

Editorial

# Why Neuroscience Does Not Pose a Threat to Moral Responsibility

David Pizarro, Cornell University

There is perhaps no other science (with the possible exception of quantum mechanics) that wreaks havoc on our intuitions the way neuroscience does. The growing knowledge it provides about the inner workings of the brain seems to threaten some of our most deeply held intuitions about how the mind works. Of particular concern to many is that learning how the brain works will necessarily cause us to alter the way we think about human freedom and moral responsibility. If our thoughts and feelings are governed by the same basic laws that govern the rest of the physical world, so the thinking goes, how can we have free will? And if we are not free, how can we hold people morally responsible for their actions? This seems to threaten not only our everyday moral judgments, but might cause us to question our legal notions of responsibility and punishment as well (Greene and Cohen 2004). In short, this concern may have real implications for everything from our daily arguments about whom to blame to the structure of our penal code.

I do not wish to debate the normative question as to whether neuroscientific findings *actually* pose a threat to a belief in free will and moral responsibility. What I wish to argue here is that regardless of whether these concerns are rationally justified, the putative threat posed by neuroscience to our moral judgments will turn out not to be a threat at all. This, I wish to suggest, is because the mechanisms that give rise to our basic moral intuitions (such as that we should hold people responsible) are too deeply entrenched in the mind for us to abandon in any meaningful way—even if we have good reason for doing so.

Why should this be the case? For one, our judgments of moral responsibility rely on very basic features of our psychology—our perception of intentionality, causality, and agency in the acts of others. Importantly, the systems that give rise to these perceptions are features of human cognition that most likely evolved because of the advantage they provided for navigating a complex physical and social world, and as such appear resistant to much top-down influence (such as arguments about deterministic brains). As evidence of the primacy of these judgments, we not only quickly and easily attribute intentionality and agency in human action, we do things like spontaneously attribute

the random movement of shapes on a screen to the figures' underlying intentions and desires (Heider and Simmel 1944). Even infants attribute goal-directed motives to shapes that perform negative and positive behaviors (Kuhlmeier, Wynn, and Bloom 2003). At a more basic level, psychologists studying visual perception have demonstrated that we seem to be hard-wired to perceive causality and animacy in the movements of simple objects (Scholl and Tremoulet 2000). In short, if we make an error about intentionality, it tends to be strongly in the direction of overattributing intentions in entities that do not possess them (Helzer, Pizarro, and Goldstein 2010). The upshot of this is that even if we have an explicit belief to the contrary, we often cannot help but attribute the properties of agency necessary for holding an individual responsible for his or her actions.

The cognitive bias toward perceiving actions as intentional and agentic is not the only force working against the suspension of moral responsibility. Our emotional systems appear to be built in such a way that we have fairly strong and rapid negative emotional responses to moral violations, which serve in turn to further amplify the judgment that an individual acted intentionally. So rather than arrive at a judgment of responsibility and blame for an act by first determining causality, intentionality, and control (as common sense and most normative theories would dictate), we often react first with blame and condemnation, then attribute intentionality and agency to the agent despite evidence to the contrary. For example, in one study by Mark Alicke, participants were told that a man got in a car accident (injuring others) as he was speeding home in a rainstorm (Alicke 2000). When asked, participants were more likely to say that he had control over the car if he was speeding home to hide cocaine from his parents than if he was speeding home to hide an anniversary gift. This and a host of other studies by Alicke and colleagues show that spontaneous judgments of blame appear to shape judgments of causality and control, rather than the other way around. An even stronger demonstration of my basic argument comes from a recent study by Nichols and Knobe (2007). In this study, individuals were presented with a description of a causally deterministic world (described as incompatible with free will) and

were asked whether murderers in this deterministic world should be held morally responsible. Most people said no. But when presented with a specific individual who murdered his entire family, individuals were more than willing to blame him—even when it was clear from the description of the world that he could not have acted otherwise. The negative emotions arising from the concrete description of a murder, the authors argued, were strong enough to push intellectual concerns about determinism to the background. In short, if we feel strongly about a moral violation we have no problem attributing blame even when we “know” better.

So while some scholars perceive determinism to be a threat to moral responsibility, human psychology appears built to handle this threat by simply ignoring it. If anything, the body of work on this topic suggests that when forced to think explicitly about free will we tend to use judgments of freedom to justify, in a post hoc fashion, our initial judgments of moral responsibility and blame. Elsewhere, I have called this feature of human psychology *stubborn moralism* (see Pizarro and Helzer 2010 for a longer treatment of the topic).

Whether the implications of this stubborn moralism are good or bad will depend, fairly obviously, on which normative theory you endorse. If, like Greene and Cohen (2004), you see a belief in free will as anachronistic and responsible for a wrongheaded view of justice and punishment, then the argument I presented poses a problem. If, on the other hand, you believe that the putative threat to moral responsibility comes from a deep misunderstanding of freedom, determinism, moral responsibility, or neuroscience, and that our

views about moral responsibility should not change one bit, then you should feel confident that you will have no problem winning in the court of public opinion.

## REFERENCES

- Alicke, M. D. 2000. Culpable control and the psychology of blame. *Psychological Bulletin* 126: 556–574.
- Greene, J. D. and J. D. Cohen. 2004. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B* (Special Issue on Law and the Brain) 359: 1775–17785.
- Heider, F., and S. Simmel. 1944. An experimental study of apparent behavior. *American Journal of Psychology* 57: 243–259.
- Helzer, E., D. A. Pizarro, and M. Goldstein. 2010. *Intentions precede accidents: Evidence for an intentionality bias in behavioral encoding*. Poster presented at the 12th annual meeting of the Society for Personality and Social Psychology, San Antonio, TX.
- Kuhlmeier, V., K. Wynn, and P. Bloom. 2003. Attribution of dispositional states by 12-month-olds. *Psychological Science* 14: 402–408.
- Nichols, S., and J. Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, 41: 663–685.
- Pizarro, D. A., and E. Helzer. 2010. Freedom of the will and stubborn moralism. In *Free will and consciousness: How might they work?*, ed. R. F. Baumeister, A. R. Mele, and K. D. Vohs, 101–120. New York: Oxford University Press.
- Scholl, B. J., and P. D. Tremoulet. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences* 4: 299–230.