R., & YOUNG, L. J. (2001). The neurobiology of attachment. *Nat Rev Neurosci, 2*, 129–136.

KAHNEMAN, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *Am Psychol, 58*, 697–720.

KANT, I. (1785/1959). *Foundation of the metaphysics of morals.* Indianapolis, IN: Bobbs-Merrill.

KIEHL, K. A. (2006). A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. *Psychiatry Res, 142*, 107–128.

KIEHL, K. A., SMITH, A. M., HARE, R. D., MENDREK, A., FORSTER, B. B., BRINK, J., & LIDDLE, P. F. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol Psychiatry, 50*, 677–684.

KOVEN, N. S. (2011). Specificity of meta-emotion effects on moral decision-making. *Emotion, 11*(5), 1255.

KNUTSON, B., TAYLOR, J., KAUFMAN, M., PETERSON, R., & GLOVER, G. (2005). Distributed neural representation of expected value. *J Neurosci, 25*(19), 4806–4812.

KOENIGS, M., KRUEPKE, M., ZEIER, J., & NEWMAN, J. P. (2012). Utilitarian moral judgment in psychopathy. *Soc Cogn Affect Neurosci, 7*(6), 708–714.

KOENIGS, M., KRUEPKE, M., & NEWMAN, J. P. (2010). Economic decision-making in psychopathy: A comparison with ventromedial prefrontal lesion patients. *Neuropsychologia, 48*(7), 2198–2204.

KOENIGS, M., & TRANEL, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *J Neurosci, 27*, 951–956.

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.

KOHLBERG, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago, IL: Rand McNally.

KOSFELD, M., HEINRICHS, M., ZAK, P. J., FISCHBACHER, U., & FEHR, E. (2005). Oxytocin increases trust in humans. *Nature, 435*, 673–676.

KOSTER-HALE, J., SAXE, R., DUNGAN, J., & YOUNG, L. L. (2013). Decoding moral judgments from neural representations of intentions. *Proc Natl Acad Sci USA, 110*(14), 5648–5653.

KRIEGESKORTE, N., GOEBEL, R., & BANDETTINI, P. (2006). Information-based functional brain mapping. *Proc Natl Acad Sci USA, 103*(10), 3863–3868.

MADDOCK, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends Neurosci, 22*, 310–316.

MARSH, A. A., CROWE, S. L., HENRY, H. Y., GORODETSKY, E. K., GOLDMAN, D., & BLAIR, R. J. R. (2011). Serotonin transporter genotype (5-HTTLPR) predicts utilitarian moral judgments. *PLoS ONE, 6*(10), e25148.

MARSH, A., FINGER, E., MITCHELL, D., REID, M., SIMS, C., KOSSON, D., & BLAIR, R. (2008). Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. *Am J Psychiatry, 165*(6), 712–720.

MENDEZ, M. F., ANDERSON, E., & SHAPIRA, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cogn Behav Neurol, 18*, 193–197.

Mill, J. S. (1861/1998). *Utilitarianism*. New York, NY: Oxford University Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci, 24*, 167–202.

Miller, M. B., Sinnott-Armstrong, W., Young, L., King, D., Paggi, A., Fabri, M., & Gazzaniga, M. S. (2010). Abnormal moral reasoning in complete and partial callosotomy patients. *Neuropsychologia, 48*(7), 2215–2220.

Mitchell, J. P. (2009). Inferences about mental states. *Philos Trans R Soc Lond B Biol Sci, 364*(1521), 1309–1316.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *J Neurosci, 22*(7), 2730–2736.

Moll, J., Eslinger, P. J., & Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arq Neuropsiquiatr, 59*, 657–664.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci USA, 103*, 15623–15628.

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proc Natl Acad Sci USA, 108*(7), 2688–2692.

Moretto, G., Làdavas, E., Mattioli, F., & di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J Cogn Neurosci, 22*(8), 1888–1899.

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron, 75*(1), 73–79.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci, 10*(9), 424–430.

Oxford, M., Cavell, T. A., & Hughes, J. N. (2003). Callous/unemotional traits moderate the relation between ineffective parenting and child externalizing problems: A partial replication and extension. *J Clin Child Adolesc Psychol, 32*, 577–585.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *J Cogn Neurosci, 23*(10), 3162–3180.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Sci, 36*(1), 163–177.

Perkins, A. M., Leonard, A. M., Weaver, K., Dalton, J. A., Mehta, M. A., Kumari, V., & Ettinger, U. (2012). A dose of ruthlessness: Interpersonal moral judgment is hardened by the anti-anxiety drug lorazepam. *J Exp Psychol Gen, 142*(3), 612–620.

Pessoa, L. (2010). Emotion and cognition and the amygdala: From "what is it?" to "what's to be done?" *Neuropsychologia, 48*(12), 3416–3429.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *J Pers Soc Psychol, 64*, 467–478.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *J Cogn Neurosci, 12*(5), 729–738.

Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. New York, NY: Viking.

Pujol, J., Batalla, I., Contreras-Rodríguez, O., Harrison, B. J., Pera, V., Hernández-Ribas, R., & Cardoner, N. (2012). Breakdown in the brain network subserving moral judgment in criminal psychopathy. *Soc Cogn Affect Neurosci, 7*(8), 917–923.

Piaget, J. (1965). *The moral judgement of the child*. New York, NY: Free Press.

Poulin, M. J., Holman, E. A., & Buffone, A. (2012). The neurogenetics of nice: Receptor genes for oxytocin and vasopressin interact with threat to predict prosocial behavior. *Psychol Sci, 23*(5), 446–452.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci USA, 98*(2), 676–682.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci, 9*(7), 545–556.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr Opin Neurobiol, 20*(2), 262–270.

Raine, A., & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Soc Cogn Affect Neurosci, 1*, 203–213.

Rilling, J., Glenn, A., Jairam, M., Pagnoni, G., Goldsmith, D., Elfenbein, H., & Lilienfeld, S. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biol Psychiatry, 61*, 1260–1271.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*, 1755–1758.

Sarlo, M., Lotto, L., Manfrinati, A., Rumiati, R., Gallicchio, G., & Palomba, D. (2012). Temporal dynamics of cognitive–emotional interplay in moral decision-making. *J Cogn Neurosci, 24*(4), 1018–1029.

Saver, J., & Damasio, A. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 29*, 1241–1249.

Schaich Borg, J., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *J Cogn Neurosci, 20*, 1–19.

Schaich Borg, J., Hynes, C., van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *J Cogn Neurosci, 18*, 803–817.

Schaich Borg, J., & Sinnott-Armstrong, W. (2013). Do psychopaths make moral judgments? In K. Kiehl & W. Sinnott-Armstrong (Eds.), *Handbook on psychopathy and law* (pp. 107–128). New York, NY: Oxford University Press.

Schaich Borg, J., Sinnott-Armstrong, W., Calhoun, V. D., & Kiehl, K. A. (2011). Neural basis of moral verdict and moral deliberation. *Soc Neurosci*, *6*(4), 398–413.

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

Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, *67*(4), 667–677.

Shenhav, A., & Greene, J. D. (2013). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *J Neurosci*, *34*(13), 4741–4749.

Singer, T., Kiebel, S., Winston, J., Dolan, R., & Frith, C. (2004). Brain response to the acquired moral status of faces. *Neuron*, *41*, 653–662.

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

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol Sci*, *19*, 339–347.

Thomas, B. C., Croft, K. E., & Tranel, D. (2011). Harming kin to save strangers: Further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *J Cogn Neurosci*, *23*(9), 2186–2196.

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

Turiel, E. (2006). Thought, emotions and social interactional processes in moral development. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 1–30). Mahwah, NJ: Erlbaum.

Viding, E., Blair, R. J., Moffitt, T. E., & Plomin, R. (2005). Evidence for substantial genetic risk for psychopathy in 7-year-olds. *J Child Psychol Psychiatry*, *46*, 592–597.

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J Neurosci*, *32*(22), 7646–7650.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*(6), 845–851.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc Natl Acad Sci USA*, *107*(15), 6753–6758.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc Natl Acad Sci USA*, *104*(20), 8235–8240.

Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Soc Neurosci*, *7*(1), 1–10.

Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proc Natl Acad Sci USA*, *108*(49), 19761–19766.