

## **Introduction, *Advances in Experimental Philosophy of Mind***

Justin Sytsma

Broadly understood, experimental philosophy involves the use of empirical methods to help answer philosophical questions while seriously engaging with the surrounding philosophical literature. More narrowly, experimental philosophy is often described as the use of empirical methods to study intuitions about philosophically interesting scenarios. While I advocate for a broad understanding of experimental philosophy, in general, there is no denying that the narrow conception does a better job of forming a coherent (if selective) whole out of the disparate methods, projects, goals, and motives that animate experimental philosophy.<sup>1</sup> This is especially clear with regard to experimental philosophy of mind, where we find both a central core focused on intuitions about conscious mental states and a periphery that extends out into a wide range of issues.

Articulating experimental philosophy of mind in such a way that it forms a coherent whole might be thought to be especially important, as the unifying threads of the sub-field threaten to snap on a broad understanding. One issue is that while it is often difficult to draw a firm dividing line between experimental philosophy and work in the sciences tackling topics of philosophical interest, this difficulty is quite pronounced for experimental philosophy of mind. Brain scientists are often concerned with questions that are of deep interest to philosophers of mind—questions about the nature of the mind, mental states, consciousness, and how these are brought about by, or otherwise relate to, the brain—and they often engage (more or less seriously) with at least parts of the surrounding philosophical literature. The result is that a good deal of scientific work on topics like consciousness, attention, theory of mind, etc., could quite

reasonably be included under the label of ‘experimental philosophy of mind’. Doing so, however, would make it easy to overlook the interesting and important projects that self-proclaimed experimental philosophers have been primarily concerned with. And it is these projects that are the central focus of this volume.

Following the broad/narrow distinction noted above, one way to restrict the domain of experimental philosophy of mind would be to focus on studies investigating people’s intuitions. And, in fact, much of what is typically thought of as experimental philosophy of mind can be construed in this way (although it depends somewhat on how one defines ‘intuitions’). In line with this approach, in the following chapter Jennifer Nado considers recent arguments concerning the use of intuitions as evidence in philosophy. She targets how these arguments relate to claims about intuitions in philosophy of mind, focusing on the use of thought experiments in the literature. For example, Nado discusses Ned Block’s (1978) nation of China thought experiment. Block asks us to imagine the population of China working together to simulate the functioning of a normal human brain. He then argues that according to functionalist accounts of the mind, such a system would have the full spectrum of mental states that you or I have; but, Block claims that our intuitions tells us otherwise—they tells us that the nation of China lacks mental states all together. Is Block correct in taking his intuition about this scenario to be widespread, though? And, if so, does this intuition constitute good evidence against functionalist theories? As we will see, the work of experimental philosophers casts some light on these questions.

With regard to Block’s thought experiment, the intuition at issue is about whether mental states can be correctly ascribed to a particular group agent (the nation of China). And, much of the experimental philosophy of mind literature has focused on intuitions of this sort:

Experimental philosophers of mind have been especially concerned with describing and explaining the mental state ascriptions that people make. In fact, the label ‘experimental philosophy of mind’ has primarily been applied to research focusing on questions raised by two seminal articles exploring lay mental state ascriptions—Knobe and Prinz (2008) and Gray, Gray, and Wegner (2007). In particular, researchers have explored how lay people classify different mental states and how such classification relates to their moral judgments.<sup>2</sup>

One way to motivate these core issues for experimental philosophy of mind is to note that while psychologists have done a great deal of work on theory of mind in recent years—work on how people ascribe mental states to themselves and others—this research has largely focused on ascriptions of mental states like beliefs and desires, and their role in reasoning about and predicting agentive behavior. There are another set of mental states, however, that have been at the forefront of philosophical debates about the mind since at least the early modern period—states like feeling pains, seeing colors, hearing sounds, and so on. Philosophers have often taken such subjective experiences, to employ the terminology used by Sytsma and Machery (2010), to be quite special. These states are thought to be phenomenally conscious: In brief, it is thought that in contrast to mental states like beliefs and desires, there is ‘something it is like’ (Nagel, 1974) to have subjective experiences, where this is often cashed out in terms of these mental states having distinctive phenomenal qualities (the pain felt, the color seen).

Having noted the common philosophical distinction between phenomenally conscious mental states and other mental states, several experimental philosophers have asked whether non-philosophers draw this distinction. More carefully, experimental philosophers have explored whether lay people tend to ascribe mental states to entities in ways that are consistent with the philosophical distinction. Responses that are consistent with this distinction are then taken as

evidence that lay people have something like the philosophical concept of phenomenal consciousness, while responses that are inconsistent with this distinction are taken to suggest against the hypothesis. For example, Knobe and Prinz (2008) present evidence suggesting that lay people do in fact have the philosophical concept of phenomenal consciousness. In one study, Knobe and Prinz gave participants a series of ten sentences ascribing different mental states to a group agent (Acme Corporation). They found that participants rated each of the five sentences ascribing mental states that philosophers typically take to be phenomenally conscious as sounding less natural than each of the five sentences ascribing mental states that philosophers typically take to be non-phenomenal. Based on these results, Knobe and Prinz argue both that lay people distinguish between mental states that are and are not phenomenally conscious, and that ascriptions of phenomenally conscious mental states depend on more than just functional properties—the composition of the entity matters as well.

Of course, Knobe and Prinz's latter conclusion is germane to Block's nation of China thought experiment, as Nado notes. Suppose, for the moment, that Block is correct and that his intuitions about the nation of China are widely shared. What should we conclude from this? If Knobe and Prinz are correct, perhaps not as much as we might have thought—it might be that our reluctance to ascribe phenomenally conscious mental states to group agents is driving the intuition that the nation of China lacks mental states all together.

In Chapter 3, Wesley Buckwalter and Mark Phelan also consider Block's nation of China thought experiment. They call on this example to illustrate the common philosophical claim that a necessary condition for an entity having mental states (or having certain sorts of mental states) is that it is made of the right kind of stuff. As we saw above, Knobe and Prinz defend a related claim—they argue that lay ascriptions of subjective experiences are sensitive to the type of body

that an entity has. Buckwalter and Phelan challenge this embodiment hypothesis, extending their previous defense of analytic functionalism as an account of lay mental state ascriptions (Buckwalter and Phelan, 2012; Phelan and Buckwalter, forthcoming). They present the results of five new experiments that test the willingness of lay people to ascribe a certain type of subjective experience (emotional states) to disembodied ghosts and spirits. The results of their studies suggest against the embodiment hypothesis.

While Buckwalter and Phelan challenge the second conclusion drawn by Knobe and Prinz (2008), another line of research challenges the first conclusion—that lay people have the philosophical concept of phenomenal consciousness. Building off of their (2009) critique of Knobe and Prinz, Sytsma and Machery (2010) argue that a common philosophical claim is mistaken: It is often claimed that the existence of phenomenally conscious mental states is obvious from first-person experience with states like seeing red and feeling pain. If this is correct, however, then we should expect that lay people will tend to treat these states similarly. For example, we would expect them to deny that a simple non-humanoid robot could be in either state. This is not what Sytsma and Machery found, however: While lay people tend to deny that such a robot feels pain, they tend to affirm that it sees red.

A number of objections have been raised against this work, but the most common is to argue that the phrase ‘sees red’ can be understood in two ways—in an informational sense (the entity makes the relevant discriminations between visual stimuli) or in a phenomenal sense (the entity is in the corresponding phenomenally conscious mental state).<sup>3</sup> It is then argued that in Sytsma and Machery’s studies, lay participants tend to adopt the informational rather than the phenomenal understanding. For example, an objection of this sort is given by Fiala, Arico, and Nichols (forthcoming) in defending their agency model of mental state ascriptions—the view

that lay mental state ascriptions result from a dual-process cognitive system, where one process focuses on cues such as having facial features, displaying interactive behavior, and moving with distinctive trajectories in determining whether an entity is a suitable target for mental state ascriptions (see Arico et al., 2011; Fiala, Arico, and Nichols, 2011).

Fiala, Arico, and Nichols present evidence suggesting that while lay people tend to understand phrases like ‘sees red’ in the phenomenal sense, participants in Sytsma and Machery’s studies answered that the robot saw red because this was the only way that they could convey that the robot performed relevant information-processing involving color detection. A response to Fiala and colleagues version of the informational/phenomenal objection is offered in Sytsma (forthcoming-b). In addition, in Sytsma (2010b) I offered a general response to this objection based on the hypothesis that lay people do not generally hold the view of colors that the objection presupposes—they do not typically associate colors with mental states, but take them to be mind-independent qualities of things in the world. I then presented empirical results supporting this hypothesis, as well as preliminary evidence that lay people also tend to adopt a similar view of pains—they tend to think of pains as being features of body parts rather than mental states.

In Chapter 4, Kevin Reuter, Dustin Phillips, and Justin Sytsma build off of this work, as well as the related results in Reuter (2011), to argue against a common view of the ordinary concept of pain—that pains are private (no one can feel anybody else’s pain), subjective (there are no unfelt pains), and that pain hallucinations are impossible. While Sytsma (2010b) and Reuter (2011) present evidence that lay people do not tend to hold that pains are necessarily private or subjective, they did not investigate lay views concerning pain hallucinations. Reuter and colleagues fill that gap in their contribution to this volume, presenting the results of a series

of new studies that indicate that the ordinary concept of pain does in fact allow for the possibility of pain hallucinations.

In addition to the two conclusions just discussed, Knobe and Prinz (2008) also argue that ascriptions of phenomenally conscious mental states are used to do more than to reason about and predict agentic behavior—they hold that they also play a role in facilitating moral judgments. A similar hypothesis is found in the work of Robbins and Jack (2006) and Gray, Gray, and Wegner (2007). The basic idea is that lay people distinguish between subjective experiences and other mental states, and that ascriptions of the former are linked to judgments of moral patiency (that morally right or wrong action can be done to the entity), while ascriptions of the latter are linked to judgments of moral agency (that an entity is capable of morally right or wrong action). The new research presented in Chapters 5 and 6 expand on this work linking mental state ascriptions with moral judgments.

In Chapter 5, Jordan Theriault and Liane Young review the literature on mental state ascriptions and moral judgments, interpreting it from the perspective of Dennett's discussion of the intentional stance and the physical stance (Dennett, 1987). Following Robbins and Jack (2006), Theriault and Young argue that a further stance is needed—the phenomenal stance—that is involved in ascriptions of subjective experiences. Specifically, Theriault and Young argue that the intentional stance alone is unable to explain the sense of moral concern that humans feel for some entities, and they point to the phenomenal stance as being critical to such attributions of moral standing. At the same time, following the work of Sytsma and Machery (2012b), Theriault and Young suggest that moral concern might not be fully explicable in terms of the phenomenal stance either, with the intentional stance also playing a role in attributions of moral standing.

In Chapter 6, Anthony Jack, Philip Robbins, Jared Friedman, and Chris Meyers build off of the distinction between the intentional stance and the phenomenal stance presented in Robbins and Jack (2006) to argue that moral cognition involves two types of psychological processes. Unlike the view put forward by Joshua Greene (e.g., Greene, 2007), however, they don't characterize these processes in terms of the contrast between reason and passion; instead, Jack and colleagues characterize the processes in terms of a contrast between two types of reasoning—between a cognitive mode that evolved for interacting with inanimate objects and a cognitive mode that evolved for interacting with conscious agents. While Jack and colleagues associate the former with Dennett's physical stance, they associate the latter with the phenomenal stance. In line with the speculation of Theriault and Young, the intentional stance is then thought of as involving both cognitive modes. As discussed in the chapter, this model receives support from recent evidence from cognitive neuroscience, as well as from five new experiments reported by Jack and colleagues.

Of course, philosophy of mind is a broad area touching on a wide array of philosophical topics. Not surprisingly, the opportunities for bringing empirical methods to bear on issues in philosophy of mind are correspondingly large; they are certainly not exhausted by issues related to mental state ascriptions and moral judgments alone. This volume closes with a pair of examples—two fascinating contributions to the literature that focus on areas beyond the core issues in experimental philosophy of mind discussed above.

In Chapter 7, Hannah Tierney, Chris Howard, Victor Kumar, Trevor Kvaran, and Shaun Nichols explore the intuitions of lay people with regard to personal identity. They note that while it often makes sense to assume conceptual monism in our philosophical theorizing, there is no guarantee that this assumption is correct in any given case. This is a point that is relevant to

many areas of experimental philosophy of mind, including work on mental state ascriptions and moral judgments. The risk of assuming conceptual monism is amply illustrated by Tierney and colleagues for the case of personal identity. They present a range of evidence indicating that lay judgments about the persistence of persons follow (at least) two different criteria—one concept of personal identity conforming to a psychological criterion, while another conforms to a biological criterion. From this, Tierney and colleagues conclude that pluralism best explains lay intuitions about personal identity; they don't stop there, however, but go on to argue that pluralism about personal identity is also a viable philosophical position.

In Chapter 8, Edouard Machery extends on his previous work (Machery 2005, 2009) arguing for a type of pluralism about concepts: According to the heterogeneity hypothesis, concepts don't form a unified kind, but instead split into three types that have little in common—prototypes, exemplars, and causal theories. If the heterogeneity hypothesis is correct, and if certain words lexicalize more than one of these distinct kinds of coreferential concepts (the polysemy hypothesis), then we would expect to find cases in which competent speakers are willing to endorse seemingly contradictory sentences because they read those sentences in terms of different coreferential concepts. Expanding on the empirical evidence given by Machery and Seppälä (2010), in this chapter Machery provides new evidence that this is in fact the case: He presents the results of a series of studies indicating that a substantial proportion of English speakers are willing to endorse a surprising range of seemingly contradictory sentences.

## References

Alexander, J. (2012), Experimental Philosophy: An Introduction. Malden, MA: Polity Press.

Arico, A., Fiala, B., Goldberg, R. F., and Nichols, S. (2011), 'The folk psychology of consciousness'. Mind & Language, 26, (3), 327–352.

Block, N. (1978), 'Troubles with functionalism', in W. Savage (ed), Perception and Cognition: Issues in the Foundations of Psychology (Minnesota Studies in the Philosophy of Science, vol. IX). Minneapolis: University of Minnesota Press, pp. 261–326.

Buckwalter, W. and Phelan, M. (2012), 'Function and feeling machines: a defense of the philosophical conception of subjective experience'. Philosophical Studies, doi: 10.1007/s11098-012-0039-9

Cohen, J. and Nichols, S. (2010), 'Colours, colour relationalism and the deliverances of introspection'. Analysis, 70, (2), 218–228.

Dennett, D. (1987), The Intentional Stance. Cambridge: MIT Press.

Fiala, B., Arico, A., and Nichols, S. (2011), 'On the psychological origins of dualism: dual-process cognition and the explanatory gap', in E. Slingerland and M. Collard (eds), Creating Consilience: Integrating the Sciences and the Humanities. Oxford: Oxford University Press, pp. 88–110.

Fiala, B., Arico, A., and Nichols, S. (forthcoming), 'I, Robot', in E. Machery and E. O'Neill (eds), Current Controversies in Experimental Philosophy. New York: Routledge.

Gray, H. M., Gray, K., and Wegner, D. M. (2007), 'Dimensions of mind perception'. Science, 315, 619.

Greene, J. D. (2007), 'The secret joke of Kant's soul', in W. Sinnott-Armstrong (ed), Moral Psychology, Vol. 3: The Neuroscience of Morality. Cambridge, MA: MIT Press, pp. 35–79.

Huebner, B. (2010), 'Commonsense concepts of phenomenal consciousness: does anyone care about functional zombies?' Phenomenology and the Cognitive Sciences, 9, 133–155.

Hurlburt, R. and Schwitzgebel, E. (2007), Describing Inner Experience? Proponent Meets Skeptic. Cambridge: MIT Press.

Knobe, J. (2008), 'Can a robot, an insect or God be aware?' Scientific American Mind, 19, 68–71.

Knobe, J. and Prinz, J. (2008), 'Intuitions about consciousness: experimental studies'. Phenomenology and the Cognitive Sciences, 7, 67–83.

Machery, E. (2005), 'Concepts are not a natural kind'. Philosophy of Science, 72, 444–467.

- Machery, E. (2009), Doing without Concepts. New York: Oxford University Press.
- Machery, E. and Seppälä, S. (2010), ‘Against hybrid theories of concepts’. Anthropology & Philosophy, 10, 97–125.
- Machery, E. and Sytsma, J. (2011), ‘Robot Pains and Corporate Feelings’. The Philosophers’ Magazine, 1st Quarter, 78–82.
- Nagel, T. (1974), ‘What is it like to be a bat?’ The Philosophical Review, 83, 435–450.
- Peressini, A. (2013), ‘Blurring two conceptions of subjective experience: folk versus philosophical phenomenology’. Philosophical Psychology, doi:10.1080/09515089.2013.793150
- Phelan, M. and Buckwalter, W. (forthcoming), ‘Analytic functionalism and mental state attribution’. Philosophical Topics.
- Phelan, M. and Waytz, A. (2012), ‘The Moral Cognition/Consciousness Connection’. Review of Philosophy and Psychology, 3, (3), 293–301.
- Reuter, K. (2011), ‘Distinguishing the appearance from the reality of pain’. Journal of Consciousness Studies, 18, (9-10), 94–109.
- Robbins, P. and Jack, A. I. (2006), ‘The phenomenal stance’. Philosophical Studies, 127, 59–85.
- Schwitzgebel, E. (2011), Perplexities of Consciousness. Cambridge: MIT Press.
- Sytsma, J. (2010a), ‘Folk psychology and phenomenal consciousness’. Philosophy Compass, 5/8, 700–711.
- Sytsma, J. (2010b), ‘Dennett’s theory of the folk theory of consciousness’. Journal of Consciousness Studies, 17, (3–4), 107–130.
- Sytsma, J. (forthcoming-a), ‘Revisiting the valence account’. Philosophical Topics.
- Sytsma, J. (forthcoming-b), ‘The Robots of the Dawn of Experimental Philosophy of Mind’, in E. Machery and E. O’Neill (eds), Current Controversies in Experimental Philosophy. New York: Routledge.
- Sytsma, J. and Livengood, J. (forthcoming), The New Experimental Philosophy: An Introduction and Guide. Peterborough, ON: Broadview Press.
- Sytsma, J. and Machery, E. (2009), ‘How to study folk intuitions about phenomenal consciousness’. Philosophical Psychology, 22, (1), 21–35.

Sytsma, J. and Machery, E. (2010), 'Two conceptions of subjective experience'. Philosophical Studies, 151, (2), 299–327.

Sytsma, J. and Machery, E. (2012a), 'On the relevance of folk intuitions: a reply to Talbot', Consciousness and Cognition, 21, 654–660.

Sytsma, J. and Machery, E. (2012b), 'The two sources of moral standing'. Review of Philosophy and Psychology, 3, (3), 303–324.

Talbot, B. (2012), 'The irrelevance of folk intuitions to the 'hard problem' of consciousness', Consciousness and Cognition, 21, (2), 644–650.

---

<sup>1</sup> See Sytsma and Livengood (forthcoming) for an introduction to experimental philosophy on the broad conception; see Alexander (2012) for an excellent introduction adopting the narrow conception.

<sup>2</sup> For a survey of work on these themes see Sytsma (2010a); see Knobe (2008) and Machery and Sytsma (2011) for short introductions. For further discussion of the connection between mental state attributions and moral judgments, see Phelan and Waytz (2012). Of course, there are exceptions to this characterization of experimental philosophy of mind, including important empirical work by experimental philosophers on other issues in philosophy of mind (see, for example, Hurlburt and Schwitzgebel, 2007; Schwitzgebel, 2011; Cohen and Nichols, 2010). In fact, as noted below, this volume closes with two new examples of such work.

<sup>3</sup> For the informational/phenomenal objection see Huebner (2010, 137) and Fiala, Arico, and Nichols (forthcoming); for other objections, see Talbot (2012) and the response in Sytsma and Machery (2012a), as well as Peressini (2013), Buckwalter and Phelan (2012), and Sytsma (forthcoming-a).