Philos Stud (2013) 166:349–361
DOI 10.1007/s11098-012-0039-9

# Function and feeling machines: a defense of the philosophical conception of subjective experience

**Wesley Buckwalter · Mark Phelan**

Published online: 26 October 2012


**Abstract** Philosophers of mind typically group experiential states together and distinguish these from intentional states on the basis of their purportedly obvious phenomenal character. Sytsma and Machery (Phil Stud 151(2): 299–327, 2010) challenge this dichotomy by presenting evidence that non-philosophers do not classify subjective experiences relative to a state's phenomenological character, but rather by its valence. However we argue that S&M's results do not speak to folk beliefs about the nature of experiential states, but rather to folk beliefs about the entity to which those experiential states are attributed. In two experiments, we demonstrate that ordinary attributions of subjective experiences (of smell and felt emotions) to a simple robot are not sensitive to valence, but instead respond to functional assumptions about the entity to which the states are (or are not) attributed.

**Keywords** Philosophy of mind · Phenomenal consciousness · Subjective experience · Experimental philosophy · Mental state attribution · Function

Philosophers typically distinguish mental states such as seeing red, feeling pain, and feeling angry, from mental states such as believing that such-and-such is the case or wanting some state of affairs to obtain. Philosophers of mind as diverse as Thomas

**Electronic supplementary material** The online version of this article (doi:10.1007/s11098-012-0039-9) contains supplementary material, which is available to authorized users.

W. Buckwalter (✉)
The City University of New York-Graduate Center, 365 Fifth Ave., New York, NY 10016, USA
e-mail: jbuckwalter@gc.cuny.edu
URL: http://wfs.gc.cuny.edu/JBuckwalter/index.html

M. Phelan
Lawrence University, Appleton, WI, USA

Nagel (1974), John Searle (1994), Ned Block (1995), and David Chalmers (1995), defend this dichotomy on the grounds that members of the first set of states each have a manifest and undeniable *phenomenal character* that the second set lacks. Accordingly, these philosophers maintain that we should group the first set of states—which we will call *experiential states* (or *subjective experiences*)—together because they all *just obviously* share the property that there is 'something it is like' to occupy them. Conversely, they maintain that we should distinguish these from the second set of states—what we will call *intentional states*—because those lack this feeling or quality.

However, some philosophers have recently claimed that if experiential states really are fundamentally different from intentional states because of their *manifest* and *undeniable* phenomenal character, then this fact may entail a series of important empirical commitments about the way ordinary people sort mental states (see Sytsma and Machery 2010; Knobe 2008; Dennett 1996). Presumably, we should expect that non-philosophers sort mental states in a similar way as philosophers. The prediction is that ordinary people will categorize states like seeing red, feeling pain, and feeling angry together despite their differences because it is like something to occupy those experiential states. And also, that they will distinguish those experiential states from intentional mental states of believing or desiring because it is not like anything to occupy those intentional states. But do non-philosophers group the diverse states that make up the philosophers' class of subjective experiences together, and do they distinguish these from intentional states?

Recently, cognitive scientists have begun to explore these questions by empirically investigating how we ordinarily conceive of subjective experience (for a review see Knobe et al. 2012). Like a number of philosophers of mind before them, these researchers investigate these questions by examining our intuitive attributions of mental states to different kinds of entities. But instead of speculating on such attributions by way of a thought experiment, these researchers actually assess the responses of ordinary people by means of controlled experiments. This method assumes that attributions of mental states reflect the categorization-schema for mental states that people tacitly accept. For example, if people think an entity can occupy the full range of mental states except for experiences, then by assessing which states they are unwilling to attribute to that entity, we should be able to determine which states people categorize as experiences. In this way, we can ascertain whether the states ordinary people categorize as subjective experiences are the same states philosophers have identified as possessing an unmistakable phenomenal character.

Building on the assumption that the intuitive schema for mental states is reflected in ordinary attributions, three mutually exclusive positions on the folk category of subjective experience are now advocated in the empirical literature. The first position holds that ordinary people's attributions of mental states to other entities reveal their tendency to group experiential states together and distinguish them from intentional states in roughly the same way that philosophers traditionally have (Robbins and Jack 2006; Gray et al. 2007; Knobe and Prinz 2008). Alternatively, a second position maintains that empirical studies give us no good reason to suppose that people distinguish between experiential and intentional states, and emphasizes that some of the same general properties that lead us to attribute intentional states

also lead us to attribute experiences (Arico et al. 2011; Phelan et al. 2012). Lastly, the third view is that people group experiential states together and distinguish them from intentional states, they just do not do this in the same way as philosophers (Sytsma and Machery 2010).

Here we challenge a prominent statement of the third position. According to Sytsma and Machery (2010) [hereafter S&M], ordinary people are willing to attribute different mental states to simple robots than philosophers traditionally have been. S&M take this as evidence that philosophers and non-philosophers have different concepts of subjective experience, suggesting that philosophers were deeply mistaken in supposing that subjective experiences were characterized by manifest phenomenality. While S&M assume that philosophers accept the traditional account, which emphasizes a dichotomy between mental states exhibiting phenomenal character and those that lack it, they argue on the basis of their own experimental data that non-philosophers are actually sorting subjective experiences based on whether or not the state in questions has a valence (or, "a hedonic value for the subject"). So S&M are engaged in two distinct projects. Their *negative* project involves demonstrating that non-philosophers do not group mental states in the manner that philosophers do, on the basis of phenomenal characters for subjective experiences. Their *positive* project involves hypothesizing about the nature of the folk concept of experience, suggesting that non-philosophers sort mental states according to an appeal to a state's valence. Our goal in this paper is to challenge these projects by questioning whether intuitive attributions of mental states to simple entities reflect the ordinary categorization-schema of mental states that people tacitly accept. Specifically, we wish to call into question the assumption, made by S&M and other experimentalists, that we should expect fundamental differences between intentional and experiential states to be reflected in mental state attributions to the sort of entities these experimentalists discuss.

Here is how we plan to proceed. In Sect. 1 we review the evidence provided by S&M in support of their negative and positive projects. In Sects. 2 and 3, we present evidence suggesting that, contra S&M, people do not ordinarily differentiate experiences according to whether they are valenced. Instead, beliefs about the functional role a mental state might play for an entity seem to be a key factor in phenomenal state attributions. Lastly, in Sect. 4 we defend philosophical orthodoxy in the philosophy of mind by arguing that these results about function undermine S&M's positive and negative projects.

## 1 The S&M robot

According to S&M, people do not ordinarily group states of subjective experience together in the same way as philosophers. If this is correct, it constitutes a fundamental challenge to a philosophical tradition that emphasizes the obviousness of phenomenal character. The assumption is that one's attributions of mental states to simple entities presumed to lack subjective experience reflect one's tacit categorization-schema for mental states. So, to get at the philosophical and folk categories of subjective experience, S&M conduct a number of experiments designed to examine

how philosophers and the folk attribute different mental states to a non-human entity: a purportedly simple robot named Jimmy. In each of their studies, S&M present experimental participants with vignettes in which Jimmy distinguishes boxes on the basis of visual or olfactory cues, receives a high-voltage shock, or interacts with a violent, competitor bot. For each scenario, they ask participants whether the robot (e.g.,) saw red, smelled bananas, felt pain, or felt angry.

In their first study, S&M compare the responses of philosophers to the responses of ordinary people. They present both sets of participants with either a case where Jimmy receives a shock and recoils in roughly the same way a biological organism might, or with a case where Jimmy successfully completes an assigned task involving identification of a red box based on its color.[1] Presented with these cases, ordinary people are willing to attribute seeing red to the robot, but not feeling pain. Professional philosophers, on the other hand, attribute mental states in a way consonant with philosophical tradition by denying Jimmy both of these experiential states.[2] According to S&M, this result suggests that people do not ordinarily group states of subjective experience together in the same way as philosophers do.

On the basis of their first study, S&M tentatively conclude that philosophers and non-philosophers have different concepts of subjective experience. The philosophical concept presumably relies heavily on the philosophers' notion of phenomenal consciousness. But to determine what underlies the folk concept, S&M go on to conduct several other studies focusing only on ordinary people's attributions across different sense modalities and for different kinds of experiential states. Surveying the results of these studies, S&M discover that ordinary participants are *sometimes* willing to ascribe subjective experiences to a robot, but that these ascriptions vary both within and across modalities (feeling, seeing, and smelling). S&M explain this pattern of responses by reference to the valence of the particular states under consideration. In one study for instance, S&M ask ordinary participants to assess cases in which Jimmy is presented with smells that participants are likely to find pleasant or unpleasant smelling (bananas and vomit) or smells with which participants are likely unfamiliar and which therefore lack either a positive or negative valence for participants (isoamyl acetate). S&M find that people are willing to say that the robot could smell the valence-neutral chemical, but were ambivalent about whether the robot could smell bananas and vomit. Figure 1 summarizes their findings across this and other of their experiments.

These results provide the evidence that S&M take to support their positive claim regarding the factors that guide folk attributions of subjective experiences. Their view is that, "in contrast to philosophers' emphasis on the phenomenology of subjective mental states, for the folk, subjective states seem to be primarily states with a valence" (320). Their hypothesis is that people do not distinguish subjective experiences by their common possession of a manifest phenomenal character, but rather in virtue of judgments concerning an experience's valence. However, is this

[1] Vignettes used in S&M's experiments, as well as in our own studies appear in the electronic supplementary materials.

[2] While there is some disagreement among professional philosophers about phenomenal consciousness, here we refer to the large consensus that thinks subjectively experienced mental states have phenomenal properties (as found in Nagel 1974; Searle 1994; Block 1995; Chalmers 1995).

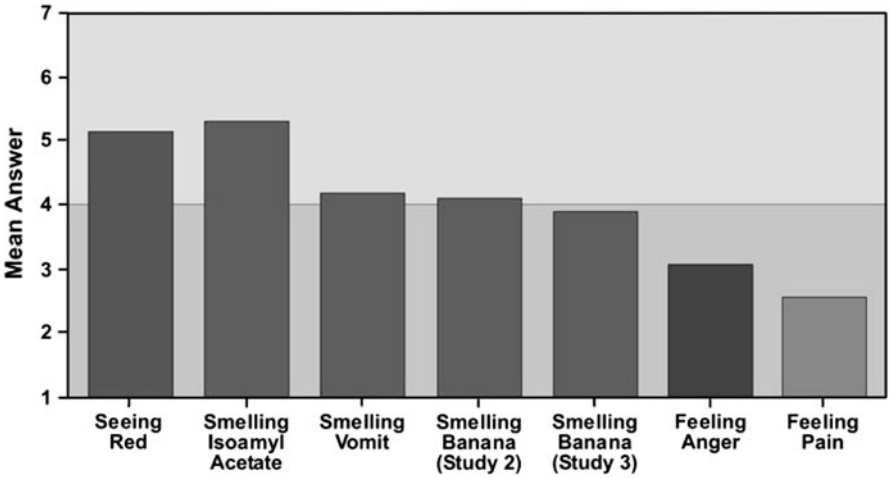


**Fig. 1** Mean folk answers for mental state ascriptions to Jimmy, within and across perceptual modality (from Sytsma and Machery 2010)

really the best explanation for S&M's pattern of results? One suggestion, recently proposed by David Rosenthal, is that ordinary attributions of mental states to simple robots might be importantly influenced by assumptions about the robot. Rosenthal points out that, "we'd need further studies to see if the folk would say that a fancier robot that, e.g., likes bananas can also smell one" (2011). We think this suggestion is very promising. Perhaps attributions of experiential states to Jimmy (in S&M's studies and beyond) might be guided, not primarily by one's beliefs about the experiential states, but rather by one's beliefs about the entity to which the experiential states are attributed. Specifically, we want to suggest that S&M's pattern of results is due to experimental participants drawing particular conclusions about the function of Jimmy relative to what particular state is being asked about in each condition. If we are correct, then the method of investigating folk categorization schemas for mental states by examining which mental states the folk will attribute to non-human entities bears rethinking.

## 2 The S&M robot smells bananas and vomit (study 1)

S&M contend that people are willing to attribute some experiential states to Jimmy because (from the participant's perspective) those states are *non-valenced*, but that they are unwilling to attribute other, *valenced* states. We want to suggest instead that this asymmetry of attributions is due to tacit assumptions on the part of experimental participants about the function for which Jimmy was created. In S&M's vignettes, function is left unspecified. Thus participants are able to draw their own conclusions about the function of the robot, and this could play an important role in explaining the asymmetry between different experiential states. For example, consider the olfactory study. While participants may be unlikely to suppose that a robot would have been

designed with the function of detecting smells salient to human beings, such as bananas or vomit, it is easy to imagine why one would invent a robot to detect a chemical with a technical-sounding name. Or consider the difference S&M find for non-philosophers between attributions of seeing red and feeling pain. In both the seeing red and feeling pain conditions it is specified that, "An instruction was then transmitted to Jimmy. It read: 'put the red box in front of the door'" (306). Thus, in both vignettes, the scientists experimenting on Jimmy treat him as though he were designed to perform visual tasks (and do not treat him as though he were designed to experience pain). On our functional hypothesis, it is no wonder then that people are more likely to attribute seeing red to Jimmy than feeling pain. In line with this hypothesis, we will argue that people attribute experiential states appropriate to satisfy the functions they assume an entity was designed to perform. And we contend that the asymmetry S&M find between the folk and philosophers is the result of the philosophers' hesitancy to draw assumptions about the entity's function when this is not explicitly mentioned in the thought-experiment-like vignettes.

To test whether experiential state attributions to a robot are guided by assumptions about the robot's function, we designed a study in which we specified functions for the robot. The robot in our vignettes was given either the function of "making smoothies" or "cleaning up bio-medical waste". Of course, specifying a robot's function might also inadvertently increase people's assumptions about that robot's level of complexity. And, it could be that the general complexity of an entity correlates with mental state ascriptions to that entity. So to control for the possible worry that any results for function might just be reflecting people's views about complex robots, rather than due to the robot's function specifically, we designed our study to manipulate the robot's function independently of its complexity. Finally, we asked participants a question intended to gauge the perceived valence associated with each of the specific olfactory objects that Jimmy interacted with (either a box that smelled like isoamyl acetate, bananas, or vomit).

This resulted in a multifactor experiment that independently manipulated the complexity of the robot, the robot's function, and objects with which the robot interacts. In our study, the description of how the robot manipulated the olfactory objects was identical to S&M's studies (so our study differed only in extra information about function and complexity). In this between-subject online experiment, 253 participants were randomly assigned to one of twelve conditions.[3] For example, below is one condition involving a non-complex robot, which has the function of disposing of bio-waste and is tasked with manipulating a box of vomit:

Jimmy (shown below) is a relatively simple robot built at a small state university. He has a scent detector, video camera for eyes, wheels for moving about, and two grasping arms with touch sensors that he can move objects with. He was created in order to clean bio-medical waste. As part of an experiment, three substances were placed under Jimmy's scent detector. The substances were presented one at a time. As they

[3] Both studies we report were run online using Amazon Mturk and Qualtrics. Because Huebner et al. (2010) found cross-cultural differences in mental state attribution in prior work; the country of origin for participants in both studies was restricted to the United States and native language to English.

were presented their names were transmitted to Jimmy: Vomit, Human Feces, and Rotting Dog Meat. The next day Jimmy was put in a room that was empty except for one box of vomit, one box of human feces, and one box of rotting dog meat. The boxes were closed, but had small holes to let the scent through. The boxes were otherwise identical. An instruction was then transmitted to Jimmy. It read: "Put the box of vomit in front of the door". Jimmy did this with no noticeable difficulty. The test was repeated on three consecutive days with the order of the boxes shuffled. Each time Jimmy performed the task with no noticeable difficulty.

In a contrasting condition, the robot had a different function ("He was created in order to make fruit smoothies"), but otherwise the vignette did not differ. After reading the vignette, participants were asked to respond on a seven-point scale (anchored at 1-clearly no, 4-not sure, and 7-clearly yes) to the question, "Did Jimmy smell vomit?" Though this was a subtle difference between the two conditions, participants were significantly more likely to claim that Jimmy smelled vomit in the condition in which his function was to clean bio-medical waste, as Fig. 2.[4]

The above figure compares only two of twelve total conditions. However, it suggests our general finding. The main result of the study was that phenomenal attribution ratings to Jimmy regarding smelling the different kinds of objects in the vignettes depend on the robot's function.[5] When the robot was designed to make smoothies, people were more likely to say that he smelled the chemical and the banana but not the vomit. Conversely, when his function was to clean up bio-medical waste, mean scores where higher for vomit than for the chemical (and to a lesser extent the banana) despite the level of complexity specified.[6] This effect is represented in Fig. 3.

Thus function seems to be an important characteristic that people consider when ascribing phenomenal states to Jimmy, independently of the level of complexity of the robot.[7] One may have wanted to explain our findings by claiming that the more complex an entity is the more experiential states it will be thought to have, and by pointing out that assumptions about functional specification are likely to correlate with assumptions about complexity. But if complexity led to greater experience judgments, one would expect greater experience ratings for our more complex robots. In fact, people's judgments about subjective experiences do not seem to depend on having a particular kind of complex robot body at all, but are instead sensitive to the robot's functional specifications. Unlike the earlier S&M result,

[4] Post hoc testing reveals that distributions in these two groups differed significantly (Mann–Whitney $U(40) = 94.0$, $Z = 3.342$, $p < 0.001$).

[5] An interaction effect was detected for the factors *Function* and *Object*: $F\ (2, 241) = 5.02$, $p < 0.01$. Further descriptive statistics are available in the supplementary materials.

[6] This difference is much smaller for banana than for chemical between conditions, though this may be explained by the fact that bananas, unlike chemicals and vomit, seem to naturally fit into both fruit and biological categories.

[7] Though it does not bear on our hypothesis directly, we also detected an incredibly complicated three-way interaction between Function, Complexity, and Object, $F\ (2, 241) = 4.67$, $p < 0.01$. This result suggests that complexity did have some impact on participants' attribution of smell to Jimmy, but that this impact of complexity depended on both the function and the type of object with which Jimmy interacted.

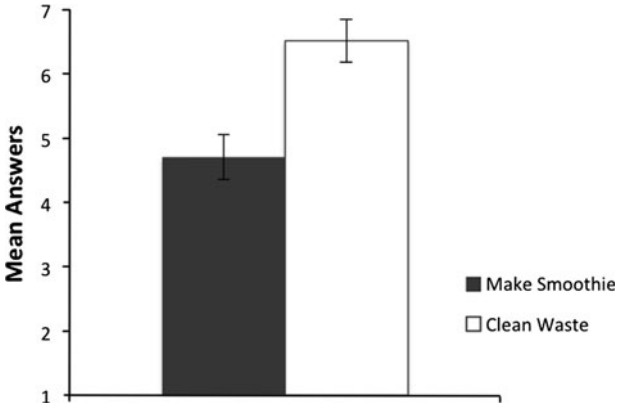


**Fig. 2** Mean answers for smell attributions of vomit in non-complex cases by function

when the function of smell is specified, the folk have no problem at all attributing valenced smell experiences to Jimmy.

But even if function is key for folk attributions of experiential states, valence could still be important in its own right. In order to gauge the level of valence associated with the different objects the robot encounters, participants were also asked a question about their own personal olfactory preferences regarding smell valence: "What kind of smell do you consider (*Banana, Isoamyl Acetate,* or *vomit*) to be?" To directly test the S&M valence hypothesis, we then compared the degree to which participants rated the valence for each smell with the degree to which they ascribed the subjective experience of smell to Jimmy.[8] However, no relationship was found between the intensity of valence ratings (of both positive and negative degree) and smell attributions.[9] Thus our findings do not seem to depend at all on assumptions about the nature of the experience itself. Instead, they seem to be explained in virtue of the robot's functional specifications. This explanation readily generalizes to S&M's findings. It is reasonable to suppose, for example, that in the absence of a specified function a robot would have been designed to detect a technical-sounding chemical, rather than to smell bananas. Therefore, the functional explanation presents the best explanation of the total pattern of evidence. It seems that the evidence for the S&M valence hypothesis is actually better explained in terms of differing beliefs about the subject of the experience, in this case

[8] Participants responded to the valence question on a seven-point scale, anchored with 1-very bad, 2-bad, 3-poor, 4-neither good nor bad, 5-fair, 6-good, 7-very good. Responses were then recoded into an "affect measure" relative to the neutral point on the scale (all responses of $4 = 0$, $3/5 = 1$, $2/6 = 2$, and $1/7 = 3$). According to the resulting affect measure, higher numbers meant a stronger valence judgment (of either positive or negative value). The strength of valence ratings were as follows: Banana ($M = 1.62$, $SD = 0.913$), Chemical ($M = 0.92$, $SD = 1.04$), and Vomit ($M = 2.51$, $SD = 0.61$). A one-way between subjects ANOVA was conducted, $F\ (2, 550) = 70.93$, $p < 0.01$, and post hoc comparisons using the Tukey HSD test indicated that the mean scores between each of these groups were significantly different ($p < 0.01$).

[9] We were not able to detect a correlation between affect (see above) and subjective experiences of smelling banana, chemical or vomit, $r\ (251) = 0.066$, $p = 0.30$.

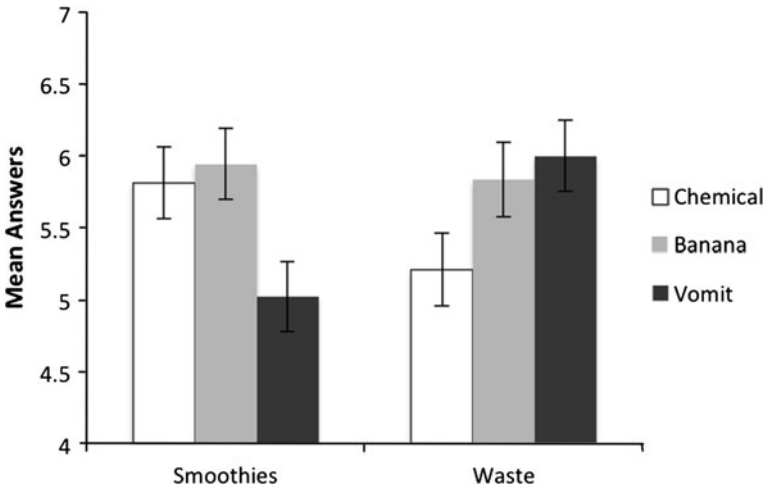


**Fig. 3** Interaction effect of object and function on smell attribution ratings in study 1 (all scales ran 1–7)

assumptions about the robot's functional specifications, rather than assumptions about the nature of the experience itself.

## 3 The S&M robot feels guilty (study 2)

So far we have seen some evidence for the claim that attributions of experiential states like smell to Jimmy depend on specifications of the robot's function, not judgments regarding smell valence. However, since S&M report findings across modalities, we also wanted to know if the assumptions people were making about the robot's function might explain their reticence to attribute subjective experiences beyond just sensory states. Perhaps people were unwilling to ascribe anger and pain to the robot in S&M's original studies because they did not see what purpose these states could have in the robot's functional repertoire. To test the range of our hypothesis, we devised a further experiment to see whether or not Jimmy would also be said to experience felt emotions when the right function was specified.

In a multifactor between subjects experiment, 118 participants again saw one of four possible cases about Jimmy. First, participants were presented with either a complex or non-complex version of the robot, very similar to the robot in the Experiment 1 smell cases. Next, they were told that the robot was either "designed to be a friend to the elderly by interpreting and responding to their emotional needs," or "to be a tool for the elderly by lifting and moving heavy objects around their houses." The remainder of each vignette read as follows:

> Jimmy has been assigned to help Abigail, an elderly woman with few possessions. An instruction was transmitted to Jimmy. It read: "Move Abigail's antique music box from the living room to the bedroom." Jimmy went to pick up the music box. But Jimmy's grip was too strong, and the box shattered into tiny pieces in his arms. Abigail walked in and began to weep

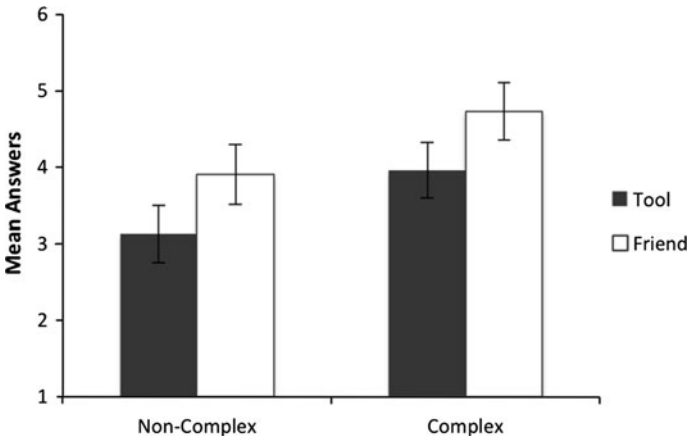


**Fig. 4** Interaction effect of object and function on guilt attribution ratings in study 2 (all scales ran 1–7)

> when she saw her shattered music box. Jimmy emitted a series of short beeps, spun around twice, and quickly drove out of the room. Later, when Abigail looked for Jimmy at his docking station, he was nowhere to be found.

After reading the vignettes, participants were asked both, "what kind of emotion do you consider guilt to be?" as well as, "did Jimmy feel guilty about breaking the music box?" For our hypothesis to be confirmed, we would expect people to be more likely to attribute the subjective experience of feeling guilt to Jimmy when they received a vignette in which a function relevant to emotional processing was specified and, contra S&M, regardless of the valence associated with guilt in the experiment.

That is exactly what we found.[10] When Jimmy's function involved being a friend for the elderly instead of being a tool for lifting, people were more likely to say the robot had the experience of feeling guilty when it broke the music box, independently of the level of complexity of the robot[11] This result is shown in Fig. 4.

As in Experiment 1, we were unable to detect any significant relationship between mental state attribution and perceived valence of the mental state.[12] Indeed, when we combine responses from both the complex and non-complex versions of

[10] Means and standard deviations for Non-complex conditions: lifting ($M = 3.13$, $SD = 2.20$), friend ($M = 3.91$, $SD = 2.10$). For complex conditions: lifting ($M = 3.96$, $SD = 2.00$), friend ($M = 4.73$, $SD = 1.80$).

[11] A two-way between-subjects analysis of variance was conducted to evaluate the effect of function and complexity on participants' attributions of guilt to Jimmy. We found a main effect for function, $F$ (1,115) = 4.19, $p < 0.05$, and a main effect for complexity $F$ (1, 115) = 4.80, $p < 0.05$.

[12] Participants responded to the valence question on the same seven-point scale as in Experiment 1 (see fn 7). They judged guilt to be moderately negatively valenced ($M = 3.39$, $SD = 1.58$, significantly less than the midpoint on the scale indicating negative valence, $t$ (118) = −4.22, $p < 0.01$). Responses were then recoded into an "affect measure" relative to the neutral point on the scale (all responses of 4 = 0, 3/5 = 1, 2/6 = 2, and 1/7 = 3). According to the resulting affect measure, higher numbers meant a stronger valence judgment (of either positive or negative value). There was no correlation detected between either the positive or negative valence associated with guilt on the affect scale ($M = 1.35$, $SD = 1.02$) and their attribution of feeling guilt to Jimmy, $r$ (118) = 0.09, $p = 0.29$.

the 'friend to the elderly' conditions, we observe that even though guilt is rated a somewhat negatively valenced emotion, participants still on average judged Jimmy to have felt this emotion.[13] These results reinforce the idea that even subjective experiences involving negatively valenced felt emotions can be attributed to a simple robot, so long as the specific emotional state is functionally useful to the robot in question.

State valence apparently does not explain these findings. However, people will attribute experiential states to a simple robot so long as such states contribute to the robot's functional repertoire. Given the aforementioned availability of relevant functional assumptions for the S&M cases, we contend that the S&M cases are explained in virtue of participants' assumptions about function rather than state-valence.[14]

## 4 Defending phenomenal consciousness

We began with the idea that the default basis for sorting mental states in philosophy of mind may have a series of important empirical commitments. If experiential states really are fundamentally different from intentional states because of their obvious and unmistakable phenomenal character, then presumably philosophers and non-philosophers will group them in similar ways. We then reviewed the work of S&M, whose studies lead them to conclude that ordinary people do not classify subjective experiences as philosophers have, relative to a state's phenomenological character, but rather by its valence. Importantly, these studies rested on the assumption that both philosophers and ordinary people conceive of the simple robot that is the target of the attributions in the same way, as incapable of the full range of subjective experience. Finally, we presented evidence suggesting that ordinary attributions of subjective experience are not sensitive to valence, but instead respond to the functional specifications of the target of the attribution, thus challenging S&M's *positive hypothesis*.

However, this still leaves the further question about what to make of S&M's *negative hypothesis*. Does our evidence also challenge the broader claim that non-philosophers group mental states differently than philosophers, and thereby provide a defense of the purportedly manifest nature of phenomenal character? To review, the key piece of evidence to support S&M's claim comes from their first experiment, in which ordinary participants (but not philosophers) are willing to attribute seeing red, but not feeling pain, to a simple robot. S&M argue that these data demonstrate that non-philosophers do not sort mental states as philosophers

[13] Mean valence rating for guilt in friend conditions ($M = 3.34$, $SD = 1.58$); mean guilt attribution rating in friend conditions ($M = 4.32$, $SD = 2.01$).

[14] Rosenthal (2011) suggests a different explanation for Sytsma and Machery's findings, which he calls the equivocation hypothesis. On this view, philosophers and the folk have at their disposal both a detection reading and an experiential reading for sensory ascriptions. Philosophers are comfortable using the experiential reading and use it in all of S&M's studies, whereas the folk tend to use the detection reading, opting for the experiential reading only when valence considerations push them to use it. See Sytsma and Machery (2010) for more discussion of this challenge.

have assumed, and that this fact challenges the manifest nature of phenomenal character. We contend, however, that phenomenal character may be an obvious fact and that the asymmetry in state attributions could be accounted for because philosophers and non-philosophers conceive of the robot in different ways.

We suggest that the difference between non-philosophers' and philosophers' responses could be due to the fact that non-philosophers (but not philosophers), lacking training in how to assess thought experiments and the philosophical tendency to restrict what is meant to what is said, are imputing likely functions to the robot, and thus attributing sensory states that seem necessary to perform these functions. Specifically, while seeing colors or smelling a technical-sounding chemical seem as though they could play a role in helping Jimmy perform whatever functions he was likely designed to perform, it is much more difficult to imagine how feeling anger or pain could play any role in reasonable robot functions. It could be that people attribute experiential states to the robot only when and in virtue of the fact that they think the states would play a certain functional role in whatever task they think the robot was designed to perform. Or it may be that people are making certain assumptions about the robot's design, such that, if they suppose that the robot is designed to perform a certain function, they will suppose that the robot was given machinery sufficient to give rise to whatever states are needed to perform that function.[15] Either explanation may underlie the differences observed by S&M between philosophers and non-philosophers without challenging the manifest phenomenal character of subjective experience.

In short, S&M's data reveal that ordinary people and philosophers exhibit different patterns of subjective state attribution when it comes to the robot in their studies. But those data support the conclusion that philosophers and the folk have different conceptions of subjective experience only on the further assumption that philosophers and the folk conceive of Jimmy in the same way across all of S&M's studies. If this assumption does not hold, then perhaps S&M's studies aren't getting at differences in how people think about mental states at all. Rather they may be due to differences in how people think about *the robot*. That is, perhaps the robot isn't the same simple robot for the folk as it is for philosophers.

Ordinary people are presumably more likely than philosophers to infer extra information beyond what is actually stated in the case of a thought-experiment-style vignette. It is easy to imagine how sense experiences could be helpful in performing typical robot functions, whereas pain or anger experiences would not. And, after all, the asymmetries S&M found in state attributions must be explained somehow, though they are apparently not explained in terms of valence. Thus, our proposal—that ordinary people (but not trained philosophers) are making different assumptions about the robot across different probes—strikes us as a plausible rival explanation of S&M's results.[16] Without further evidence against the purported manifest nature of

[15] This explanation suggests that despite our attempts to control for the robot's level of general complexity across our experiments, specifying functions may have also increased people's assumptions about the complexity of the robot *relative to those functions*.

[16] Related arguments have been given in this emerging discussion of the differences between philosophers and non-philosophers (see Williamson 2007; Devitt 2011; Nagel, forthcoming).

phenomenal experience, the philosophical conception of subjective experience stands.

**Acknowledgments** Our thanks to Richard Brown, Bryce Huebner, Joshua Knobe, Edouard Machery, Jesse Prinz, David Rosenthal, Hagop Sarkissian, Justin Sytsma, and Josh Weisberg, as well as audiences at the 2012 meetings of the Southern Society for Philosophy and Psychology and The Online Consciousness Conference for detailed and helpful comments on previous drafts of this paper.

## References

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

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227–247.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

Dennett, D. (1996). *Kinds of minds: Toward an understanding of consciousness*. Cambridge: MIT Press.

Devitt, M. (2011). Experimental semantics. *Philosophy and Phenomenological Research, 82*, 418–435.

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

Huebner, B., Bruno, M., & Sarkissian, H. (2010). What does the nation of China think about phenomenal states? *Review of Philosophy and Psychology, 1*(2), 225–243.

Knobe, J. (2008). Can a robot, an insect, or God be aware? *Scientific American: Mind,* 33–35.

Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., & Sommers, T. (2012). Experimental philosophy. *Annual Review of Psychology, 63*, 81–99.

Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and Cognitive Science, 7*, 67–83.

Nagel, T. (1974). What is it like to be a bat*?*. *Philosophical Review, 4*, 435–450.

Nagel, J. Intuitions and experiments: A defense of the case method in epistemology. Forthcoming in *Philosophy and Phenomenological Research.*

Phelan, M., Arico, A., & Nichols, S. (2012). Thinking things and feeling things: On an alleged discontinuity in folk metaphysics of mind. *Phenomenology and the Cognitive Sciences*. Online first. doi:10.1007/s11097-012-9278-7

Robbins, P., & Jack, A. (2006). The phenomenal stance. *Philosophical Studies, 127*(1), 59–85.

Rosenthal, D. (2011). *Mental quality, valence, and intuition: Comments on Edouard Machery*. Presented at the metropolitan experimental research group's annual experimental philosophy conference, March 26, 2011. https://wfs.gc.cuny.edu/DRosenthal/www/DR-MERG.pdf

Searle, J. (1994). *The rediscovery of the mind*. Cambridge: MIT Press.

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

Williamson, T. (2007). *The philosophy of philosophy*. Malden: Blackwell.