

Philosophizing the Social Brain

Action editor: Leslie Marsh

Philip Robbins

*Department of Philosophy, University of Missouri, 426 Strickland Hall, Columbia, MO
65211, United States*

Abstract: Introduction to themed issue on philosophical approaches to social neuroscience.

Keywords: Social neuroscience; philosophy; mindreading; consciousness; moral psychology

1. Introduction

The last few decades have witnessed an explosive growth of research in social neuroscience, the interdisciplinary study of the brain mechanisms and processes that underlie social cognition and behavior (Cacioppo & Berntson, 1992). Social psychology has taken a biological turn, and as a result, our understanding of ourselves as social beings is undergoing a profound transformation (Cacioppo & Patrick, 2009; Churchland, 2012, 2014; Greene, 2014; Lieberman, 2014).

In this brief essay, we consider some of the philosophical implications of this development. In particular, we look at three domains in which findings from social neuroscience have been brought to bear on philosophical issues. Our objective is twofold. First, on a positive note, we want to motivate the idea that social neuroscience — like neuroscience in general (Churchland, 1986, 2002) — matters to philosophy, in the sense that research in the field has the potential to illuminate a number of philosophical issues. At the same time, we want to warn against the idea that progress in social neuroscience obviates the need for careful philosophical reflection, or that philosophical problems can

be neatly resolved (or dissolved) by brain science. Indeed, where some areas of philosophy are concerned, a healthy dose of neuroskepticism may be the order of the day.

2. Mindreading

Nowhere in philosophy is the influence of social neuroscience more apparent than in theorizing about how we read minds — that is, how we make sense of the thoughts, feelings, and behavior of others (and ourselves). A major point of controversy in the philosophical literature is whether mindreading is driven primarily by theoretical inference, as claimed by ‘theory theory’, or by mental modeling, as per ‘simulation theory’. According to theory theorists, we arrive at an understanding of a friend’s emotional state, for example, by observing their behavior (e.g., their facial expression or tone of voice) and inferring which emotion best fits that behavior by tapping an internally represented ‘theory of mind’. By contrast, a simulation theorist might say that observing our friend’s behavior causes us to experience their emotional state first-hand, by a process of mental mirroring. At that point, the simulationist story goes, we identify the state in ourselves on the basis of introspection and then project it onto the target.

In recent years, findings from social neuroscience have played an important role in advancing our understanding of mindreading, especially where the theory versus simulation debate is concerned. For example, Goldman and Sripada (2005) note that selective impairment of face-based emotion recognition (e.g., loss of the ability to recognize fear) is paired with selective impairment of emotional experience (i.e., loss of the ability to feel fear), as predicted by simulation-based, but not theory-based, accounts of emotion recognition. In a similar vein, patients with damage to brain areas associated with the inhibition of one’s own epistemic perspective are especially prone to egocentric errors in mindreading — a neat fit with simulation theory (which posits inhibition of self-perspective as an essential component of successful third-person mindreading), but something of an embarrassment for theory theory (which assigns no special role to executive function in the mindreading process) (Goldman, 2006). The discovery of ‘mirror neurons’ in the premotor cortex of macaque monkeys — cells that fire both when a goal-directed action is performed and when another individual is seen performing an

action of the same type — also lends itself to interpretation in simulation-friendly terms, insofar as it suggests that action understanding is mediated by motor empathy (Gallese & Goldman, 1998). Though the significance of these findings, and related results from human studies, have been contested by philosophers and scientists alike (Hickok, 2009; Jacob, 2008), research on mirror neurons has helped to advance the simulationist agenda to the point where theory-based approaches to mindreading have lost much of their appeal.

Theory theory and simulation theory, however, are not the only games in town. Both views have come in for criticism on grounds that they fail to capture the essentially interactive character of everyday social cognition, focusing instead on mindreading from a detached, spectatorial point of view (Gallagher, 2001). According to interaction theory, the primary mode of social understanding involves adopting a second-person, rather than a third-person, stance towards others — that is, actively engaging with them in a reciprocal social exchange, rather than passively observing their behavior from a distance. This alternative theoretical paradigm has inspired a new wave of research investigating the neural underpinnings of mindreading in the context of social interaction (Schilbach, Costall, Bente, Schlicht, & Vogeley, 2013). Interestingly, the principal direction of influence in this case appears to be from philosophy to neuroscience, rather than the other way around (De Jaegher, Di Paolo, & Gallagher, 2010; Przyrembel, Smallwood, Pauen, & Singer, 2012).

3. Consciousness

Another area of philosophy illuminated by social neuroscience is theorizing about consciousness. Of particular interest in this area is the ‘hard problem’ popularized by Chalmers (1996), who presses the intuition that scientific accounts of phenomenal consciousness will inevitably leave out the subjective, ‘what-it’s-like’ aspect of phenomenal experience. For some philosophers, the intuition of an unbridgeable ‘explanatory gap’ between mind and brain reveals a fatal flaw in physicalist approaches to the mind-body problem and ultimately points in the direction of dualism. For others, the intuition itself cries out for explanation. This is where social neuroscience enters the

picture, in the form of evidence from neuroimaging and neuropsychology which suggests that the gap intuition may originate in the functional architecture of the brain (Jack, 2014; Robbins & Jack, 2006).

The core idea has three components: first, thinking about minds recruits activity in a functionally specialized brain network; second, thinking about brains recruits activity in a functionally specialized network distinct from the first; and third, these two networks operate in a mutually antagonistic, push-pull fashion. Thinking about something as a subject of conscious experience is associated with activation of the Default Mode Network (DMN), which comprises a set of regions including ventromedial prefrontal, posterior cingulate, and lateral parietal cortex. By contrast, thinking about something in terms of its causal-mechanical properties activates the Task Positive Network (TPN), which includes regions of the dorsal attention system and the frontoparietal control network. Since activity in the DMN tends to suppress activity in the TPN, and vice versa, thinking about consciousness and thinking about brain mechanisms tend to conflict. As a result, it becomes difficult to conceive of consciousness as a biological phenomenon, or to imagine how mere physical processes in the brain, however complex, could give rise to subjective experience.

One way of taking the philosophical import of this hypothesis for consideration of the ‘hard problem’ is by way of a debunking argument, akin to arguments from evolutionary biology against the objectivity of morality (Joyce, 2007; Street, 2006). First, there is no reason to expect that human neuroanatomy would reveal anything of significance about the metaphysics of consciousness. (There is every reason to expect that it would tell us something about the psychology of consciousness, but that is a different matter.) Hence, if the intuition of a mind-brain explanatory gap can be traced to features of our neuroanatomy, then there is no reason to give that intuition much credence. And in that case, we should be wary of arguments against physicalism about the mental which rest on such intuitions.

Whatever the merits of this account, it faces a number of challenges. One issue has to do with underspecification. What do high-level cognitive constructs like ‘thinking about consciousness’ and ‘thinking about physical mechanisms’ correspond to in terms of operations at the computational-functional level, and how are those operations realized in

specific patterns of brain activity? Pending a relatively fine-grained specification of the proposed mapping between cognitive domains and neural processes, the hypothesis that the gap intuition arises from neuroanatomical constraints will be difficult to test empirically. Relatedly, in what sense are phenomenal concepts and physical concepts intrinsically opposed, given that we appear to have no trouble thinking of ourselves (and others) as having both sorts of properties at the same time? Third, given that thinking about intentionality is associated with activity in certain regions of the DMN (e.g., the temporoparietal junction), why isn't there a comparable gap intuition regarding that aspect of the mental? Finally, with respect to philosophical debates about the metaphysics of consciousness, one might worry about the force of neuroscientific arguments aimed at debunking the gap intuition, in parallel with worries about evolutionary arguments targeting intuitions about morality (Kahane, 2010).

4. Moral psychology

Large swaths of philosophy are concerned, either directly or indirectly, with social cognition. Perhaps the clearest example of this concern comes from the domain of ethics, where philosophers have long been interested in questions about the nature of moral judgment. Of particular interest to moral philosophers is the issue of how such judgments originate in our psychology. According to the sentimentalist tradition, the source of our judgments of right and wrong is primarily emotional or affective. On this view, we judge an action to be right or wrong depending upon whether we feel approval or disapproval towards it. According to the rationalist tradition, on the other hand, judgments of right and wrong derive chiefly from reasoning: we assess the moral character of an action by thinking about whether it satisfies the dictates of morality. This contrast between sentimentalist and rationalist perspectives on moral judgment has important epistemological implications, insofar as judgments based on feeling have often been regarded as epistemically inferior to judgments based on the exercise of reason. Indeed, the idea that moral judgment is grounded in feeling, rather than thinking, has long been associated with doubts about the objectivity of morality (Prinz, 2009).

Alongside the distinction between sentimentalist and rationalist perspectives on the nature of moral judgment lies the distinction between utilitarian and deontological judgments. According to the utilitarianism, the moral status of an action is determined by whether performing the action promotes the general welfare in a maximally impartial way. According to deontological ethics, the permissibility of an action depends on whether its performance is consistent with the duty to respect the rights of individuals. The distinction between these views can be seen clearly in cases where they deliver conflicting verdicts about the permissibility of intentionally harming individuals in the service of advancing the greater good. Puzzle cases of the ‘trolley problem’ variety, in which an individual is sacrificed in order to save the lives of a group, are the standard bearer here. (In the classic example, an individual is pushed to their death from a footbridge in order to block the progress of a runaway trolley that would otherwise cause the death of several individuals in its path.) According to utilitarianism, actions of this sort are not just permitted but required, whereas deontology forbids them. Thus, utilitarian judgments differ from deontological judgments in that according to the former, but not the latter, such actions are morally appropriate.

Based on findings from a series of ingenious neuroimaging studies, Joshua Greene and colleagues have argued that divergent moral judgments about trolley-type puzzle cases reflect the functional anatomy of the social brain. In Greene’s dual-process model, inspired by the work of Kahneman and Tversky (Kahneman, 2003), utilitarian judgments about puzzle cases result from ‘cool’, controlled, analytic thinking, whereas deontological judgments reflect ‘hot’, automatic, intuitive processes. Early support for this idea came from studies showing, in the case of utilitarian responding to personal-harm scenarios, selective activation of brain areas associated with deliberative cost-benefit reasoning (e.g., the dorsolateral prefrontal cortex), and in the case of deontological responding, selective activation of areas linked to emotion processing (e.g., ventromedial prefrontal cortex, medial parietal cortex, and amygdala). Further evidence included studies of moral judgment in psychopathy (in which amygdala and ventromedial prefrontal functioning is impaired) and acquired sociopathy (due to damage to ventromedial frontal cortex), suggesting an association between utilitarian reasoning and diminished emotionality. The upshot was an uneasy rapprochement between rationalist

and sentimentalist views of moral judgment, with utilitarian responses to puzzle cases reflecting the more cognitive, rational side of the equation, and deontological responses the more affective, emotional side — an apparent epistemic win for utilitarianism (Greene, 2007).

As it turns out, however, evidence for Greene's model is mixed. Consider, for example, the idea that utilitarian judgments about puzzle cases are driven by deliberative reasoning, whereas deontological judgments are driven by emotional responding. If this were true, then individuals in whom deliberative reasoning is compromised and emotional reactivity is heightened should be less inclined to choose the utilitarian option — but a recent study of moral judgment in alcohol-intoxicated individuals suggests otherwise (Duke & Bègue, 2015). Indeed, a growing body of evidence suggests that the main source of utilitarian responses to puzzle cases is not superior reasoning, but deficits in socio-emotional processing linked to low levels of empathic concern (Côté, Piff, & Willer, 2013; Gleichgerrcht & Young, 2013; Jack, Robbins, Friedman, & Myers, 2014).

In addition to these evidential limitations, Greene's account faces problems of a more philosophical character. We'll mention just two of those problems here. First, even if utilitarian and deontological judgments are driven by controlled and automatic processes, respectively, that by itself doesn't tell us much about their epistemic credentials (Berker, 2009). The fact that automatic processes tend to be unreliable in one domain (for example, probability) doesn't show that they are unreliable in another (morality). Second, utilitarian judgments in puzzle cases need not reflect distinctively utilitarian reasoning, that is, reasoning characterized by application of the principle of utility maximization (Kahane, 2012). Indeed, utilitarian responding to puzzle cases is associated with systematic deviations from utilitarian reasoning about more quotidian situations, such as cases involving charitable giving (Kahane, Everett, Earp, Farias, & Savulescu, 2015). Hence, even if controlled processing (associated with utilitarian judgment) is more reliable in the moral domain than automatic processing is, it doesn't follow that utilitarian reasoning is epistemically superior to the deontological alternative.

The scientific and philosophical issues here are complex, and not easily summarized. Suffice it to say that, despite the initial appeal of a brain-based argument for utilitarianism, Greene's account runs into difficulties. That said, his research and its

critical reception exemplifies the kind of productive interplay between social neuroscience and philosophy which is likely to advance our understanding of morality in the future.

5. Conclusion

Philosophy has long been a socially minded field. As such, it comes as no great surprise that social neuroscience has directly and deeply influenced recent thinking about core philosophical issues, and vice versa. This strikes us as a salutary development — a sign of the progressive naturalization of philosophy, of a piece with philosophy's 'experimental turn' (Knobe, 2015). And while these are early days, there is a very real possibility that future interaction and collaboration between social neuroscience and philosophy will have transformative effects on both fields. Time will tell.

References

- Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy and Public Affairs*, 37: 293-329.
- Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain. *American Psychologist*, 47: 1019-1028.
- Cacioppo, J. T., & Patrick, W. (2009). *Loneliness: Human Nature and the Need for Social Connection*. New York: W. W. Norton.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.
- Churchland, P. S. (2002). *Brain-wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.
- Churchland, P. S. (2012). *Braintrust: What Neuroscience Tells Us About Morality*. Cambridge, MA: MIT Press.
- Churchland, P. S. (2014). *Touching a Nerve: Our Brains, Our Selves*. New York: W. W. Norton.

- Côté, S. Piff, P. K., & Willer, R. (2013). For whom do the ends justify the means? Social class and utilitarian moral judgment. *Journal of Personality and Social Psychology*, 104: 490-503.
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14: 441-447.
- Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition*, 134: 121-127.
- Gallagher, S. (2001). The practice of mind: Theory, simulation, or primary interaction? *Journal of Consciousness Studies*, 8: 83-108.
- Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences*, 2: 493-501.
- Gleichgerricht, E., and Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLOS ONE*, 8: e60418.
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
- Goldman, A. I., and Sripada, C. (2005). Simulationist models of face-based emotion recognition. *Cognition*, 94: 193-213.
- Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality* (pp. 35-79), MA: MIT Press.
- Greene, J. D. (2014). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin.
- Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21: 1229-1243.
- Jack, A. I. (2014). A scientific case for conceptual dualism: The problem of consciousness and the opposing domains hypothesis. In T. Lombrozo, J. Knobe, , and S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy* (Vol. 1). Oxford: Oxford University Press.
- Jack, A. I., Robbins, P., Friedman, J. F., & Myers, C. D. (2014). More than a feeling: Counterintuitive effects of compassion on moral judgment. In J. Sytsma (ed.), *Advances in Experimental Philosophy of Mind* (pp. 125-79). London: Bloomsbury.

- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language*, 23: 190-223.
- Joyce, R. (2007). *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kahane, G. (2010). Evolutionary debunking arguments. *Noûs*, 45: 103-125.
- Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind and Language*, 27: 519-545.
- Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., and Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, 134: 193-209.
- Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, 58: 697-720.
- Knobe, J. (2015). Philosophers are doing something different now: Quantitative data. *Cognition*, 135: 36-38.
- Lieberman, M. (2014). *Social: Why Our Brains Are Wired to Connect*. New York: Crown.
- Przyrembel, M., Smallwood, J., Pauen, M. & Singer, T. (2012). Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in Human Neuroscience*, 6: 190.
- Prinz, J. J. (2009). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Robbins, P., & Jack, A. I. (2006). The phenomenal stance. *Philosophical Studies*, 127: 59-85.
- Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127: 109-66.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36: 393-462.