

## The Mechanics of Moral Judgments

By Liane Young

Tags: Emotions, Ethics, fMRI, Judgment, Morality, Neuroscience



If you realize you never received an invitation to your friend's housewarming party, you might wonder — accidental omission or purposeful slight?

If you turn on the news and discover that an explosion close to home has caused death and destruction, a question likely to cross your mind is — tragic accident or terrorist act?

We spend a great deal of time trying to decipher what's going on inside the heads of our friends, our enemies, and other people around us. The inferences we make about people's beliefs and motivations shape our moral judgments.

When you discover that the explosion wasn't simply a manhole cover explosion but the result of a carefully placed bomb, you might react not only with grief but with moral outrage. When you realize your party invitation simply got sent to the wrong address, you

might feel sheepish about your earlier doubts and buy your friend an especially nice gift.

Brain imaging technology is now revealing the neural mechanisms that underpin the moral judgments we make about others' intentions and actions. When people evaluate others' actions, a certain brain region — the right temporo-parietal junction (RTPJ) — shows an especially interesting pattern, we've found.

Using functional magnetic resonance imaging (fMRI), my research team has scanned healthy, college-aged students while they read a series of scenarios in which protagonists *accidentally* cause harm. One scenario, for example, describes a person who hurt her friend by serving her poison that she had mistaken for sugar. Is this understandable, or unforgivable?

In one study, some of our participants made harsh judgments about these types of accidents, pointing directly to the bad outcome. Others judged the situations more leniently because the people depicted didn't mean to cause harm.

The RTPJ responds robustly during all moral calculations, but the intensity of that response depends on the type of judgments made. In our study, those who made harsh, outcome-based judgments of accidents (e.g., she *poisoned* her friend) had lower RTPJ responses, whereas those who made more lenient belief-based judgments (e.g., she *thought* it was sugar) had higher RTPJ responses.

This indicates that our ability to forgive depends on the neural mechanisms that allow us to consider, in the face of harmful consequences, another person's innocent mistakes and benign intentions.

But *how* exactly does the RTPJ tell intentionally inflicted harm from accidents?

In another series of experiments, we used a more sophisticated technique for analyzing fMRI data called multi-voxel pattern analysis. MVPA allows us to see not only where, but how, brain activity changes in response to certain cues.

Using this approach, we found that specific patterns in the RTPJ indeed allow a person to *identify* harmful actions as being either deliberate or inadvertent. In addition, the more the RTPJ discriminates between intentional and accidental harms, the more that information determines the individual's moral conclusion.

But how critical is the RTPJ to this process? Are there other neural routes to such judgments?

Answering this question involves disrupting activity in the RTPJ and observing how moral judgment changes. To this end, we used a technique called transcranial magnetic stimulation (TMS) to disrupt activity in participants' RTPJs as they read and then considered the moral issues of different scenarios. In a flip of the scenario mentioned earlier, the participants read about a person who maliciously attempted but failed to poison her friend after mistaking sugar for poison. In this case, we found a subtle but systematic effect on moral judgment — participants formed more outcome-based rather than intent-based opinions. They viewed the failed attempt to poison as more morally tolerable — no harm, no foul.

In another approach to the causal question, we set out to examine individuals with specific impairments in reasoning about others' intentions. We tested high-functioning individuals with autism spectrum disorders (ASD) — individuals known to have impairments in social cognition, including reasoning about the mental states of others. Compared to neurotypical participants, those with ASD delivered 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 were more likely to say, for example, that it was morally forbidden for the person to accidentally poison her friend. Moreover, when we scanned a different sample of participants with ASD, we found that the activity within their RTPJs did *not* discriminate between intentional and accidental harms (in striking contrast to our neurotypical participants). These findings suggest that the atypical functioning of the RTPJ in ASD is involved in the atypical, outcome-based moral judgments observed in ASD.

Interestingly, our recent work on individuals with psychopathy reveals another route to "forgiving" accidents. Participants with impaired emotional processing and a clinical diagnosis of psychopathy were even more likely to "forgive" accidental harms, compared to healthy control participants. Such individuals have a blunted emotional response to the harmful outcome, rather than an especially strong read on someone's mental state.

### Interpersonal Harms Versus Victimless Violations

Do mental states also matter more for some categories of moral judgments and less for others? We all recognize that manslaughter is a far cry from murder, but do we feel the same about other behaviors that aren't so obviously harmful — eating culturally taboo

foods or performing socially proscribed sexual acts (e.g., incest)? Taboo behaviors or “purity” violations are often condemned even in the absence of clear victims — when the agents themselves are the only ones who are directly affected by their actions. Typically, we react to victimless violations with disgust, whereas we react to interpersonal harms with anger. Purity violations such as incest can disgust us regardless of the context or the intent of the people involved. While people tend to see a moral difference between murder and manslaughter, they make less of a distinction between incest that occurs accidentally (say by two strangers who don’t know they’re related) and intentionally.

Why might we put less weight on intentions when judging impure acts? Rules against eating taboo foods or committing incest may have evolved as a means for us to protect ourselves from possible contamination. In contrast, norms against harmful actions may have evolved to regulate our impact on one another. In the case of accidents, knowing someone’s true intentions helps us reliably predict the person’s future behavior, leading to either forgiveness or condemnation. In short, norms against harms govern how we act toward *others*; norms against purity violations govern how we behave toward *ourselves*.

This theory finds support in a recent series of experiments in our lab. Those studies showed that people react with anger to deviant actions directed at others (regardless of whether they are damaging or impure), but view self-directed actions as disgusting. Moreover, moral judgments of other-directed violations (splashing either sterile urine or painfully hot water on someone else) rely on intent information to a greater extent than do moral judgments of self-directed violations (splashing the same fluids on oneself). More recently we have examined moral attitudes toward suicide, the ultimate self-harm. We have found that people perceive suicide to be immoral insofar as they see it as tainting the soul. However, they *think* they judge it to be immoral because it causes harm (for example, to friends and family left behind). Our ongoing work extends this broad approach to interpersonal purity violations in which the victim herself may be blamed, as in the case of rape in cultures of honor.

### The Impact of Moral Beliefs on Moral Behavior

Much of the work in moral psychology, including our own work on the role of mental states, has focused on how people deliver judgments of others. Moral psychologists are now beginning to examine the impact of our moral beliefs on our own moral behavior. Recently, we identified three cases in which altering people’s beliefs — about specific moral values, about whether morality is “real,” and about one’s own moral character — alters people’s actual moral behavior.

In one demonstration, we primed participants with specific moral values — fairness versus loyalty. We instructed participants to write either an essay about the value of fairness over loyalty or an essay about the value of loyalty over fairness. Subsequently, participants who had written pro-fairness essays were more likely to engage in fair behavior — in this case, 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 another demonstration, we focused participants’ attention not on specific moral values like loyalty or fairness but on broader metaethical views. We primed them 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 in this experiment 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 to a charitable organization represented by the street canvasser.

Why might a simple belief in moral realism lead to better moral behavior in this context? Moral rules that are 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, who make the right sorts of moral decisions and who behave in accordance with moral rules.

In our third demonstration, we primed some participants to think of themselves as good, moral people by asking them to write about their recent good deeds, and we asked others to write about either neutral events or their recent bad deeds. Those whose positive self-concept had been reinforced were nearly twice as likely to donate money to charity than participants in the other conditions. Furthermore, within the good deeds condition, participants who did not mention being appreciated or unappreciated by others were the most likely to donate money. Thinking of ourselves as good people who do good for goodness’s sake may lead to even more of that good behavior.

Certainly we take our moral values to be a defining feature of ourselves — a topic of ongoing investigation in our lab. But, as studies now show, our morality is somewhat malleable. We can alter moral decisions by priming people in different ways.

Is this cause for concern? Does this mean we lack a moral core? I think not. Instead, we should embrace a moral psychology that can be flexibly deployed across diverse contexts — in dealing with interpersonal harms and victimless violations, issues of fairness, and issues of loyalty. We should embrace a moral psychology that allows us to stretch our capacity as moral agents and judges — to reinforce our own good behavior and to hone our moral intuitions. Indeed, if our moral psychology is malleable, then so are we — and there is always room for improvement. This is certainly a moral psychology worth studying.

### References and Recommended Reading

Chakroff, A., Dungan, J., & Young, L. (in press). Harming ourselves and defiling others: What determines a moral domain? *PLOS ONE*, 8(9), e74434.

Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. (2013). Decoding moral judgments from neural representations of intentions. *PNAS*, 110(14), 5648–5653.

Moran, J., Young, L., Saxe, R., Lee, S., O’Young, D., Mavros, P., & Gabrieli, J. (2011). Impaired theory of mind for moral judgment in high functioning autism. *PNAS*, 108, 2688–2692.

Rottman, J., Kelemen, D., & Young, L. (in press). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition*.

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower’s dilemma and the fairness-loyalty tradeoff. *Journal of Experimental*

*Social Psychology*, 49, 1027–1033.

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

Young, L., & Durwin, A. (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., Koenigs, M., Kruepke, M., & Newman, J. (2012). Psychopathy increases perceived moral permissibility of accidents. *Journal of Abnormal Psychology*, 121(3), 659–667.

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

**Observer Vol.27, No.2 February, 2014**

---

Copyright © Association for Psychological Science