

The Irrelevance of Folk Intuitions to the “Hard Problem” of Consciousness

Word Count: 3,900 (text without references or notes)

Abstract: Recently, a number of philosophers have turned to folk intuitions about mental states for data about whether or not humans have qualia or phenomenal consciousness. In this paper I argue that this is inappropriate. Folk judgments studied by these researchers are mostly likely generated by a certain cognitive system – System One – that will ignore qualia when making these judgments, even if qualia exist. If experimental research has any hope of shedding light into the existence of qualia, it needs to be better founded in an understanding of how we make judgments.

Certain mental states, such as perceptual ones (seeing red), emotions (feeling anger), or bodily sensations (pain), are supposed by many philosophers to have a qualitative aspect – a “what it is like” to be in those mental states. The experience of this qualitative aspect to these mental states – of *qualia* – is phenomenal consciousness. The existence of phenomenal consciousness gives rise to what David Chalmers (1996) calls “the hard problem” of consciousness: explaining how it is that we have phenomenal consciousness.

Recently there have been a number of papers that look to folk intuitions (intuitions of non-philosophers) for data to help understand phenomenal consciousness and possibly resolve the hard problem (Knobe & Prinz, 2008, Sytsma & Machery, forthcoming, Huebner, forthcoming). In this paper I will argue that folk intuitions as currently studied shed no light on the hard problem. I am not, however, making an argument against the use of folk intuitions generally in philosophy, nor against the use of folk intuitions as data about consciousness if properly gathered and used. My discussion will focus on one particular series of experiments on folk intuitions, carried out by Justin Sytsma and Edouard Machery (forthcoming); while their research is among the strongest experimental philosophy work so far on this topic, they still make a mistake common to almost all experimental philosophy work on consciousness, a mistake that is also symptomatic of a more general problem in the philosophical use of intuitions.

I will start by first giving Sytsma and Machery’s argument, and describing the data they generated to support this argument. I will then explain why research of the sort they are doing is unlikely to really be able to support the claims they make; I will argue that the cognitive system that likely generates the judgments studied by Sytsma and Machery is not equipped to give useful data on phenomenal consciousness. At the end, I will generalize from this and discuss ways one could do research so as to avoid this problem.

Sytsma and Machery’s Argument

Sytsma and Machery make the following argument: if we have phenomenal consciousness, then it is “the most central and manifest aspect of our mental lives.”¹ This

¹ Sytsma & Machery (forthcoming, 2), citing Chalmers, 1995, 207. All page references to Sytsma & Machery are to the pre-print version of their paper.

claim is an adoption of one made by those who think that there *is* a hard problem. If qualia or phenomenal consciousness are the most central and manifest aspect of our mental lives, then we should expect the folk to be aware of them. If the folk are aware of qualia, then this awareness should be reflected in folk judgments about mental states. Specifically, it ought to play a role in how the folk ascribe mental states (or their lack) to various things in the world. However, Sytsma and Machery claim, it seems that it does not. So, by *modus tollens*, they conclude that there is no hard problem, or at least we have some good evidence that there is none.

To support their claim that an awareness of qualia, or phenomenal consciousness, plays no role in folk judgments about mental states, Sytsma and Machery performed a series of experiments on the folk. Through these experiments, they learned that folk subjects are willing to ascribe certain types of supposedly qualitative mental states to beings that are not conscious – to very simple robots. However, they were not willing to ascribe other qualitative mental states to these beings. The folk reported that a robot could see red and smell certain smells (that of Isoamyl Acetate), but could not feel pain or smell other smells, such as that of a banana.²

This data seems to indicate that folk ascriptions of mental states are not based on those mental states being qualitative or non-qualitative, since some qualitative mental states but not others were attributed to robots. This also seems to indicate that the presence or absence of phenomenal consciousness does not play a role in these ascriptions. Further, those sympathetic to qualia can give no obvious, straightforward explanation of this data. It cannot be attributed to robots' lack of the appropriate perceptual faculties, since they were judged able to smell *some* smells but not others. It cannot be attributed to the folk taking robots to be conscious, since robots were judged *unable* to smell other things, and if robots were conscious and able to smell, there is no apparent reason for this inability. It seems unlikely that folk interpret the words used like “see” or “smell” differently than do philosophers (e.g. using them to refer to purely functional states), since this does not explain why they reported that robots could smell something but not others, or could see but not smell.

Sytsma and Machery's claim is that these judgments track the distinction between valenced and non-valenced mental states, and not the qualitative/non-qualitative or conscious/non-conscious distinction. Valenced mental states are those that are associated with goodness or badness (for the agent in question); often they are pleasurable or painful mental states. Non-valenced mental states are not associated with goodness or badness. Seeing red is non-valenced, but smelling a banana or feeling pain are valenced; the mental states not attributed to robots were valenced ones. Sytsma and Machery argue that the valenced/non-valenced distinction is independent of the qualitative/non-qualitative distinction, and that valence does not generate a hard problem

² Subjects in each experiment were presented with some variation of the following stimulus: “Jimmy (shown below) is a relatively simple robot built at a state university. He has a video camera for eyes, wheels for moving about, and two grasping arms with touch sensors that he can move objects with. As part of a psychological experiment, *he was put in a room that was empty except for one blue box, one red box, and one green box (the boxes were identical in all respects except color)*. An instruction was then transmitted to Jimmy. It read: ‘*Put the red box in front of the door.*’ Jimmy did this with no noticeable difficulty. *Did Jimmy see red?*” (Sytsma & Machery, forthcoming, emphasis mine). The italicized portions were changed to reflect the different qualitative states being tested (smelling banana, etc.) The pictured robot was very simple and bore no resemblance to a human.

of consciousness. Since whether a mental state is qualitative or not did not seem to make a difference in when or to whom subjects ascribe that mental state, it seems that phenomenal consciousness plays no role in our ascriptions of mental states. Thus, it seems that the folk are not aware of qualia or phenomenal consciousness, and it seems that the hard problem is really the invention of philosophers.

In what follows, I will not question Sytsma and Machery's data or that these folk judgments are based on the distinction between valenced and non-valenced experiences, rather than that between qualitative and non-qualitative experiences. I think their experimental work is quite good³ (and I think there is a more fundamental problem with their research). Instead, I will question the conclusion they draw from it. I will argue that, even if folk judgments of others' mental states do track the valenced/non-valenced distinction, and this distinction has nothing to do with qualia or phenomenal consciousness, we are not justified in concluding that the folk are not aware of phenomenal consciousness or qualia. This is because the research they have done can shed no light into folk subjects' (non-)experience of qualia. Even if the folk *do* experience phenomenal consciousness, this experience should play no role in the sorts of judgments Sytsma and Machery (and other experimental philosophers working in this area) study.

System One Judgments about Mental States

We make judgments in a number of ways. Consider the judgment "Cats are animals." One might judge this to be true because it simply strikes them as true. One might also judge this to be true because they recall having been told that it is true by a trusted source. One also might judge it is true because one has consciously recalled a number of traits cats share, and has compared these traits to those one believes animals must possess. Each of these ways of making a judgment draws upon, to some extent, distinct cognitive systems. The cognitive system that generates immediate, snap judgments is (to a large extent) distinct from that which generates our consciously-reasoned-to judgments.

Different cognitive systems draw upon different sources of information to make the judgment they make. If, for example, I have just been taught the rules to a complex game, I may be able to consciously draw upon these rules to make judgments about what moves in the game are legal, but it is unlikely that my snap-judgment generating system will be able to use these rules. However, if I have played the game for years, I will at some point have automatic judgments about what moves I can and cannot make. The snap-judgment (or intuition) generating system can, on the other hand, draw upon information about quite subtle patterns that I have been exposed to that I am not consciously aware of (Lewicki, et al, 1992) – for example, we all make automatic judgments about the grammaticality of sentences in our native languages, even when we do not consciously know the rules that make these sentences (un)grammatical.

This discussion, while no doubt fascinating, is also central to the point I want to make about Sytsma and Machery's research (and about experimental philosophy of mind more generally). I want to argue that a certain cognitive system is generating the judgments Sytsma and Machery are studying, and that that cognitive system is limited in the information it can or will use. Specifically, we should not expect it to use information

³ This has been confirmed by discussion with a colleague in the psychology department (Schwartz, personal communication).

about the qualitiveness or non-qualitiveness of others' mental states, even if that information is (in some sense) available. Thus, the fact that the judgments of this system are not based on the qualitative/non-qualitative distinction tells us nothing about whether or not the folk experience qualia.

While there are a number of cognitive systems, I am going to focus on two here to make my argument. One is slow, conscious, and largely oriented towards rational, rule-based thought. This is the system that would generate the judgment "Cats are animals" by consciously recalling traits of cats and traits of animals, and explicitly comparing the two. This is the cognitive system most of us are most familiar with, since it's the one we are most directly aware of; let's call it "System Two," borrowing the name from the psychologists Keith Stanovich and Richard West (2000). The other cognitive system I want to discuss is faster than System Two, does most of its work unconsciously, and makes judgments based largely on associations. This latter system is often called "System One" by psychologists, and is typically thought of as the intuitive system. I will argue in a minute that System One is mostly likely responsible for the folk judgments Sytsma and Machery study. But first, I want to argue that System One (that is, intuitive) judgments about others' mental states are not likely to employ information about qualia.⁴

System One judgments are associative. When System One judges some object O to have trait T, it is because O bears a strong resemblance to other objects which have trait T and/or because O is found in a context that we have encountered other T things in in the past. The stronger the resemblance between O and other T possessing things, or the more often T possessing things have been found in O's context, the more likely System One is to judge O to have T. So, if I see some object that looks and acts just like a dog, then System One is going to tell me that it is a dog; if I find something that looks mostly like a dog, and I'm in a dog park, then System One is likely to judge that it is a dog (there are a number of *caveats* to this account that are not important here).

Experience, both present and past, plays a huge role in System One's judgments. Most of the information we have about things we have encountered in the past comes from experience of those things; our senses almost never stop working and giving us information, and so, generally speaking, most of the information we have about any object comes from experience of that object. Thus, *experienced* similarity between objects will generally be the basis for System One judgments.

Let's apply this to System One judgments about mental states. When System One ascribes a mental state to some object, it looks for similarities between this object and past objects that had that mental state. These are mostly going to be similarities between what we *experience* of the object and what we have experienced of objects in the past that had that mental state. In other words, to decide if some thing is seeing red, System One compares experienced facts about this thing to experienced facts about past objects that have been judged to see red.

This means that we should not expect information about qualia (or their absence) to play much a role in System One judgments about others' mental states, even if qualia exist. This is because when we encounter things with mental states we necessarily never experience their qualia, whether or not qualia exist. From time to time we might think about others' qualia, but the number of times we do this versus of the number of times we only see or hear or smell or touch (or sometimes taste) these people and things is

⁴ For more on Systems One and Two, see Sloman, 1996, or Kahneman & Frederick, 2002.

vanishingly small. Thus, we will have much stronger associations between the external features of objects and their mental states than we do between data about qualia and mental states. These external features will be largely behavioral; if something behaves as if it sees red (as the robots in Sytsma and Machery's experiments do) then System One should judge that it sees red.

Now, we could form strong associations between qualia (if they exist) and mental states through first-person experiences of our own qualia and mental states. However, this is likely to play relatively little role in System One judgments about *others'* mental states. The first- and third-person perspectives are quite different, and there is very little similarity between the sorts of stimuli that prompts one type of judgment (first person, non-perceptual experience of our own mental life) and the other (third person, mostly perceptual, experience of others). Given this lack of similarity, we should not expect an associative process to connect third person stimuli to first person experience; thus, System One judgments of others' mental states will draw largely on past third-person experience.

Valence and System One

Sytsma and Machery found that the folk did not ascribe some mental states to robots. How do the judgment that robots don't feel pain, or can't smell bananas, even though they behave as if they do, fit into the two systems account I have been giving?

When we encounter others that are experiencing valenced mental states – pleasure or pain, or anger or joy, etc. – we generally experience more than their behavior or appearance. We have a set of mental systems that often cause us to simulate the experience of valenced or emotionally charged mental states when we see another who is experiencing that mental state. Here's one way that occurs: when we experience positive or negative mental states, we typically have characteristic facial expressions, whereas we do not for non-valenced (neutral) mental states. When we see another person's facial expression, we often simulate making that expression in our own minds.⁵ The relationship between mental states and facial expressions is a two-way street: having these facial expressions is not only caused by the associated mental state, it also causes it (to a much lesser degree): smiling actually makes you a little happier. This happens even when we *simulate* these expressions internally. Thus, we should expect people to associate others' non-valenced mental states largely with behavior, but others' valenced mental states will be associated with behavior, facial expression, and with their *own* emotional experience.⁶

⁵ For an overview of the literature on the material discussed in this paragraph, as well as a more detailed analysis of the evidence for and mechanisms behind these processes, see Goldman & Sripada, 2005. For the sake of simplicity I have left out some nuances of the different accounts of just how this works; for example, some theorists think that we simulate emotions *first* and then the appropriate facial expressions. The effect is the same for my purposes.

⁶ This does not require that every valenced mental state have a corresponding facial expression, or be accompanied by some emotion-signalling facial expression. Associations exhibit some degree of transitivity (Lewicki, et al, 1994). So, if we have strong associations between many valenced mental states and characteristic facial expressions, and between these mental states and other mental states without characteristic facial expressions, then how we attribute one mental state will affect how we attribute the other.

This means that when System One judges whether or not another person has a certain mental state, it looks for different things when making judgment about valenced and non-valenced mental states. Further, some of the things it looks for when making judgments about robots' valenced states are not there. Robots (at least, the very simple ones described in Sytsma and Machery's experiments) lack facial expressions, and should also not cause us to simulate emotions in our own heads.

If the judgments Sytsma and Machery studied are System One judgments, we would expect just the results they found. Non-valenced mental states should be ascribed based largely on behavior, which means that they can be ascribed to appropriately behaving robots. Valenced mental states should only be ascribed to beings that are more like humans, and which have facial expressions.⁷ Notice that this has nothing to do with qualia. The System One generated judgment that a robot cannot feel pain but that a person can is not based on any consideration of the other's qualia or phenomenal consciousness at all, but rather on whether or not it is similar to other things that we have made these judgments about in the past. We would expect exactly the same judgments whether or not we have qualia of our own.

Why think System One is responsible for Sytsma and Machery's data?

So, 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 consider qualia even if they exist. This is only a problem for their argument if we have reason to think that the judgments they elicited were (largely) generated by System One. Why think this?

System One judgments are automatic and effortless (Kahneman & Fredrick, 2002). This means that they occur whether we want them to or not, even when we are thinking about something else: "The associative system... always has its opinion heard..." (Sloman, 1996, 15) This does not mean that we always use or report our System One judgments. We can suppress them or replace them with judgments made in other ways. But this only happens when we are a) aware of reasons to mistrust a given System One judgment, b) we care about getting the right answer in that case, and c) when we are aware of what we think is a better of making the judgment in question. Otherwise, people tend to simply make and report System One judgments (see, e.g. Kahneman & Fredrick, 2002, Sloman, 1996). So, if we have no reason to think that the subjects in Sytsma and Machery's study met conditions a, b, and c, we should take their judgments to be System One judgments.

The folk subjects Sytsma and Machery studied likely did not meet these three conditions. It was a low stakes situation, a psychological experiment about a seemingly easy question where almost no one will know the answer the subjects gave and the answer will affect no one (as far as the subjects know). There is no obvious reason to mistrust one's gut judgment on the matter, and no clearly better way to approach the question. Now, each of these is somewhat variable; some people are more inclined to

⁷ For reasons of personal history, idiosyncratic interpretation of the stimulus, or idiosyncratic tendencies to engage System Two during studies, we should expect some deviation from this. Further, if we throw odder situations at subjects – e.g. human bodies with computer brains, or robot bodies with human brains (see Huebner, forthcoming), we should expect a greater percentage of subjects to engage System Two (conscious reasoning) or to get ambivalent results from System One.

second guess their answers, or to take experiments more seriously. And that is reflected in the degree of variation in responses to some of the prompts. But we should expect subjects' responses to be heavily weighted to System One responses, and thus to be irrelevant to the issue studied.

Sytsma and Machery do bring up other evidence for their conclusion: they claim that it is hard to teach non-philosophers about phenomenal consciousness, and hard to make them aware of qualia, both of which (it is claimed) are surprising if qualia are really as manifest as their proponents claim. I find this claim about surprisingness dubious myself: it seems to me that the most fundamental aspects of our experience *should* be hard to talk about, and thus to teach non-philosophers about or to get them to talk about. It's hard to put the simplest things into simpler terms, so talking about qualia should be much harder than having them. But even if this were a good argument, the data is only anecdotal and briefly sketched; a strong argument requires more research.

Generalizing and How to Improve

The problems I have raised here should generalize, and will be issues for any experimental research into consciousness via folk judgments. System One judgments are not likely to be sensitive to the presence or absence of qualia even if they exist. One way to try to address this is to elicit judgments from other mental systems. I am skeptical about this option. In order to do this, we need to give subjects a reason to doubt their intuitive judgments, and to think they can do better otherwise. However, it does not seem to me that most people have thought much about how to categorize and ascribe mental states. This means that they are not likely to know of a better way to ascribe them other than just going with what intuitively seems right. Further, if we call their intuition into question, this is likely to call their judgment about the issue more generally into question, given that it is natural to make these judgments using System One. This is likely to get subjects to second guess whatever answers they give, and generate bad data.

Another way to approach this research would be to allow subjects to use System One, but to ask them to make judgments about their *own* mental states. If qualia exist, subjects do have extensive experience of their own qualia, and it would make sense for System One to be able to use the presence or absence of qualia as a grounds for ascribing mental states to one's self. The difficulty here is in replicating the structure of Sytsma and Machery's study with judgments about our own mental states. Sytsma and Machery compare ascriptions of mental states to beings who are taken to be conscious and to those that are taken to not be conscious. To replicate this, we would need to compare subjects' judgments about their own conscious and unconscious mental states. However, it is in a certain sense hard to imagine ourselves being unconscious, and it is likely that subjects make judgments about their unconscious selves just as they would make judgments about others. If this cannot be avoided, it would mean that System One judgments about our own mental states are just as likely to ignore qualia as System One judgments of others' mental states. Further, it will be impossible to replicate using this method other experiments in the literature on consciousness, such as those on group entities (Knobe & Prinz, 2008).

A way to avoid this might be to pursue research on System Two (conscious) judgments about qualia. We might look at ways to get the folk to articulate an awareness of their own qualia; if the folk cannot be brought to do so (as Sytsma and Machery

suggest) then we have good evidence that there are no qualia. The danger here is that we will, in a sense, talk people into believing in qualia, rather than get them to talk about something they already believe in. After all, a great many philosophers believe in qualia; if there really are none, then there is something about thinking about philosophy of mind that tends to cause one to believe in qualia. Putting the folk in mind of the distinctions they need to talk about what we are interested in will always run the risk of taking from them that which makes them useful as subjects.

Conclusion

I am not trying to say that studying the hard problem via folk judgments is impossible; however, there are clearly significant difficulties in this line of research. There are certain kinds of phenomena that System One is not likely to pick up on (and there are certain kinds that System One is much better at using than System Two). Experimental philosophy is very likely to elicit System One judgments, and there are significant dangers in trying to get System Two judgments from the folk. This means that experimental philosophers need to be extremely careful when studying folk judgments, both about mental phenomena and also in general, that the mental systems engaged in their studies are appropriate to the task at hand.

References

- Chalmers, D. (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies*. 2, 200-219.
- Chalmers, D. (1996) *Consciousness*. Oxford: Oxford University Press.
- Goldman, A.I. & Sripada, C.S. (2005) Simulationist models of face-based emotion recognition. *Cognition*. 94, 193-213.
- Huebner, B. (forthcoming) Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences*.
- Kahneman, D. & Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich, T., Griffin, D., Kahneman, D. (eds). *Heuristics and Biases*. New York: Cambridge University Press.
- Kauppinen, A. (2007) The rise and fall of experimental philosophy. *Philosophical Explorations*. 10, 119-122.
- Knobe, J. & Prinz, J. (2008) Intuitions about consciousness: experimental studies. *Phenomenology and the Cognitive Sciences*. 7, 67-85.
- Lewicki, P., Hill, T., Czyzewska, M. (1992) Nonconscious acquisition of information. *American Psychologist*. 47, 796-801.
- Sloman, S.A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin*. 119, 3-22.

Stanovich, K.E. & West, R.F. (2000) Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*. 23, 645-726.

Sytsma, J. & Machery, E. (forthcoming) Two conceptions of subjective experience. *Philosophical Studies*.