AJOB Neuroscience, 6(3): 41–46, 2015

ISSN: 2150-7740 print / 2150-7759 online
DOI: 10.1080/21507740.2015.1058866

Target Article

How the Mind Matters for Morality

Alek Chakroff, Harvard University
Liane Young, Boston College

Attributing minds to people and reasoning about the contents of those minds are crucial components of moral judgment and social interaction. This article provides a review of recent work seeking to illuminate the psychological and neurobiological processes that guide human moral cognition. First, we review the role of social cognitive processes in moral judgment, including the role of mental state reasoning or theory of mind (ToM)—reasoning about people's beliefs, intentions, and motivations. Second, we explore how social cognitive processes such as ToM are deployed for different kinds of moral judgments, supporting the proposal that distinct moral norms are associated with distinct adaptive functions. Third, we examine not only how people represent others' beliefs and intentions but also how people's own moral beliefs influence their actual behavior. Finally, we conclude with a discussion of how understanding the psychological and neural processes that guide human moral cognition can contribute to bioethics more broadly.

Keywords: morality, theory of mind, social cognition, cooperation, competition, meta-ethics

If you hear about an explosion causing death and destruction, you might ask: tragic accident or act of terrorism? If you find you weren't invited to a friend's baby shower, you might wonder: accidental omission or purposeful slight? Serious events and ordinary occurrences alike prompt us to consider the mental states of moral agents, both friends and enemies. When you discover the blast was due to a bomb, you might react with outrage and not simply grief. When you realize your invitation got lost in the mail, you might feel sheepish and buy your friend an especially nice gift.

This article highlights three sets of questions about human moral cognition. First, we review the role of social cognitive processes in moral judgment, including the role of mental state reasoning or theory of mind (ToM)—reasoning about people's beliefs, intentions, and motivations. For example, we ask, what is the role of intent information in moral judgment? How do people make moral judgments when they lack information about intent? Second, we explore how social cognitive processes such as ToM are deployed for different kinds of moral judgments. Does intent information matter more for some judgments than for others? Are distinct moral norms (e.g., norms against harmful actions vs. norms against taboo behaviors) associated with distinct adaptive functions? Third, we examine not only how people represent others' beliefs and intentions but also how people's own moral beliefs influence their actual behavior. What is the precise relationship between beliefs—about the self, about specific values, about broader meta-ethical claims—and behavior? This article concludes with a discussion of how understanding the psychological and neural processes that guide human moral cognition can contribute to bioethics more broadly.

THE ROLE OF INTENT INFORMATION IN MORAL JUDGMENT

Legal institutions distinguish between murder and manslaughter, primarily based on the mind and intentions of the person responsible for a death. When determining friend or foe, it is not enough to evaluate agents on the basis of their external, observable actions; moral judgment depends on an assessment of internal mental states. In our work, we have investigated mental state reasoning or theory of mind (ToM) for moral cognition across different populations. Typically, we have done so by providing participants with hypothetical scenarios describing agents, their intentions, their actions, and the outcomes of their actions. So, for example, in one scenario, an agent accidentally poisons her friend after mistaking poison for sugar. In another version of this scenario, an agent attempts but fails to poison her friend after mistaking sugar for poison. Innocent intentions in the case of accidents decrease blame, whereas malicious intentions even in the absence of actual harm increase blame (Cushman 2008; Inbar, Pizarro, and Cushman 2012; Young et al. 2007). Notably, the failure to process emotionally salient intentions, that is, malicious

Address correspondence to Liane Young, Department of Psychology, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. E-mail: liane.young@bc.edu

intentions, results in abnormally lenient judgments of attempted harms including attempted murder, as observed in individuals with focal lesions to the ventromedial prefrontal cortex (VMPFC), a brain region implicated in social–emotional processing (Young, Bechara, et al. 2010). Mental state information can inform moral judgments of not only individuals but also entire groups of people (e.g., corporations, unions; Waytz and Young 2012).

Functional magnetic resonance imaging (fMRI) studies on the neural basis of ToM have consistently implicated a network of brain regions including the medial prefrontal cortex (MPFC), right and left temporo-parietal junction (RTPJ, LTPJ), and precuneus (Dodell-Feder et al. 2011). Among these brain regions, the RTPJ supports the initial encoding of mental state information and its integration with task-relevant information (e.g., about actions and outcomes) for moral judgment (Young and Saxe 2008). At the time of integration, the magnitude of the RTPJ response is correlated with moral judgment (Young and Saxe 2009a). In one fMRI study, participants who made harsh outcome-based judgments of accidents (e.g., she poisoned her friend) had lower RTPJ responses, whereas participants who made more lenient belief-based judgments (e.g., she thought it was sugar) had higher RTPJ responses. Our ability to forgive or deliver exculpatory judgments depends on the neural mechanisms that allow us to consider, in the face of adverse consequences, another person's innocent intentions and false beliefs.

Although the RTPJ encodes intent information in the context of moral judgments, it is robustly recruited for moral judgments of intentional and accidental harms—its response is high for both. Is there evidence that the RTPJ distinguishes between intentional and accidental harms? In another series of experiments, we used a more sophisticated technique for analyzing fMRI data called multivoxel pattern analysis —we examined not just the overall magnitude of response in specific brain regions (as we did in the preceding) but also the spatial pattern of neural activity across voxels (three-dimensional units of brain space) within brain regions, including the RTPJ (Koster-Hale et al. 2013). We did this in order to determine how specific brain regions represent harmful actions as intentional or accidental. Indeed, we found that the spatial pattern of voxels within the RTPJ contains information about whether a harmful act is intentional or accidental. The spatial pattern of activity across voxels within the RTPJ differentiates between intentional and accidental harms, and, importantly, individual differences in participants' neural discriminability correlate with individual differences in participants' behavioral responses—the extent to which participants distinguish between intentional and accidental harms in their moral judgments. Convergent findings using electroencephalography (EEG) revealed that activity stemming from RTPJ can distinguish intentional from accidental harms 62 ms after stimulus presentation (Decety and Cacioppo 2012).

These findings reveal dedicated neural circuitry for representing the mental states of moral agents. But the information detected using fMRI or EEG cannot tell us whether the RTPJ is causally necessary for mental state-based moral judgments. To address this causal question, we used a neuromodulatory technique called transcranial magnetic stimulation (TMS) to transiently disrupt activity in participants' RTPJs as they read moral scenarios and made moral judgments (Young, Camprodon, et al. 2010). Recall the preceding scenario in which the agent maliciously attempted but failed to poison her friend after mistaking sugar for poison. What we found was a subtle but systematic effect on moral judgment—when activity in the RTPJ was disrupted, participants made more outcome-based rather than intent-based moral judgments. They viewed the failed attempt to poison as more morally permissible—no harm, no foul. Taken together with the neuroimaging findings described earlier, these results reveal that the intent information encoded in the RTPJ is causally linked to moral judgments.

Another approach to the causal question is to examine individuals with impairments in mental state reasoning. Individuals with autism spectrum disorders (ASD)—individuals known to have impairments in social cognition, including reasoning about the mental states of others (Moran et al. 2011)—deliver more outcome-based moral judgments in the case of accidental harms, basing their judgments more on the bad outcome than on the innocent intention. They are more likely to report, for example, that it is morally forbidden for someone to accidentally poison her friend. Moreover, the pattern of voxels within their RTPJs did not reliably discriminate between intentional and accidental harms (in contrast to neurotypical participants) (Koster-Hale et al. 2013). Consistent with the findings described earlier, these findings suggest that the atypical functioning of the RTPJ in ASD is involved in the atypical, outcome-based moral judgments observed in ASD. Scenarios describing accidents may pose a particular challenge because they pit salient information about bad outcomes against neutral information about false beliefs.

Forgiving accidents may be challenging not simply for individuals with cognitive impairments but also for typically developing individuals—forgiving accidents may require an especially robust representation of the agent's innocent intent in order to overcome a prepotent response to a very salient aspect of the situation: the bad outcome (cf. Miller et al. 2010). Recall too our finding of individual differences in the extent to which participants forgive agents who cause harm accidentally. Remarkably, we found that the individuals most capable of "forgiveness," that is, ignoring accidental outcomes in favor of "hyperrational," intent-based judgments, were those with impaired emotional processing and a clinical diagnosis of psychopathy (Young, Koenigs, et al. 2012). "Forgiveness" in this case may be due to a blunted emotional response to the harmful outcome, rather than an especially robust mental state representation. These results highlight the multiple mechanisms that may interact to guide moral

judgment, that is, ToM and emotional processing (cf. Young, Bechara et al. 2010), consistent with evidence for the role of emotion in moral cognition.

The simple point that mental states matter for moral judgment may appear uncontroversial. Thus, our research has also examined key cases in which intentions appear to count less—first, when we assign blame for accidents in spite of the actor's innocent intentions, and second, when explicit intent information is absent. Our work shows that even in these cases, moral judgments depend crucially on mental state inferences. Imagine that a neighbor feeds your peanut-allergic child a peanut butter sandwich; she heard you say "almonds" and her false belief leads to your child's death. Now imagine a scenario in which your neighbor had held a true belief resulting in no harm being done. Although a key difference between these two scenarios is the difference between life and death (outcome), another difference is whether the belief is true or false (mental state). But imagine a third scenario in which your neighbor holds a false belief—she believes, falsely, that your child is allergic to almonds; she prepares the sandwich, but it gets eaten by the dog, so no harm is done. Participants assigned substantial blame in this scenario (Young, Nichols, and Saxe 2010); they assessed the false belief as unjustified, and it was this assessment of negligence that led to moral condemnation.

Second, when information about mental states isn't accessible, do people base their judgments on observable actions and outcomes? Mounting work reveals that people spontaneously infer mental states for moral judgment. The RTPJ is recruited for morally relevant versus morally irrelevant facts about an action (Young and Saxe 2009b). Participants may be motivated to infer mental states—for example, did she know about the allergy? But how might participants infer mental states? Behavioral research suggests that information about an agent's moral character or prior record informs assessments of the agent's harmful and helpful actions as intentional or unintentional (e.g., Knobe 2004). Research combining behavioral and neural measures offers convergent evidence. In one fMRI study, participants interacted with "other players" who behaved fairly or unfairly in an economic game (Kliemann et al. 2008). Participants then read, in the scanner, a series of stories, presented as written by the players about their past actions (e.g., broke roommate's lamp, shrunk friend's sweater); intent information was absent. Participants judged harmful actions performed by previously unfair players as more blameworthy and more intentional. These judgments were associated with increased RTPJ activity, reflecting inferences of blameworthy intent based on negative prior record. Additional fMRI studies reveal broadly similar patterns in which increased RTPJ activity reflects inferences of negligence (Young, Nichols, and Saxe 2010) and negative intent or the absence of positive intent (Young, Scholz, and Saxe 2011).

In sum, moral neuroscience has identified a network of brain regions that support mental state reasoning for moral judgment. Notably, the RTPJ rapidly and automatically encodes intent information in the context of moral judgments of accidental and intentional harm. Critically, the intent information encoded in the RTPJ is causally important for moral judgment: When RTPJ activity is disrupted via TMS or impaired in participants with ASD, intentions are assigned less weight in moral judgments. The work reviewed thus far supports the view that reasoning about the minds and, specifically, the intentions of others is central to moral cognition (Gray, Young, and Waytz 2012). People even infer nefarious intentions where they are absent, particularly for agents who cause bad outcomes or who are seen as having poor moral character.

DISTINCT DOMAINS OF MORALITY AND DISTINCT MORAL JUDGMENTS

The picture we have painted so far suggests that moral judgments of interpersonal harms rely on information about agents' mental states—though there are clear individual and group differences in the degree of this reliance on mental states. Do mental states also matter more for judgments of some categories of moral issues and less for others? We all recognize that manslaughter is a far cry from murder, but do we feel the same about other unintentional versus intentional behaviors that aren't obviously harmful—eating taboo foods (e.g., rat meat) or performing taboo sexual acts (e.g., incest, bestiality)? Taboo behaviors or "purity" violations are often condemned even in the absence of victims—when the agents themselves are the only ones who are directly affected by their actions (Graham et al. 2011; Haidt, Koller, and Dias 1993). Typically, we react to victimless violations with disgust, whereas we react to interpersonal harms with anger. Furthermore, purity violations such as incest elicit strong disgust reactions regardless of the context of the act or the intent of the agent (e.g., Russell and Giner-Sorolla 2013). Do intentions matter less for moral judgments of "impure" versus harmful acts? Is the perceived moral difference between murder and manslaughter greater than that between intentional and accidental incest, for example?

We have begun to investigate whether people deploy social cognitive capacities, including theory of mind (ToM), differently for different kinds of moral judgments, judgments of harmful actions, and judgments of actions that violate purity norms (e.g., consensual incest, consumption of taboo foods)—victimless violations that appear to defile or contaminate the actors themselves. This investigation serves a broader one: Is there evidence for distinct moral domains? As hypothesized, we found that mental states matter more for moral judgments of harmful versus impure actions (Young and Saxe 2011): Participants perceived a large moral difference between intentional and accidental harms, compared to purity violations (e.g., knowingly vs. unknowingly sleeping with a long-lost sibling). Participants also delivered harsher moral judgments of failed attempts to harm others (based on false beliefs and guilty intentions) than failed attempts to commit

incest (e.g., sleeping with someone falsely believed to be a sibling).

We have proposed an adaptive account for this cognitive difference (Young and Tsoi 2013): Distinct moral norms serve distinct functions—for regulating interpersonal relationships, versus for protecting the self. Harm norms (i.e., don't harm others) are aimed at limiting people's negative impact on each other. Indeed, paradigmatic cases of harm feature at least one agent (the violator) who harms at least one patient (the victim) (Gray et al. 2012). The victim may demand an explanation from the violator, who might appeal to innocent intentions. Information about intent supports not only explanations and evaluations of other people's past actions, but also reliable predictions of their future behavior. Typically, only knowing a person's true intentions can afford an accurate identification of friend or foe. By contrast, purity norms against sleeping with blood relatives or eating taboo foods may have evolved as a means for us to protect ourselves, for our own good, from possible contamination. Researchers have proposed that disgust reactions, elicited by purity violations, evolved for pathogen avoidance and food rejection (Russell and Giner-Sorolla 2013; Schaller and Park 2011; Tybur et al. 2013). When we worry about negatively impacting ourselves, we may care less about whether the impact is accidental or intentional; the key is to avoid the contamination. Thus, purity violations like consensual incest or eating taboo foods may be deemed morally offensive even in the absence of victims. Often, impure acts directly affect only the actors themselves. This account is consistent with mounting work associating person-based evaluations with purity violations. People judge impure agents to have a corrupted moral character (Uhlmann and Zhu 2014), and even judge impure agents to be the primary cause of their acts (rather than the situation; Chakroff and Young 2015).

We have found initial support for the link between harmful actions and other-focus, on the one hand, and impure or defiling actions and self-focus, on the other. Harmful actions and other-directed actions elicit more anger, whereas impure actions and self-directed actions elicit more disgust. Intent matters more not only for moral judgments of harmful versus impure actions but also more for moral judgments of other-directed versus self-directed actions (Chakroff, Dungan, and Young 2013). Other work on attitudes toward suicide, the ultimate self-directed harm, offers convergent evidence. In this work, moral judgments of suicide were correlated with (1) endorsement of purity morals, (2) ratings of disgust in response to obituaries of individuals who committed suicide, and (3) judgments that these individuals had tainted the purity of their souls (Rottman, Kelemen, and Young 2014). Moral judgments of suicide were uncorrelated with harm concerns (e.g., harm to others, God), in contrast to moral judgments of homicide. Although conservative, religious participants judged suicide as more immoral, compared to liberal, secular participants, all participants perceived suicide to be wrong insofar as they perceived suicide to be a purity violation. At an explicit level, participants reported that their moral judgments of suicide were based on assessments of harm, indicating that participants do not always have conscious access to the reasons for their judgments (Cushman, Young, and Hauser 2006; Nisbett and Wilson 1977).

THE IMPACT OF MORAL BELIEFS ON BEHAVIOR

Much of moral psychology has focused on how people deliver judgments of others. Our recent work targets the impact of people's moral beliefs on their own behavior. In an initial demonstration, we primed some participants to think of themselves as good moral people by asking them to write about their recent good deeds; others wrote about neutral events or bad deeds (Young, Chakroff, and Tom 2012). Participants whose positive self-concept had been reinforced were nearly twice as likely to donate money to charity. Furthermore, within the good deeds condition, participants who did not mention being appreciated or unappreciated by others, that is, earning (or not earning) reputational credit, were the most likely to donate money. Thinking of ourselves as good people who do good for goodness's sake may lead to subsequent good behavior. Meanwhile, on the flip side, we have also found that people are more likely to behave badly when they engage in strategies to perceive their own bad behavior as more permissible, for example, when people use indirect speech to make unethical propositions (e.g., bribes). Indirect speech use enhances speakers' perceptions of their own behavior and increases their likelihood of making unethical propositions (Chakroff et al. 2014).

In another demonstration of the impact of moral beliefs on moral behavior, we primed participants with specific moral values—fairness versus loyalty (Waytz, Dungan, and Young 2013). Participants were instructed to write an essay about either the value of fairness over loyalty or the value of loyalty over fairness. Participants who had written pro-fairness essays were more likely to blow the whistle on unethical actions committed by other members of their community. Participants who had written pro-loyalty essays were more likely to keep their mouths shut in solidarity.

In a final demonstration, we primed participants with broader meta-ethical views (Young and Durwin 2012). We primed participants to adopt either *moral realism*, the view that moral propositions (e.g., murder is wrong) can be objectively true or false, similar to mathematical facts, or *moral antirealism*, the view that moral propositions are subjective and generated by the human mind. The participants were passersby primed by a street canvasser who in the realism condition asked "Do you agree that some things are just morally right or wrong, good or bad, wherever you happen to be from in the world?" and in the antirealism condition asked "Do you agree that our morals and values are shaped by our culture and upbringing, so there are no absolute right answers to any moral questions?" Participants primed with moral realism were twice as likely to donate money. Moral rules perceived as "real"

may be more psychologically costly to break; people may be more sensitive to possible punishment by peers, a divine being, or even themselves. After all, people are highly motivated to think of themselves as good moral people.

In a related project, we have asked whether, in the absence of experimental primes, moral propositions are spontaneously processed more like objective facts (e.g., 2 + 2 = 4) or subjective preferences (e.g., chocolate is better than vanilla; Theriault et al. n.d.). We scanned participants as they viewed statements about morals, facts, and preferences, designed to elicit three levels of agreement among participants (high/mid/low agreement). First, participants rated morals, overall, as more preference-like than fact-like. Furthermore, the neural profile for morals emerged as more similar to that for preferences versus facts, across brain regions for ToM in terms of both the spatial pattern and the overall magnitude of activity. Statements about morals and preferences alike may invite inferences about the mind of the speaker. Second, high-agreement morals were rated as more fact-like than mid-/low-agreement morals, and, similar to facts, elicited reduced neural activity across brain regions for ToM. High-agreement morals, like facts, may provide less social information. Reduced activity in precuneus, RTPJ, and LTPJ also tracked with ratings of morals as fact-like, and RTPJ activity in particular mediated the perception of high-agreement morals as more fact-like. Finally, temporarily disrupting activity in the RTPJ using TMS led to participants' perception of mid-agreement morals as more fact-like. Together, these results reveal the neural signatures of meta-ethical beliefs.

BIOETHICAL IMPLICATIONS

The work described in the preceding has focused on the role of social cognitive processes for third-party moral judgment and the influence of people's beliefs on their own behavior. Some have argued that an increasing understanding of people's moral psychology ought to directly inform moral philosophy (e.g., Greene 2013). This article has instead taken a descriptive approach, focusing on elucidating when minds and mental states matter for moral judgments and behavior, and when they do not. While our adaptive account (e.g., Young and Tsoi 2013) aims to explain differences in moral judgments across different kinds of moral issues (e.g., harm versus purity violations), it need not speak to the normative status of particular behaviors (e.g., suicide), or approaches to moral judgment (e.g., weighting intent information).

However, we do note that a large number of controversial topics in bioethics concern "purity violations." For example, acts such as suicide, cloning, body modification, sexual reassignment, birth control, and human enhancement are considered unnatural by some (Chakroff et al. 2013; Rottman et al. 2014; Russell and Giner-Sorolla 2011). While condemnations of such acts could be cashed out in terms of the potential harm caused to possible victims (Gray et al. 2012), the original causes of the condemnations may be driven not by perceived harmfulness of the act, but instead by intuitions that the moral agent is disgusting or tainted (Graham et al. 2011; Haidt et al. 1993; Rottman et al. 2014; Russell and Giner-Sorolla 2011). Thus, disagreements about the ethical propriety of such acts may not reflect different opinions about how best to reach a single moral goal, such as the reduction of harm and suffering, but instead reflect different weights given to distinct moral concerns such as harm versus purity (Graham et al. 2011). The literature reviewed here suggests that purity concerns are ultimately geared toward protecting the self (e.g., from pathogens), rather than increasing others' welfare (Young and Tsoi 2013). Bioethical debates regarding the moral status of putatively victimless acts could be informed by the ultimate functions that our moral intuitions serve.

Finally, as the reviewed studies show, our morality may be malleable. Scientists can change the way people make moral decisions—how to judge others and how to behave toward others—using, for example, financial incentives, experimental primes, and neuromodulatory techniques such as TMS. While some might worry that malleable morals aren't morals at all, we conclude with two remarks. First, we think that the ways in which people make moral decisions, just like all sorts of other decisions, should surely depend on context—whether people are responding to interpersonal harms or victimless violations, or to issues of fairness or issues of loyalty. Second, information about any unwanted forces on moral psychology, as well as positive influences, can only facilitate moral self-improvement—in the domain of judgment and behavior alike. Future work would do well to focus on how people engage in and improve upon their moral cognition as active participants of the complex social world.

REFERENCES

Chakroff, A., J. Dungan, and L. Young. 2013. Harming ourselves and defiling others: What determines a moral domain? *PLoS ONE* 8(9): e74434. doi:10.1371/journal.pone.0074434

Chakroff, A., K.A. Thomas, O. S. Haque, and L. Young. 2014. An indecent proposal: The dual functions of indirect speech. *Cognitive Science* 2014: 1–13. doi:10.1111/cogs.12145

Chakroff, A., and L. Young. 2015. Harmful situations, impure people: An attribution asymmetry across moral domains. *Cognition* 136: 30–37.

Cushman, F. 2008. Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 1082: 353–380.

Cushman, F., L. Young, and M. Hauser. 2006. The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm. *Psychological Science* 17(12): 1082–1089. doi:10.1111/j.1467-9280.2006.01834.x

Decety, J., and S. Cacioppo. 2012. The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology* 10811: 3068–3072.

Dodell-Feder, D., J. Koster-Hale, M. Bedny, and R. Saxe. 2011. fMRI item analysis in a theory of mind task. *Neuroimage* 552: 705–712.

Graham, J., B. A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P. H. Ditto. 2011. Mapping the moral domain. *Journal of Personality and Social Psychology* 1012: 366.

Gray, K., L. Young, and A. Waytz. 2012. Mind perception is the essence of morality. *Psychological Inquiry* 23(2): 101–124. doi:10.1080/1047840X.2012.651387

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

Haidt, J., S. H. Koller, and M. G. Dias. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 654: 613.

Inbar, Y., D. A. Pizarro, and F. Cushman. 2012. Benefiting from misfortune when harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin* 381: 52–62.

Kliemann, D., L. Young, J. Scholz, and R. Saxe. 2008. The influence of prior record on moral judgment. *Neuropsychologia* 46: 2949–2957. doi:10.1016/j.neuropsychologia.2008.06.010

Knobe, J. 2004. Intention, intentional action and moral considerations. *Analysis* 64282: 181–187.

Koster-Hale, J., R. Saxe, J. Dungan, and L. Young. 2013. Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences of the United States of America* 110(14): 5648–5653. doi:10.1073/pnas.1207992110

Miller, M. B., W. Sinnott-Armstrong, L. Young, et al. 2010. Abnormal moral reasoning in complete and partial callosotomy patients. *Neuropsychologia*, 48: 2215–2220. doi:10.1016/j

Moran, J., L. Young, R. Saxe, et al. 2011. Impaired theory of mind for moral judgment in high functioning autism. *Proceedings of the National Academy of Sciences of the United States of America* 108: 2688–2692. doi:10.1073/pnas.1011734108

Nisbett, R. E., and T. D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 843: 231.

Rottman, J., D. Kelemen, and L. Young. 2014. Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition* 130 (2): 217–226. doi:10.1016/j.cognition.2013.11.007

Russell, P. S., and R. Giner-Sorolla. 2011. Moral anger, but not moral disgust, responds to intentionality. *Emotion* 112: 233.

Russell, P. S., & Giner-Sorolla, R. 2013. Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. *Psychological bulletin* 139: 328.

Schaller, M., and J. H. Park. 2011. The behavioral immune system and why it matters. *Current Directions in Psychological Science* 202: 99–103.

Theriault, J., A. Waytz, L. Heiphetz, and L. Young. n.d. Perceptions of morals as subjective or objective rely on theory of mind. Manuscript submitted for publication.

Tybur, J. M., D. Lieberman, R. Kurzban, and P. DeScioli. 2013. Disgust: Evolved function and structure. *Psychological Review* 1201: 65.

Uhlmann, E. L., and L. Zhu. 2014. Acts, persons, and intuitions person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science* 5: 279–285.

Waytz, A., J. Dungan, and L. Young. 2013. The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology* 49: 1027–1033. doi:10.1016/j.jesp.2013.07.002

Waytz, L., and L. Young. 2012. The group-member mind tradeoff: attributing mind to groups versus group members. *Psychological Science* 23: 77–85. doi:10.1177/0956797611423546

Young, L., A. Bechara, D. Tranel, et al. 2010. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron* 65: 845–851. doi:10.1016/j.neuron.2010.03.003

Young, L., J. Camprodon, M. Hauser, A. Pascual-Leone, and R. Saxe. 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. doi:10.1073/pnas.0914826107

Young, L., A. Chakroff, and J. Tom. 2012. Doing good leads to more good: The reinforcing power of a moral self-concept. *Review of Philosophy and Psychology* 3(3): 325–334. doi:10.1007/s13164-012-0111-6

Young, L., F. Cushman, M. Hauser, and R. Saxe. 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 (20): 8235–8240. doi:10.1073/pnas.0701408104

Young, L., and A. J. Durwin. 2013. Moral realism as moral motivation: The impact of meta-ethics on everyday decision-making. *Journal of Experimental Social Psychology* 49: 302–306.

Young, L., M. Koenigs, M. Kruepke, and J. Newman. 2012. Psychopathy increases perceived moral permissibility of accidents. *Journal of Abnormal Psychology* 21(3): 659–667. doi:10.1037/a0027489

Young, L., S. Nichols, and R. Saxe. 2010. Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology* 1: 333–349. doi:10.1007/s13164-010-0027-y

Young, L., and R. Saxe. 2008. The neural basis of belief encoding and integration in moral judgment. *NeuroImage* 40: 1912–1920. doi:10.1016/j.neuroimage.2008.01.057

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

Young, L., and R. Saxe. 2009b. An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience* 21: 1396–1405.

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

Young, L., J. Scholz, and R. Saxe. 2011. Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts. *Social Neuroscience* 6: 302–315. doi:10.1080/17470919.2010.529712

Young, L., and L. Tsoi. 2013. When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass* 7(8): 585–604. doi:10.1111/spc3.12044