

A Neural Network for Moral Decision Making

O FeldmanHall, New York University, New York, NY, USA

D Mobbs, Columbia University, New York, NY, USA

© 2015 Elsevier Inc. All rights reserved.

Introduction

Some of the most fundamental psychological questions concerning human relations center on morality and altruism. Reinvigorated by the advent of modern technologies such as magnetic resonance imaging (MRI), these questions have been extended to humans, which has resulted in a recent flurry of empirical studies examining the psychological and neural architecture of moral cognition. The depth of this subject is remarkable, and there is insufficient room here to touch on all facets. As a result, this review will cover two questions central to human moral behavior: How do we arrive at our moral judgments? And what is the neural circuitry that underpins these moral judgments?

How Do We Arrive at Our Moral Judgments? Emotions and the Role of Empathy

An initial theory in moral research proposed that moral judgments were largely the product of applied reasoning (Kohlberg & Hersh, 1977). However, accumulating research illustrates that a confluence of factors – including emotion (Damasio, 1994), reason (Kohlberg, 1969), environmental context (Isen & Levin, 1972), social frameworks (FeldmanHall, Mobbs, Evans, Hiscox, Navardy, & Dalgleish, 2012), empathy and compassion (FeldmanHall, Dalgleish, Evans, & Mobbs, accepted for publication; Preston & De Waal, 2001), sensitivity to harm (Kedia, Berthoz, Wessa, Hilton, & Martinot, 2008), perceived intentions of another (Berthoz, Armony, Blair, & Dolan, 2002), and putative consequences (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006) – all contribute to guiding our moral judgments. In particular, the proposal that emotion is a fundamental precursor of prosocial behavior (Penner, Dovidio, Piliavin, & Schroeder, 2005) has generated a flourish of research cataloging the putative relationship between emotion and morality. Even with consideration to arguments to the contrary (Hauser, 2006; Huebner, Dwyer, & Hauser, 2009), emotions appear to be central in motivating moral judgments.

Take, for example, the classic trolley dilemma (Foot, 1978): while people are averse to physically pushing a man off a bridge in order to save the lives of five others, they are more willing to contemplate killing five if they only have to flip a switch (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). This particular set of dilemmas suggests that the emotional impact of tactilely shoving another engages emotional processes differentially than the more removed act of flipping a switch. Other research has examined how more specific emotions, like disgust (Hutcherson & Gross, 2011; Ugazio, Lamm, & Singer, 2011; Wheatley & Haidt, 2005), guilt (Finger, Marsh, Kamel, Mitchell, & Blair, 2006; Takahashi et al., 2004), and contempt (Rozin, Lowery, Imada, & Haidt, 1999), can

modify our moral judgments (Eisenberg, 2000; Pizarro & Salovey, 2002; Tangney, Stuewig, & Mashek, 2007). While partaking in immoral behavior can generate guilt or shame, watching others behave immorally incites reactions that give rise to moral appraisals and judgments, like contempt and disgust (Moll et al., 2002a). Each of these distinct moral emotions can promote prosocial behavior. For instance, guilt – a negative evaluation of a behavior that violates moral rules or principles – can motivate people to make amends when they have done something wrong (Chang, Smith, Dufwenberg, & Sanfey, 2011; Giner-Sorolla & Espinosa, 2011). Similarly, exhibiting disgust or contempt (the moral condemnation of others) implements social policing as ultimately, the harming agent is ostracized and morally offensive behavior is minimized (Tangney, Miller, Flicker, & Barlow, 1996).

In the same way that guilt impels people to repair relationships, empathy can be seen as a moral motivator because it ‘moves us to partake of the pleasure or pain, the happiness or misery, of others’ (Davis, 1900). Empathy, the capacity to have feelings that reflect the emotional dynamics of another’s situation (Hoffman, 2000), is a multidimensional psychological concept (Zaki & Ochsner, 2012) composed of distinct, yet related, cognitive and affective processes (Batson, Fultz, & Schoenrade, 1987; Davis, 1983; Penner et al., 2005; Preston & De Waal, 2001). One study in which subjects were trained to become more compassionate and empathic resulted in increased altruistic helping (Leiberg, Klimecki, & Singer, 2011). Similarly, subjects who were part of a group – and thus felt greater positive empathy towards other group members – helped to reduce another member’s distress by choosing to endure the physical pain themselves (Hein, Silani, Preuschoff, Batson, & Singer, 2010). Indeed, individuals high in trait levels of empathic concern report greater harm aversion (Crockett, Clark, Hauser, & Robbins, 2010), and this general tendency towards expressing empathic concern for another in distress predicts greater altruistic behavior (FeldmanHall et al., accepted for publication). Together, this work demonstrates how the ability to know, understand, and empathize with what another is feeling can be a strong motivation to help others (Decety, 2011).

Neural Architecture of Moral Cognition

Neuroscientists – predominantly through the use of neuroimaging methods such as functional magnetic resonance imaging (fMRI) and MRI lesion mapping – have identified a set of brain regions that map onto a ‘moral circuit.’ The earliest imaging experiments contrasted simple written statements containing moral content against statements that contained no moral content (Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Moll, Eslinger, & Oliveira-Souza, 2001; Moll

et al., 2002b). These studies revealed that moral stimuli preferentially engage the prefrontal cortex (PFC). More recently, research has begun to explore if discrete brain regions support the various psychological parameters known to influence moral cognition (i.e., emotion, reason, and social context). The assumption is that these different psychological parameters can be functionally localized – and dissociated – within the brain. Indeed, moral neuroscientists have found that cognitive load (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008), semantic content (Takahashi et al., 2004), emotional arousal and regulation (Decety, Michalska, & Kinzler, 2011; Moll & De Oliveira-Souza, 2007), probability (Shenhav & Greene, 2010), intent (Berthoz et al., 2002; Young & Saxe, 2011), and harm (Kedia et al., 2008), to name a few, are associated with distinct activation patterns within the wider moral network.

The Prefrontal Cortex

Through the course of evolution, the capacity for sophisticated reasoning has progressively developed into highly complex and well-tuned cognitive skills. The PFC – which can be subdivided into many subcomponents including the ventromedial PFC (vmPFC), dorsolateral PFC (dlPFC), and orbital frontal cortex (Figure 1) – is considered to underlie higher-order cognition. The cortical expansion of the PFC has permitted humans to engage in more sophisticated capacities, such as abstract reward and punishment contingencies (O’Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001), self-regulation (Raine & Yang, 2006), and top-down control (Mansouri, Tanaka, & Buckley, 2009). By logical extension, it is unsurprising then that moral decision making, which requires the ability to process and integrate highly complex and conflicting stimuli, consistently engages the PFC.

Broadly speaking, the vmPFC and dlPFC demonstrate reciprocal connectivity with different posterior and subcortical regions of the brain, which, in turn, impacts each region’s functionality (Wood & Grafman, 2003). The dlPFC – which receives inputs from temporal and parietal regions and projects to motor regions (Ongur & Price, 2000) – is most closely associated with processes that include cognitive control and deliberative thinking (Mansouri et al., 2009) like executing motivated selfish behavior (FeldmanHall, Dalgleish,

Thompson, Evans, Schweizer and Mobbs, 2012). The dlPFC has also been implicated in processing abstract moral principles (Moll, De Oliveira-Souza, Bramati, & Grafman, 2002) such as sensitivity to unfairness (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). One seminal study showed that when individuals contemplate what they would do during the trolley dilemma (i.e., flipping a switch that reroutes an out of control trolley from a track with five people to a track that only has one person), the dlPFC preferentially engages (Greene et al., 2001). This is a different neural result from when subjects contemplate actively pushing another person onto the tracks to stop the trolley, a more ‘emotional’ response and one that engages the vmPFC.

The vmPFC plays an integral role in social cognition and moral decision making and has shown to be especially crucial for monitoring internal states and motivations (Forbes & Grafman, 2010) like self–other reflection (Van Overwalle, 2009) and perspective taking (Amodio & Frith, 2006). The imaging literature has further emphasized the vmPFC and its role in encoding socioemotional knowledge, such as stereotypes (Gozzi, Raymond, Solomon, Koenigs, & Grafman, 2009) and moral emotions, like pride (Tangney et al., 2007), embarrassment (Zahn et al., 2008), and guilt (Moll et al., 2011). The vmPFC also engages when individuals select the clearly prosocial option in a relatively easy moral dilemma (FeldmanHall, Mobbs, & Dalgleish, 2013) and when they make rapid decisions during emotionally evocative moral challenges (Greene et al., 2001). Similarly, violations of social norms (Berthoz et al., 2002) and adherence to social norms (i.e., greater altruistic behavior) (FeldmanHall, Dalgleish, et al., 2012) preferentially engage the vmPFC. Together, these findings suggest that as individuals navigate through a complex and social world, the vmPFC plays a critical role in evaluating conflicting moral phenomena and assigning moral value (Preston & De Waal, 2011).

Cingulate Cortex and Amygdala

The cingulate cortex and amygdala (Figure 1) – seemingly integral structures for successful moral responding – have extensive connections with one another (Derubeis, Siegle, & Hollon, 2008) and directly project to the anterior portions of

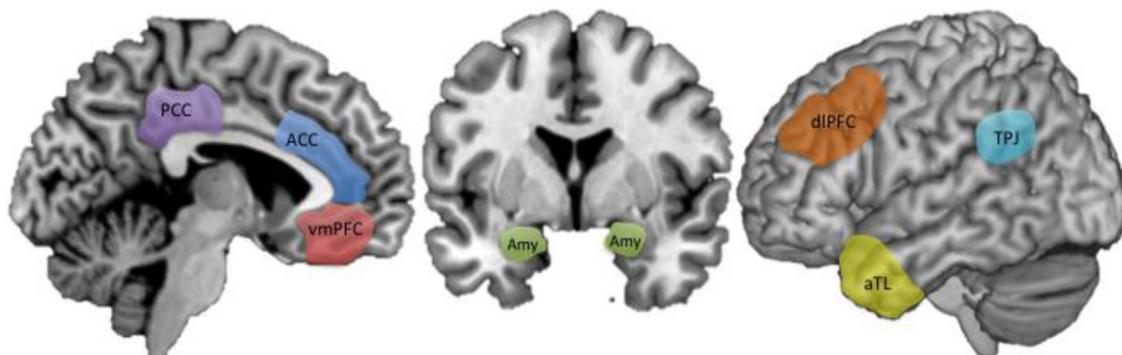


Figure 1 The brain’s ‘moral network’: comprising the prefrontal cortex, which includes the dorsolateral PFC (dlPFC) and ventral medial PFC (vmPFC); the temporal lobes, which includes the anterior temporal lobe (aTL) and temporoparietal junction (TPJ); the cingulate cortex, which is subdivided into the posterior cingulate cortex (PCC) and the anterior cingulate cortex (ACC); and the amygdala (Amy).

the PFC (Van Hoesen, Pandya, & Butters, 1972). The cingulate cortex is divided into the anterior, middle, and posterior cingulate cortices (PCC) and does not exhibit uniform functionality (Devinsky, Morrell, & Vogt, 1995). The anterior cingulate cortex (ACC) has been broadly implicated in a number of cognitive functions that influence moral decision making, including emotional responsiveness (Ochsner & Gross, 2005) and conflict monitoring (Kerns et al., 2004) – especially for difficult moral dilemmas (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Kahane et al., 2011). Research further demonstrates that the ACC engages during evaluations of moral transgressions that involve the self and others (Kedia et al., 2008), possibly indexing the associated feelings of guilt (Shin et al., 2000). The ACC also activates when subjects make moral decisions that result in real outcomes, compared with decisions made in purely hypothetical contexts (FeldmanHall, Dalgleish, et al., 2012). In contrast, the PCC has been more closely linked to reasoning through moral judgments (Greene et al., 2001), hypothetical moral choices (FeldmanHall, Dalgleish, et al., 2012), and simple moral decisions (as opposed to more complex, semantic ones) (Heekeren et al., 2003).

Like the cingulate, the amygdala is anatomically organized into distinct regions that are divided along the basis of evolutionarily newer regions (basolateral) and evolutionarily more primitive areas (corticomедial) (Ledoux, 2007). While the amygdala consistently engages when processing socially relevant stimuli (Adolphs, Tranel, & Damasio, 1998), research has found inconsistent activation of the amygdala for explicit moral phenomena (Raine & Yang, 2006). Some work has illustrated that the amygdala engages for processing the consequences of moral actions (Decety et al., 2011), especially for harmful stimuli like physical and emotional pain. Interestingly, such harmful stimuli – and the moral emotions that arise from situations involving interpersonal harm – do appear to reliably engage the amygdala (Berthoz, Grezes, Armony, Passingham, & Dolan, 2006; Decety & Porges, 2011; Decety et al., 2011; Kedia et al., 2008; Luo et al., 2006). However, the fact that the amygdala is not consistently activated across different types of moral phenomena – especially in contrast to the PFC, which is consistently active across a myriad of moral tasks – suggests that the amygdala may be more sensitive to a particular subset of moral decision-making parameters. For example, imaging studies exploring moral phenomena differ widely in their task requirements (i.e., passive versus active) and emotional engagement. Thus, it is possible that unlike the PFC, the amygdala may only engage when there is salient harm or active engagement during the task (FeldmanHall, Dalgleish, et al., 2012).

The Temporal Lobes

More recently, research has illustrated that the temporal lobes, and in particular the anterior temporal lobe (aTL) and temporoparietal junction (TPJ) (Figure 1), play significant roles in guiding moral cognition. The aTL is thought to encode the understanding of abstract conceptual representations (Moll, De Oliveira-Souza, & Zahn, 2008) – such as ‘honorable’ or ‘brave’ (Zahn et al., 2007) – an ability that allows moral phenomena to be conceptualized across a wide range of

situations (Tse, 2008). That the aTL has been implicated in synthesizing semantic representations and social conceptual information (Lambon Ralph, Pobric, & Jefferies, 2009) dovetails nicely with the evidence that damage to this area can cause inappropriate social behavior (Moll et al., 2008). The TPJ, which is situated posteriorly to the aTL, generally behaves in a similar way; its functions include decoding social cues (i.e., facial expressions and bodily movements) (Saxe & Kanwisher, 2003) and agency and intentionality (i.e., theory of mind) (Moll et al., 2008). This fits with the evidence that disrupting TPJ engagement, through either lesion or transcranial magnetic stimulation, results in the diminished capacity to understand another’s beliefs (Samson, Apperly, Chiavarino, & Humphreys, 2004; Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010).

A recent study comparing effortful and cognitively demanding moral decisions (such as suffocating your baby in order to save the rest of your family from being discovered by enemy soldiers) with rapid, easy moral decisions found increased activation in the TPJ (FeldmanHall, Mobbs, & Dalgleish, 2013). This finding accords with the proposal not only that the TPJ is crucial for attending to, shifting between, and weighing up salient and competing stimuli (Mitchell, 2008) but also that the TPJ is integral for understanding another’s intentions (Young, Cushman, Hauser, & Saxe, 2007). Individuals diagnosed with autism, who are known to have difficulty attributing mental states to others, demonstrate abnormal TPJ activity (Baron-Cohen et al., 1999) and even sometimes fail to activate the TPJ altogether (Castelli, Frith, Happe, & Frith, 2002). One recent study found that, the TPJ was relatively deactivated in individuals who have trouble identifying and experiencing emotions (a subclinical personality construct called alexithymia) (Sifneos, 1973) when observing another in distress (FeldmanHall, Dalgleish, & Mobbs, 2013).

A Clinical Perspective

Disturbances in socioemotional responding – observed to varying degrees across multiple clinical populations including psychopaths, alexithymics and frontal lobe patients – can significantly affect social behavior (FeldmanHall, Dalgleish, & Mobbs, 2013). For example, psychopathy is characterized by a profound lack of empathy, guilt, and remorse (Hare, Hart, & Harpur, 1991) – emotions that are critical for dictating how individuals experience interpersonal dynamics. It is thought that a psychopath’s inability to behave morally is related to a lack of empathic concern or a reduced ability to emotionally respond to the observation of distress in others (Blair, Peschardt, Budhani, Mitchell, & Pine, 2006) – abilities that are attributed to the functional integrity of the frontal lobe (Blair, 2007). Early PFC damage, specifically to the anterior PFC, alters the development of social cognition and social emotions, creating deficits in emotional and moral behavior (Eslinger et al., 2009). Research has also found significant reduced PFC gray matter volume in patients diagnosed with psychopathy (Weber, Habel, Amunts, & Schneider, 2008). A recent fMRI study found that the network subserving moral judgment, and in particular the PFC, is relatively underactive in psychopathic individuals when deliberating over moral

dilemmas (Pujol et al., 2011). In addition to the PFC, dysfunction in the amygdala has also been linked to psychopathy (Blair, 2007). Psychopaths exhibiting flagrant immoral behavior without remorse show reduced activity in the amygdala in response to emotional words (Blair, 2007), and another study correlated smaller amygdala volume with increasing degree of psychopathy traits in violent offenders (Blair, 2003).

Evidence from frontal lobe lesions strongly parallels the findings of moral impairment in psychopathy and continues to indicate that robust vmPFC functioning is necessary for effective moral decision making. The earliest and perhaps most famous case comes from Phineas Gage, who in 1848 survived an accident that almost completely obliterated his left frontal lobe, which resulted in somewhat antisocial and morally deficient behavior (Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994). Other patients with lesions to the vmPFC (Damasio, 1994) commonly illustrate blunted affect, irrational responses to minor provocation, and irritable, angry, and abusive behavior (Koenigs & Tranel, 2007). Research shows that compared with healthy individuals, vmPFC patients judge moral violations as more acceptable (Ciarra, Muccioli, Ladavos, & Di Pellegrino, 2007; Young et al., 2010). One study found that patients with bilateral vmPFC lesions made more utilitarian judgments in high-conflict moral dilemmas (Koenigs et al., 2007) – that is, conflicts where emotional reactions would be likely to motivate a choice incongruent with an ideally ‘rational’ analysis. It is thought that because these patients are unable to harness emotional responsiveness to highly aversive stimuli (e.g., pushing a man in front of a trolley in the trolley dilemma), they make more utilitarian decisions because they have no internal emotional signal indicating that it is aversive to push a man to his death.

Together, the clinical data support the proposal that moral cognition relies on a qualitatively distinct emotional signal. When emotions – which are arguably central to social behavior (Huebner et al., 2009) – fail to develop or cannot be deployed, callous, antisocial behavior can ensue. This could be explained by the theory that a negative emotion can act as an antecedent to a moral decision, alerting the individual to the moral salience of a situation (Decety et al., 2011) and indicating that a certain action should be taken. Without this reflex-like emotional precursor, individuals may have difficulty registering that a relevant event is occurring as they lack the requisite capacity that ‘emotions are, in essence, impulses to act’ (Goleman, 1995).

Conclusions

What has emerged from the growing literature is that moral cognition arises from a complex, diverse, and decentralized neural network that selectively engages distinct brain regions depending on the associated emotions, framework, and context. The vmPFC appears intimately linked with emotional arousal (Harenski & Hamann, 2006; Moll et al., 2007); the TPJ seems to subserve the capacity to think about another’s perspective when deliberating over a moral dilemma (FeldmanHall, Dalgleish, & Mobbs, 2013; Young et al., 2007;

Young & Saxe, 2008; Young, Scholz, & Saxe, 2011); and the amygdala emerged as a region sensitive to salient stimuli and, in particular, harmful actions (Anderson & Phelps, 2001; Berthoz et al., 2006; Blair, 2007; FeldmanHall, Dalgleish, et al., 2012; Finger et al., 2006; Harenski, Harenski, Shane, & Kiehl, 2010; Kedia et al., 2008).

Inevitably, questions remain. Is moral cognition the result of a bottom-up neural process where higher-order rational processes, presumably within the PFC, are modulated by lower-order socioaffective regions (Naqvi, Shiv, & Bechara, 2006)? Or do moral decisions arise from a top-down modulation of subcortical networks (Moll, Zahn, De Oliveira-Souza, Krueger, & Grafman, 2005)? It is also possible that moral cognition is the result of a bidirectional process where a number of specialized neural regions (limbic, prefrontal, and temporal) interact with one another in a reciprocal – and, by extension, influential – manner. To more completely characterize the psychological and neurobiological determinants of moral cognition, future work needs to address these questions.

Issues relating to morality, honesty, integrity, and virtue have been, and will continue to be, relevant throughout humanity’s tenure. This is largely because the issues central to morality revolve around our capacity for successful social living. And as a fundamentally social species, our moral faculty is paramount to our interpersonal experiences. To that end, moral research can inform us about which psychological processes most strongly motivate prosocial behavior, which, in turn, has significant implications for both the legal and clinical domains.

References

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–474.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, *7*, 268–277.
- Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*, 305–309.
- Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., et al. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *The European Journal of Neuroscience*, *11*, 1891–1898.
- Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, *55*, 19–39.
- Berthoz, S., Armony, J. L., Blair, R. J., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, *125*, 1696–1708.
- Berthoz, S., Grezes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one’s own moral violations. *NeuroImage*, *31*, 945–950.
- Blair, R. J. (2003). Neurobiological basis of psychopathy. *The British Journal of Psychiatry*, *182*, 5–7.
- Blair, R. J. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*, 387–392.
- Blair, R. J., Peschardt, K. S., Budhani, S., Mitchell, D. G., & Pine, D. S. (2006). The development of psychopathy. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*, 262–276.
- Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI study. *Journal of Cognitive Neuroscience*, *5*, 803–817.
- Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*, 1839–1849.
- Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, *70*, 560–572.

- Ciaramelli, E., Muccioli, M., Ladavas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2, 84–92.
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17433–17438.
- Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam, NY: Penguin Books.
- Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science*, 264, 1102–1105.
- Davis, N. K. (1900). *Elements of psychology*. Boston, MA: Silver, Burdett & Company.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113–126.
- Decety, J. (2011). The neuroevolution of empathy. *The Annals of the New York Academy of Sciences*, 1231, 35–45.
- Decety, J., Michalska, K. J., & Kinzler, K. D. (2011). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral Cortex*, 22, 209–220.
- Decety, J., & Porges, E. C. (2011). Imagining being the agent of actions that carry different moral consequences: An fMRI study. *Neuropsychologia*, 49, 2994–3001.
- Derubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Opinion – Cognitive therapy versus medication for depression: Treatment outcomes and neural mechanisms. *Nature Reviews: Neuroscience*, 9, 788–796.
- Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain*, 118(Pt 1), 279–306.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, 51, 665–697.
- Eslinger, P. J., Robinson-Long, M., Realmuto, J., Moll, J., Deoliveira-Souza, R., Tovar-Moll, F., et al. (2009). Developmental frontal lobe imaging in moral judgment: Arthur Benton's enduring influence 60 years later. *Journal of Clinical and Experimental Neuropsychology*, 31, 158–169.
- FeldmanHall, D., Dalgleish, T., & Mobbs, D. (2013). Alexithymia decreases altruism in real social decisions. *Cortex*, 49(3), 899–904.
- FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs. hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, 7(7), 743–751.
- FeldmanHall, O., Mobbs, D., & Dalgleish, T. (2013). Deconstructing the brain's moral network: Dissociable functionality between the temporoparietal junction and ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 9(3), 297–306.
- FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navardy, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition*, 123(3), 434–441.
- Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *NeuroImage*, 33, 414–421.
- Foot, P. (1978). *The problem of abortion and the doctrine of the double effect in virtues and vices*. Oxford: Basil Blackwell.
- Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience*, 33, 299–324.
- Giner-Sorolla, R., & Espinosa, P. (2011). Social cuing of guilt by anger and of shame by disgust. *Psychological Science*, 22, 49–53.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- Gozzi, M., Raymond, V., Solomon, J., Koenigs, M., & Grafman, J. (2009). Dissociable effects of prefrontal and anterior temporal cortical lesions on stereotypical gender attitudes. *Neuropsychologia*, 47, 2125–2132.
- 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.
- Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology*, 100, 391–398.
- Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, 30, 313–324.
- Harenski, C. L., Harenski, K. A., Shane, M. S., & Kiehl, K. A. (2010). Aberrant neural processing of moral violations in criminal psychopaths. *Journal of Abnormal Psychology*, 119, 863–874.
- Hauser, M. D. (2006). The liver and the moral organ. *Social Cognitive and Affective Neuroscience*, 1, 214–220.
- Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, 14, 1215–1219.
- Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68, 149–160.
- Hoffman, M. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge, UK: Cambridge University Press.
- Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13, 1–6.
- Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functional account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, 100, 719–737.
- Isen, A., & Levin, P. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*, 21, 384–388.
- Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2011). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, 7(4), 393–402.
- Kedia, G., Berthoz, S., Wessa, M., Hilton, D., & Martinot, J. L. (2008). An agent harms a victim: A functional magnetic resonance imaging study on specific moral emotions. *Journal of Cognitive Neuroscience*, 20, 1788–1798.
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303, 1023–1026.
- Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *Journal of Neuroscience*, 27, 951–956.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (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*. Chicago, IL: Rand McNally.
- Kohlberg, L., & Hersh, R. (1977). Moral development: A review of the theory. *Theory Into Practice*, 16, 53–59.
- Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex*, 19, 832–838.
- Ledoux, J. (2007). The amygdala. *Current Biology*, 17, R868–R874.
- Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS One*, 6, e17798.
- Luo, Q., Naki, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. (2006). The neural basis of implicit moral attitude – An IAT study using event-related fMRI. *NeuroImage*, 30, 1449–1457.
- Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nature Reviews: Neuroscience*, 10, 141–152.
- Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, 18, 262–271.
- Moll, J., & De Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences*, 11, 319–321.
- Moll, J., De Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, 16, 696–703.
- Moll, J., De Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., et al. (2002a). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22, 2730–2736.
- Moll, J., De Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., et al. (2002b). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience*, 22, 2730–2736.
- Moll, J., De Oliveira-Souza, R., Garrido, G. J., Bramati, I. E., Caparelli-Daquer, E. M., Paiva, M. L., et al. (2007). The self as a moral agent: Linking the neural bases of social agency and moral sensitivity. *Society for Neuroscience*, 2, 336–352.

- Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition: Sentiments, concepts, and values. *Annals of the New York Academy of Sciences*, 1124, 161–180.
- 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. *Arquivos de Neuro-Psiquiatria*, 59, 657–664.
- Moll, J., Zahn, R., De Oliveira-Souza, R., Bramati, I. E., Krueger, F., Tura, B., et al. (2011). Impairment of prosocial sentiments is associated with frontopolar and septal damage in frontotemporal dementia. *NeuroImage*, 54, 1735–1742.
- Moll, J., Zahn, R., De Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). Opinion: The neural basis of human moral cognition. *Nature Reviews. Neuroscience*, 6, 799–809.
- Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science*, 15, 260–264.
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9, 242–249.
- O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, 4, 95–102.
- Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, 10, 206–219.
- Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, 56, 365–392.
- Pizarro, D. A., & Salovey, P. (2002). On being and becoming a good person: The role of emotional intelligence in moral development and behaviour. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education*. San Diego, CA: Academic Press.
- Preston, S. D., & De Waal, F. B. (2001). Empathy: Its ultimate and proximate bases. *The Behavioral and Brain Sciences*, 25, 1–20 discussion 20–71.
- Preston, S. D., & De Waal, F. (2011). Altruism. In J. Decety & J. Cacioppo (Eds.), *The handbook of social neuroscience*. New York, NY: Oxford University Press.
- Pujol, J., Batalla, I., Contreras-Rodríguez, O., Harrison, B., Pera, V., Hernandez-Ribas, R., et al. (2011). Breakdown in the brain network subserving moral judgment in criminal psychopathy. *Social Cognitive and Affective Neuroscience*, 7(8), 917–923.
- Raine, A., & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience*, 1, 203–213.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76, 574–586.
- Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7, 499–500.
- 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.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people – The role of the temporo-parietal junction in 'theory of mind'. *NeuroImage*, 19, 1835–1842.
- Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, 67, 667–677.
- Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., et al. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry*, 48, 43–50.
- Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, 22, 255–262.
- Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: An fMRI study. *NeuroImage*, 23, 967–974.
- Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, 70, 1256–1269.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, 58, 345–372.
- Tse, P. U. (Ed.). (2008). *Symbolic thought and the evolution of human morality*. Cambridge, MA: The MIT Press.
- Ugazio, G., Lamm, C., & Singer, T. (2011). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion*, 12(3), 579–590.
- Van Hoesen, G. W., Pandya, D. N., & Butters, N. (1972). Cortical afferents to the entorhinal cortex of the rhesus monkey. *Science*, 175, 1471–1473.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30, 829–858.
- Weber, S., Habel, U., Amunts, K., & Schneider, F. (2008). Structural brain abnormalities in psychopaths – A review. *Behavioral Sciences & the Law*, 26, 7–28.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16, 780–784.
- Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews. Neuroscience*, 4, 139–147.
- 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, 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. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 6753–6758.
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 8235–8240.
- Young, L., & Saxe, R. (2008). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, 21(7), 1396–1405.
- Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, 120, 202–214.
- Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for 'intuitive prosecution': The use of mental state information for negative moral verdicts. *Social Neuroscience*, 6, 302–315.
- Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 6430–6435.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., et al. (2008). The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex*, 19(2), 276–283.
- Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, 15(5), 675–680.