

## Comments on Brian Talbot,

### “The irrelevance of folk intuitions to the ‘hard problem’ of consciousness”

Bryce Huebner, Department of Philosophy, Georgetown University\*

---

Does the methodology advanced by experimental philosophers entail a revolutionary, radical, or at least revisionary re-conceptualization of how philosophy ought to be done? It's hard to say. But, when philosophers adopt experimental methodologies to deny “the most fundamental and manifest aspect of our mental lives”—i.e., qualia—the affirmative answer seems far more plausible! Sytsma and Machery (2009) attempt to deflate the call to 'take consciousness seriously' by showing that commonsense judgments suggest that the 'hard problem' is a philosophical invention. Similarly, I've argued that philosophical intuitions about phenomenal consciousness uncover little more than mere ideology—that's my attempt at philosophical radicalism (Huebner, in press). But many people seem to think that the assumption that commonsense intuitions reveal interesting facts about the furniture of the world *radically tempers* the philosophical import of experimental philosophy. Perhaps a more careful look at the causal mechanisms behind the production of these intuitions can explain why such quasi-democratic appeals to intuitions—for all their revolutionary fervor—do little more than obscure the real philosophical questions. As we all know, democratic appeals to 'folk' intuitions are not always philosophically *apropos*.<sup>1</sup>

Without expressing hostility toward experimental methods, Brian Talbot has offered a novel criticism of experimental philosophy. He wants to improve experimental methodologies by embedding them in a range of highly confirmed psychological generalizations; and he offers interesting suggestions about how commonsense judgments about phenomenal consciousness could be employed to deflate the 'hard problem of consciousness'. While I can't be sure, I gather that he's not optimistic that this will ever happen. Strangely, I agree. So long as philosophers are convinced that there is a 'hard problem', no amount of commonsense data is likely to allow us to 'solve' it, or even to dissolve it.

Yet, while I agree with Talbot on a number of methodological issues, I am not convinced that the data collected by experimental philosophers should be treated as System-1

---

\* Acknowledgments: I have recently learned quite a bit about the role of experimental methods in the study of the mind through conversations with people inside and outside of the emerging experimental movement in philosophy. I would like to offer a special thanks to Robert Briscoe, Dan Dennett, Aaron Garrett, Mark Lance, Eric Mandelbaum, Justin E. H. Smith, Susanne Sreedhar, Kate Withy, and Tad Zawidski in this regard. I'm almost positive that none of them will agree with the use to which I have put their insights and ideas—but such is the hazard of providing a philosopher with tools for constructing and deconstructing a view of how the world 'hangs together'.

<sup>1</sup> As an aside, let me note that experimental philosophy can only be slightly more democratic than its armchair cousin given its reliance on the intuitions of people who are largely Western, Educated, Industrialized, Rich, and Democratic—that is, people who are globally WEIRD (cf., Heinrich 2009).

judgments. In these comments, I explain why not and I offer some suggestions about the sorts of mechanisms that I believe are likely to be responsible for the production of commonsense and philosophical intuitions about phenomenal consciousness.

### **The argument:**

Let me begin by noting that I agree with Talbot that a plausible interpretation of the experimental data on phenomenal consciousness requires a clear understanding of the mechanisms that are responsible for the production of human judgments more broadly. I doubt that any experimental philosopher (or anyone interested in the nature of intuition, for that matter) would disagree. But, it's an important reminder! We ignore research on judgments, decision-making, counterfactual reasoning, meta-representation, dual-process reasoning, and imagination only at our own peril; and it is with this plausible reminder in hand, that Talbot proceeds to argue roughly as follows.

Suppose that judgments about mental states are reflexive and associative System-1 judgments. If judgments about whether a simple robot can be in various mental states are System-1 judgments, then they will depend on associations between the physical features of the robot and the physical features of entities that have previously been judged to be in that sort of state.<sup>2</sup> Of course, past experience provides numerous associative links between physical features that can be implemented in a simple robot and non-qualitative mental states. So, the System-1 judgment that even a simple robot can be in non-qualitative mental states should be readily produced. However, things are different in the case of qualia—if there are qualia, we only have first personal experiences of them. This being the case, associative links between an entity's physical features and its qualia are likely to be both relatively rare and relatively weak. So, a System-1 judgment that a simple robot can be in qualitative mental states should be relatively rare as well. However, the fact that we reflexively simulate affective states means that past experience will provide associative links between certain physical features and these affectively valenced states. Here, however, simple robots lack the physical features associated with affectively valenced states; so people should be less likely to ascribe affectively valenced states to simple robots. This being the case, "if Sytsma and Machery inadvertently studied System One judgments, their data sheds no light on the hard problem of consciousness, because System One judgments of others' mental states would not [or would shallowly and unreliably] consider qualia even if they exist."

The argument is elegant. But as Talbot notes, it only goes through if judgments about mental states are (largely) System-1 judgments. Unfortunately, I find the argument for this point to be too quick. Talbot claims that System-1 judgments are likely to dominate a participant's judgments unless the participant has some reason to mistrust the System-1 judgment, some motivation to get the right answer, and is aware of a better decision strategy. Talbot suggests that these criteria are not met in experimental situations like those used by Sytsma and Machery, so he assumes that the relevant judgments must derive from a System-1 process. But is he really warranted in making this assumption?

In answering this question, it is important to remember just how strange these sorts of

---

<sup>2</sup> Throughout this argument I use 'physical features' as a stand-in for 'physical features or behavior capacities'.

experiments really are. While most of the participants in these studies are unlikely to have thought much about the mental lives of simple robots, as ‘natural psychologists’ they are likely to possess a heterogeneous array of tools, heuristics, and background assumptions—many of which are likely to be poorly articulated—that allow them to make sense of the mental lives of unfamiliar entities. Of course, we have some initial intuitions about the mental capacities of various entities—and these intuitions are likely to be driven by System-1 mechanisms of precisely the sort to which Talbot appeals. However, it seems unreasonable to assume that *most* people will just rest satisfied with these intuitions. In ecologically valid contexts, judgments about subjective experience tend to be deeply embedded in social environments rich with corrective feedback from target entities. However, without this feedback, I contend that participants are likely to rely on System-2 strategies to evaluate the plausibility of their initial judgments.

I find it difficult to believe that participants in these tasks—participants who might take 5 seconds or more to respond to an experimental prompt—do not evaluate the plausibility of their initial intuitions by constructing inferential models and examining the implications of their intuition in inner speech (a paradigmatic System-2 process). Having listened to my share of participants who preferred to do this deliberation out loud, my guess is that it tends to go something like this: “I guess it seems like Jimmy wouldn’t smell bananas. Oh, I don’t know. Damn, this question is hard! Ah come on, nobody would build a robot that smells bananas! Whatever. I’ll go with a 2...no a 3”. My guess is that participants engage in a great deal of revision on the basis of the initial System-1 judgment that they have made about the presence of mental states.

### **Both System-1 and System-2:**

Let me go a bit further down this rabbit hole. But first, let me note that I agree with Talbot that the judgments that are offered in experimental contexts often depend on associative mechanisms—they just don’t stop there. To ask participants whether a simple robot feels pain, or smells bananas, is to ask (if only implicitly) whether the robot has a mental life that is like ours in some respect. Without this frame of reference, the question doesn’t even make sense (cf., Dennett 1997). Moreover, facts about a person’s “experience, both present and past” are likely to play a prominent role in producing their judgments about the mental lives of non-human entities. As Sherry Turkle’s research has clearly shown, interactions with simple robots and other technologies can have a significant impact on people’s judgments about the mental lives of these entities. But I wish to draw a more radical conclusion from these facts: philosophical thought experiments, as well as the probes that are used by experimental philosophers, are nothing more than tools for imagination management.

When an experimental philosopher asks whether a simple robot can be in perceptual or emotional states, she is asking a question that is incredibly difficult to answer. I worry that it is not even clear *what* question she is asking. Such questions are likely to lead participants in experimental studies to engage in a series of inferences (both personal and subpersonal) that are intimately bound up with their theoretical commitments and presuppositions about what counts as evidence for the claim that the robot is in that state. So, I worry that Talbot’s appeal to the distinction between System-1 and System-2 judgments is too simplistic to capture the rich complexity of human judgments. Whether

we like it or not, experimental probes lead participants to reflexively construct simplified and idealized mental models that provide a set of initial assumptions from which participants can engage in deliberative reasoning in attempting to interpret previously unexamined cases and possibilities. Unfortunately these simplified and idealized mental models are likely to provide only a partial and incomplete representation of the case at hand, thereby focusing attention on a narrow range of seemingly salient features, which in turn structure the space of possibilities that they consider in evaluating the experimental prompt (Fauconnier & Turner, 1998, 2003). Experimental prompts, like philosophical thought experiments, succeed where they evoke responses that rely on elicited assumptions that lead people “(either reflectively or unreflectively) to represent relevant non-thought experimental content in light of the thought experimental conclusion” (Gendler, 2007, p. 69).

While we always start from default assumptions about the way that our world happens to be, we can revise our impressions in various ways that extend to the realm of intuitive plausibility. Good writers have always used clever writing techniques to facilitate the construction of ‘blended mental spaces’ (Fauconnier & Turner, 2003), which allow us to make sense of the events that can occur in merely possible worlds. This is how many of us learned that, in distant galaxies, robots can be in mental states just like our own. By learning to place ourselves in the context of a science fiction world, we learn to effortlessly interpret the actions of various non-human entities in terms of their reasons, beliefs, intentions, goals, and feeling—at least once we have gotten a handle on the rules that are operative within their universe. By constructing localized understandings of fictional worlds, and borrowing liberally from the structure of our everyday experience, we are able to revise our judgments about what is possible on the basis of the information provided by the structure of a fictional world. Of course, there are deep questions about how we construct mental spaces that allow us to appeal to alternative possible worlds. However, by the age of five, children show a robust capacity to distinguish between the range of possibilities that different mental spaces allow, including the features that are shared across various spaces as well as the features that differ between spaces (Skolnick & Bloom, 2006a, 2006b). I suggest that thinking about fictional worlds provides an important heuristic strategy that we can adopt in constructing mental spaces as we attempt to make sense of the world of our everyday experience (Fauconnier & Turner, 1998).

Consider the way in which the two ‘smell’ cases discussed by Sytsma and Machery provide tools for narrowing the focus of a participant’s attention, highlighting certain facts about the relevant system, and making certain possibilities seem more plausible. Suppose that a participant is asked whether a simple robot can smell Isoamyl Acetate. She might initially judge that robots don't smell anything; but as she reflectively evaluates the plausibility of her judgments, considering the reasons why someone would build a robot that could smell Isoamyl Acetate, she might revise this initial judgment. After all, it's easy to imagine a robot that was built to detect a strange chemical substance (even one, unbeknownst to her, that smells a lot like bananas). However, had she been asked, instead, whether a simple robot could smell bananas, she might reflectively evaluate the plausibility of her judgment by considering the reasons why someone would build a robot that could smell bananas and in this case she might come up blank. While it is *conceivable* that someone would build a banana-sniffing robot, it's much harder to

imagine any plausible reason why someone would do so. The participants in these experiments were likely to have constructed (perhaps subpersonally) different spaces of relevant possibilities by conjuring simplified images of the robot and its maker. This would then have allowed them to check the range of possibilities against their reflexive commonsense impressions of what it takes to be a minded entity. But, where they reflectively found that the analogy between an exotic system and a run-of-the-mill individual did not look plausible, they rejected the possibility of mental states in the simple robot.

#### **4. Concluding remarks**

I want to close with a quick plea to all experimental philosophers and their critics: abandon the use of the term 'the folk' in discussing the participants who take part in experimental studies. I readily acknowledge that the term 'folk psychology' is deeply embedded in our disciplinary discourse, and so there are compelling reasons for it to stay; but it is a deep mistake to appeal to 'the folk' as though there were some unitary tribe whose intuitions we are excavating when we carry out various experiments. The supposition that 'the folk' constitute a respectable entity threatens to lead us astray in our philosophical attempts to conceptualize the generalities encountered in excavating commonsense concepts. If the account of judgment that I have briefly offered is anywhere close to correct, references to 'the folk' are likely to both obscure differences in the structure of mental spaces that are constructed in responding to the underdeveloped intuition pumps offered by experimental philosophers, and to obscure similarities in mechanisms that are responsible for the production of philosophical and commonsense intuitions. No doubt, philosophers come in distinct flavors—and the mental spaces that we have learned to construct have robust structural similarities grounded in the possibilities that are opened up (as well as closed off) by our academic training. But, while thought experiments and the philosophical intuitions that they evoke can lead us to better ways of theorizing, left unchecked they are equally likely to solidify philosophical prejudices, to reaffirm theoretical commitments foisted upon us in graduate school, and to further entrench the dominant trends in our discipline. As Dan Dennett often puts the point, becoming a philosopher requires that we learn strategies of intuition management; but we must remember that merely adding additional structure to our mental spaces doesn't introduce a categorical difference in kind between *us* and *the folk*. This, I contend, is the truly radical import of experimental philosophy!