

0. Abstract

Two key questions dominate the experimental philosophy of mind literature: Do ordinary people possess a concept of phenomenal consciousness? What features of an entity lead people to attribute phenomenally conscious (and non-phenomenally conscious) mental states? Five responses to one or both of these questions are discussed: 1) an Embodiment View, which claims that agents capable of phenomenal consciousness are thought by the folk to possess a certain kind of biological body; 2) a Valence View, which emphasizes the importance of considerations of affective valence in mental state attribution; 3) a “Naïve” View, which claims that pains and colors are just thought by the folk to be physically localized properties; 4) a Functionalist View, which argues that functionalist considerations predominate ordinary peoples attributions of all varieties of mental states; and 5) a Dual-Systems View, which claims that all of the other views are investigating rational mental state attributions and have missed the significant part of the story involving automatic mental state attributions. Experimental philosophical evidence for and against each response is considered. The chapter concludes with some brief remarks about other projects in the experimental philosophy of mind.

1. Introduction: Experimental Philosophy of Mind, Positive and Negative

Philosophers have used a variety of experimental methods to investigate topics squarely within the philosophy of mind. Indeed, in one of the earliest and most influential papers of contemporary experimental philosophy, Knobe (2003) investigated how people attribute the mental capacity of intentionality and uncovered a now famous asymmetry in our attributions of intentional states. Much experimental philosophy of mind has followed this early example, investigating folk theory of mind, in general, and when and why people attribute mental states of

different kinds, in particular. In an important early paper, Knobe and Prinz (2008) investigated the attribution of mental states as such, and posited the existence of a folk psychological concept of phenomenal consciousness. They argued that such a posit was needed to explain their finding that people were less willing to attribute presumed phenomenal mental states, such as feeling sad, to group entities, such as corporations, than they were non-phenomenal mental states, such as thinking (whereas they were willing to attribute both kinds of states to other, non-group entities). In recent years, numerous other works of experimental philosophy have expanded and complicated our understanding of mental state attribution.

If one is inclined to accept that the identity conditions for at least some mental states are imposed by our intuitive folk psychological theories, then one will likely regard the results of the previous examples of experimental philosophy of mind as of instrumental value. Experimental philosophical projects of this kind purport to augment the philosopher of mind's toolbox with the tool of scientifically-informed, experimental studies of people's intuitions about mental states. Thus, they can be seen as of a piece with more traditional, non-experimental, analytic philosophy of mind, since both have as their primary goal the investigation and further understanding of the metaphysics of minds and the epistemology of mental state attribution and since both typically appeal to philosophical intuitions to advance their arguments.¹ Such projects, then, make positive contributions to analytic philosophy of mind. However, a negative project also exists in experimental philosophy of mind, which aims to undermine arguments that appeal to philosophical intuitions. Negative projects intend to show that some purportedly generally held philosophical intuition, to which some philosophical argument appeals, is not, in fact, generally

¹ Recently, a number of theorists have argued for the surprising thesis that analytic philosophers don't actually use intuitions as evidence, including Williamson (2007) and Cappelen (2012). However, a number of others have offered compelling challenges to this position, including Nado (2015) and Climenhaga (2018).

held, or is not generally held in such a way as to support the argument it is intended to support.

Because both non-experimentalists and positive experimentalists appeal to intuitions in making their arguments, either may end up as the target of a negative experimental philosopher of mind (as we shall see below).

In this chapter, I will focus, first and primarily, on the experimental philosophy of conscious state attribution. A mental state may be said to be conscious in a variety of ways, but experimental philosophers of mind have primarily been interested in phenomenal consciousness. As I explain below, experimental philosophers of mind have asked whether there are folk correlates to the philosopher's concept of phenomenal consciousness and when and why we attribute phenomenally conscious mental states to some entities but not others. In Section 2, the main part of this chapter, I examine several different accounts of phenomenal state attribution that have emerged from the experimental philosophy literature. In Section 3, I will briefly introduce several other topics of interest in the experimental philosophy of mind.

2. The Experimental Philosophy of Conscious State Attribution

Theorists have distinguished between mental states that are, in some sense, conscious and other states that are not conscious, in that sense. For example, Freud claimed that repressed mental attitudes are not first-person accessible, and hence are unconscious, in a sense. At this point, many psychological studies have partially vindicated Freud and demonstrated that individuals often lack first-person awareness of various mental attitudes and processes (e.g., judgments, decisions, belief formation processes) that are, nonetheless, operative in their behavior.² The idea

² Nisbett and Wilson (1977) cover numerous studies demonstrating that “we may have no direct access to higher order mental processes such as those involved in evaluation, judgment, problem solving, and the initiation of behavior” (232). More recently, arguments by analogy from research on commissurotomy (“split-brain”) subjects (see especially Gazzaniga 1995, 2000) to the conclusion that we quite generally lack first-person awareness of our

that people possess mental attitudes that are unconscious—in the sense of first-person inaccessible—is now widespread in western culture, to the point that we routinely attribute unconscious attitudes, as when we attribute implicit bias to someone.

However, many philosophers of mind have been interested in so-called phenomenal consciousness. According to Block (1995), “phenomenal consciousness is experience; what makes a state phenomenally conscious is that there is something ‘it is like’ (Nagel, 1974) to be in that state” (p. 228). Chalmers (1995) writes of the ‘subjective aspect’ of experience, claiming that sensory states, bodily sensations, emotions, and mental images are all united in that “there is something it is like to be in them. All of them are states of experience” (p. 201). Not only do philosophers of mind such as these countenance phenomenal consciousness, they further maintain that its existence is manifestly obvious to anyone who experiences phenomenally conscious states—that is, to all of us. For example, Chalmers (1995) goes on to claim that “experience is the most central and manifest aspect of our mental lives” (p. 206). Also, Jackson’s (1986) *Knowledge Argument* relies on the undefended premise that Mary will come to know about the phenomenal experience of seeing red upon her release from her monochromatic internment. Presumably, Jackson supposes that his readers will find this just obvious. Finally, these philosophers of mind are committed to the assumption that people attribute phenomenally conscious mental states to some but not all other entities. For example, Jackson (1986) goes on to claim that

...after Mary sees her first ripe tomato, she will realize how impoverished her conception of the mental life of *others has been all along*. [...] All along their experiences (or many of them, those got from tomatoes, the sky, ...) had a feature conspicuous to them but until now hidden from her... (pp. 292-293).

attitudinal states have been offered by Carruthers (2013: 39–45). (Carruthers 2013 is, in fact, a compelling argument from numerous pieces of experimental evidence involving a variety of populations for this conclusion.) See also Carruthers (2010).

And Nagel's (1974) "What is it Like to Be" Argument relies on our tacit acceptance that it *is* like something to be a bat (but maybe not some other living things):

I assume we all believe that bats have experience. After all, they are mammals, and there is no more doubt that they have experience than that mice or pigeons or whales have experience. I have chosen bats instead of wasps or flounders because if one travels too far down the phylogenetic tree, people gradually shed their faith that there is experience there at all (438).

Presumably, neither Jackson nor Nagel would maintain that we attribute phenomenally conscious states to a stone, a dead person, or a plant. Block's (1978) Nation of China Argument appeals to our intuition that the citizens of the nation of China, organized so as to instantiate the machine-table description of, say, the state of pain, would not thereby instantiate a pain state.

As these quotations suggest, some prominent analytic philosophers of mind suppose that there is a class of phenomenally conscious mental states that are obvious to those who experience them and are also attributed by those who experience them to some but not all other entities. Inspired by these discussions of phenomenally conscious mental states, experimental philosophers of mind have investigated two main questions:

1. Do ordinary people possess a concept analogous to the concept of phenomenal consciousness employed by analytic philosophers of mind?
2. What features of an entity lead people to attribute the mental states that philosophers identify as phenomenally conscious mental states?

In the remainder of this section, I'll survey some projects that have attempted to answer one or both of these questions.

2.1 The Embodiment Hypothesis

An early inspiration for the experimental philosophy of conscious state attribution was provided by the psychologists Gray, Gray, and Wegner (2007), who surveyed ordinary people's preferential attributions of different sorts of mental states to different sorts of entities. They had thousands of participants complete one survey each, consisting of 78 assessments as to which of two characters was more capable of a particular mental state. Participants selected the state they would assess, out of 18 possibilities (e.g., feeling pain, experiencing pride, exercising restraint, telling right from wrong). Each participant assessed half of the 156 possible pairings of 13 different characters (e.g., an infant, a chimpanzee, a robot, a dead woman, God). So, if a participant had selected the survey on the affective state of *pride*, they might be asked whether and to what degree a robot is more capable of experiencing pride than an infant, then asked whether a chimpanzee is more capable of experiencing pride than a person in a persistent vegetative state, and so on. Gray and colleagues subjected these results to a principle component factor analysis, and identified a two-factor model that accounted for 97% of the variance in people's assessments. One factor, which they called "Experience", explained 88% of the variance and included capacities such as feeling fear and feeling hunger. Another factor, which they called "Agency", explained 8% of the variance and included capacities such as self-control and moral reasoning.³ As shown in Figure 1, participants assessed some entities (a five-year old girl, the family dog) as more capable of Experiential states than some other entities (a robot, God). And they assessed some entities (God, an adult woman) as more capable of agentic states than some other entities (a fetus, a dead woman).

³ The experience dimension was subsequently confirmed in replications by Takahashi, Ban, and Asada (2016) as well as Weisman, Dweck, and Markman (2017), though the agency dimension has proved less robust. See Malle (2019) for further discussion.

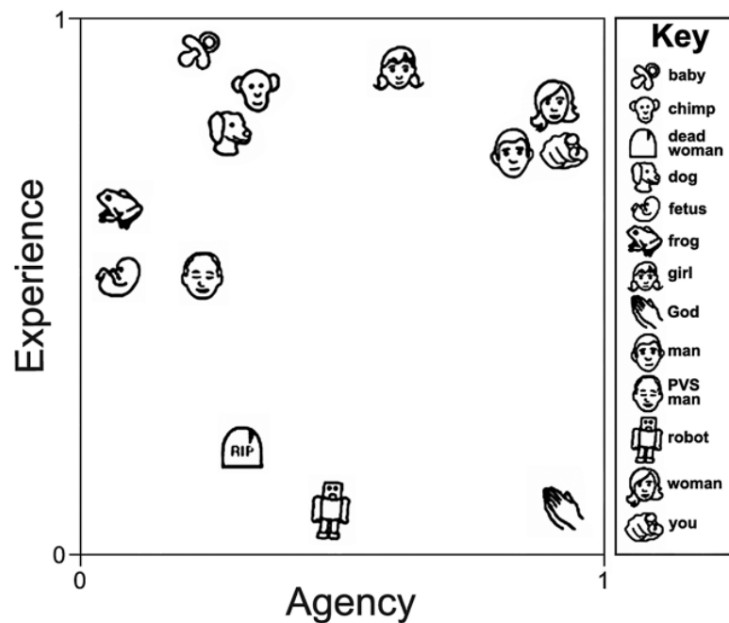


Figure 1: Adjusted character factor scores on the dimension of mind perception (from Gray, Gray and Wegner 2007, pp. 619)

One thing that stands out about Gray, Gray, and Wegner's findings is that the entities assessed as least capable of experiential states—a dead woman, a robot, and God—are all alike in lacking a functioning biological body. But, despite lacking a biological body, God was still deemed highly capable of agentic states. The experiential states used by Gray and colleagues are bodily sensations and emotions, so, they are prototypical phenomenally conscious mental states (and their agentic states are not). Their results, then, seem to suggest that considerations of bodily constitution weigh heavily in ordinary attributions of at least some phenomenally conscious mental states, but not in the attribution of other, non-phenomenally conscious mental states.

Building on these results, Knobe and Prinz (2008) used a series of experimental studies to argue that ordinary people possess an implicit concept of phenomenal consciousness and that “information about physical constitution plays a special role in those ascriptions that require phenomenal consciousness—a role it does not play in other kinds of mental state ascription” (p. 71). Knobe and Prinz’s studies focused on people’s attributions of mental states to groups—to corporations, specifically. They pointed out that groups, being composed of individual agents, have a radically different constitution than a human being. Yet we seem willing to attribute some mental states to group entities—we are perfectly willing to talk about what Apple thinks or what the Republican Party wants, for example. In one study Knobe and Prinz gave participants sentences that ascribed phenomenal states or (purportedly non-phenomenal) intentional states (such as believing or intending) to a fictional *Acme Corporation*. They then asked participants to rate how natural each sentence sounded on a scale anchored at “sounds natural” and “sounds weird”. Group phenomenal state ascriptions sounded significantly weirder than group intentional state ascriptions to these participants. Knobe and Prinz inferred that these results “seem to indicate that people are unwilling to ascribe to group agents states that require phenomenal consciousness” (p. 75). In another study, Knobe and Prinz showed that utterances such as “Acme Corp. is feeling upset” or “Acme Corp. is feeling regret” are judged to sound much weirder than utterances such as “Acme Corp. is upset about the court’s recent ruling” or “Acme Corp. regrets its recent decision”. They concluded from this result that, in addition to the folk psychological concepts of *regret* and *upset*, ordinary people also possess a concept of phenomenal consciousness, or *feeling*. And that, “people’s reluctance to say that a group agent ‘feels upset’

stems not from the criteria associated with their concept of upsetness, but rather from the criteria associated with their concept of phenomenal consciousness” (p. 78).⁴

Reflecting on the results from Gray, Gray and Wegner (2007) as well as Knobe and Prinz (2008),⁵ we may ask ourselves what common feature a robot, God, and a corporation share that inhibits people’s attributions of phenomenally conscious mental states to them? According to Knobe and Prinz, this cannot be a feature related to how these entities function, since “a state of a corporation easily could have a functional role similar to one that people ordinarily associate with feeling depressed” (p. 73). Intuitively, God or a robot could react to their environment in exactly the same way as an enraged person. At least *prima facie*, then, it seems that these three entities can occupy states functionally equivalent to the ordinary humans to whom we reflexively attribute phenomenal states. What God, a robot, and a corporation lack, however, and what a human possesses, is a unified, biological body. These findings, then, may support an Embodiment Thesis, which regards unified biological embodiment as a psychological cue to attributions of phenomenally conscious mental states. The Embodiment Hypothesis claims that ordinary people tacitly possess a concept of phenomenal consciousness roughly equivalent to the one philosophers of mind use and that preferentially apply it to entities that exhibit the non-functional feature of unified, biological embodiment.

Evidence for the Embodiment Hypothesis has been challenged. For example, Huebner, Bruno, and Sarkissian (2009) presented versions of Knobe and Prinz’s corporation probes to students at Hong Kong University, who were significantly more comfortable attributing

⁴ Knobe and Prinz use these and additional results to argue for a somewhat radical conception of human nature. They accept that information about the non-phenomenal, intentional mental states of other entities—information about their beliefs and desires—is useful for predicting and explaining behavior. But they maintain that we primarily exploit information about the phenomenal mental states of other entities to determine what our moral obligations to those entities are, and not primarily to predict and explain their behavior.

⁵ See also Knobe (2008) as well as Robbins and Jack (2006).

phenomenal states to distributed group entities than were their American counterparts.

Embodiment tendencies, then, insofar as they exist, seem to be malleable and mutable, not universally held. Sytsma and Machery (2009) challenged the idea that people possess a concept of phenomenal consciousness that they express with the word *feeling*. They claimed that the sentence, “Acme Corp. is upset about the court’s recent ruling,” sounds more natural than the sentence, “Acme Corp. is feeling upset”, simply because the mental state attribution is contextualized with a prepositional phrase in the former but not the latter. They were able to reverse the effect Knobe and Prinz found, getting participants to judge that the sentence “Acme Corporation is feeling upset about the court’s recent ruling” sounds more natural than the sentence “Acme Corporation is upset”. Similarly, Arico (2010) found no difference between people’s naturalness assessments of “McDonalds is upset” and “McDonalds is feeling upset” when each was contextualized with the addition of the prepositional phrase “about the court’s recent ruling”. Finally, Phelan et al. (2013) found that participants were equally likely to choose ‘they’ (rather than ‘it’) to paraphrase a group name in group phenomenal state attributions and group intentional (non-phenomenal) state attributions. They were significantly more likely to choose ‘it’ to paraphrase a group name in non-mental state attributions. This suggests that people understand *all* group mental state attributions—phenomenally conscious or not—distributively, as attributions of states to the constituents of groups. If correct, Knobe and Prinz’s results tell us very little about how people think about the mental states of groups, as opposed to how they think about the people who constitute groups.⁶

⁶ Concurrently, with these negative projects, positive projects within experimental philosophy of mind seemed to demonstrate that, contra Knobe and Prinz, people are perfectly happy to attribute phenomenally conscious states to entities that lack a biological body. We’ll review such projects in Sections 2.2 and 2.3 below.

Despite these challenges to the Embodiment Hypothesis, the role that an entity's body plays in our attributions of emotions, perceptual states, bodily sensations, and other phenomenally conscious states has remained a topic of interest. For example, Forstmann and Burgmer (forthcoming), have recently purported to provide evidence that people conceptualize consciousness as a spatio-temporally located process in the human brain. Though, it is doubtful that their method of asking people to color the part(s) of a two-dimensional picture of a brain that they think "contribute" to consciousness delivers much evidence as to where people think consciousness is located, since something can clearly contribute to a process without being collocated with it. Meanwhile, Diaz (2021a) has presented experimental evidence that "people report anger, sadness, and fear in the absence of bodily feelings" (24). If correct, this suggests that bodily responses are not essential to ordinary people's conception of emotions and, thereby, undermines one source of support for a sort of embodiment view.

2.2 The Valence View and the Naïve View

The Embodiment Hypothesis answered the first of experimental philosophy of mind's key questions affirmatively. People tacitly employ a concept of phenomenal consciousness analogous to that employed by analytic philosophers of mind, it claimed. Further, it maintained an answer to the second question that when we decide to predicate phenomenally conscious mental states to entities, we do not merely consider the functional capacities of those entities. We also consider physical properties that are irrelevant to functional capacities. The Valence View and the Naïve View, however, are alike in giving a negative answer to the first of experimental philosophy of mind's key questions. Both maintain that ordinary people do not employ a concept of phenomenal consciousness analogous to that employed by analytic philosophers of mind.

Instead, they contend that the diverse set of sensory states, bodily sensations, emotions, and mental images that analytic philosophers have identified as phenomenally conscious are conceptualized in a variety of different ways by ordinary people, no one of which approximates the analytic philosopher's concept of a phenomenally conscious mental state. Both views maintain that the analytic philosopher's commitment to the manifest obviousness of phenomenal consciousness was, in fact, the product of biased post-theoretic reflection on what is pre-theoretically obvious. However, none of this means that the Valence View or the Naïve View are off the hook when it comes to the second question. In fact, they have complicated the task, since either view would seem to entail that we need a variety of different accounts for how people apply each of the motley set of mental states philosophers mistakenly unify under the heading 'phenomenally conscious states'.

According to the philosophically prominent conception of phenomenal consciousness described above, we must accept the existence of phenomenal consciousness because its existence is obvious to anyone who experiences it and we all continually experience it. Proponents of this prominent conception have, further, identified both bodily sensations (e.g., the experience of physical pain) and sensory states (e.g., the state of seeing something red) as prototypical instances of phenomenally conscious mental states. According to this view, then, it seems that anyone who occupies the mental state of physical pain should experience that state as phenomenally conscious—that is, as a state that it is like something to occupy. And the same follows for anyone who occupies the state of seeing something red. Presumably, given the obviousness of phenomenal consciousness, after a little experience of bodily sensations and sensory states like these, most everyone would associate the two sorts of mental states together as being phenomenally conscious—that is, as being like something to occupy. Thus, Sytsma and

Machery (2010) argue that ordinary people should tend to treat attributions of paradigmatic examples of phenomenally conscious mental states similarly and similarly to how people with philosophical training treat them.

With this argument in mind, Sytsma and Machery designed a set of vignettes to test whether ordinary people attribute bodily sensations and sensory states in the same way as philosophers. In their first study, Sytsma and Machery split their sizable sample into two groups: “Philosophers”—including undergraduate majors and participants with some graduate training in philosophy—and “Non-Philosophers”. Participants in each group were presented with one of four vignettes. Two vignettes described a normal human (Timmy), the other two described a simple, non-humanoid robot (Jimmy). In every vignette, the agent was instructed to place a red box in front of a door as part of a psychological experiment. The agent successfully completed the task in two of the vignettes. In the other two, the box gave the agent “a strong electric shock” and, “He let go of the box and moved away from it” (306). For each vignette Gray, Heather, Kurt Gray, and Daniel Wegner (2007). “Dimensions of Mind Perception.” *Science*, 315, 619.

, the participant was asked whether the agent ‘saw red’ or ‘felt pain when he was shocked’, and answered on a 7-point scale anchored at 1 with “clearly no”, at 4 with “not sure”, and at 7 with “clearly yes”. As the results in Figure 2 show, the responses of the philosophers diverged from those of the non-philosophers when it came to the purported phenomenal states of seeing red.

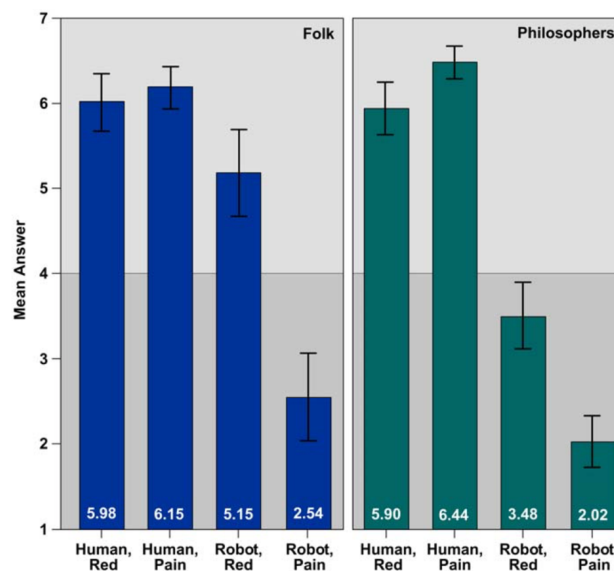


Figure 2: Mean attributions of the mental states of feeling pain and seeing red (from Sytsma and Machery 2010, pp 307)

As Sytsma and Machery report, philosophers' mean attributions of the mental states of feeling pain and seeing red were significantly above the midpoint for Timmy, the normal human, and significantly below the midpoint for Jimmy, the simple robot. For non-philosophers, mean attributions of both mental states were significantly above the midpoint for Timmy, as was the mean attribution of seeing red to Jimmy. Only attributions of feeling pain to Jimmy were below the midpoint for non-philosophers.

On the basis of their first study, Sytsma and Machery tentatively conclude that non-philosophers lack the philosopher's concept of phenomenal consciousness. They then strengthen this tentative conclusion with a series of experiments that serve to develop their positive view of how people think about sensory states and bodily sensations. They ultimately conclude that philosophers and ordinary people agree that simple robots are not capable of what we may describe as subjectively rich experiences. It's just that philosophers and ordinary people part

ways on what counts as ‘subjectively rich’. Philosophers identify the category of phenomenally conscious mental states with subjectively rich experiences. Ordinary people, on the other hand, sort experiential states according to their valence and identify positively or negatively valenced mental states with subjectively rich experiences.⁷

This Valence View explains Sytsma and Machery’s findings. Because feeling pain is negatively valenced, ordinary people, like the philosophers operating with their concept of phenomenal consciousness, regard this as a subjectively rich experience to be denied a simple robot. But ordinary people are willing to attribute the state of seeing red to the simple robot because they do not regard it as a subjectively rich (that is, valenced) experience. Philosophers, though, are unwilling to attribute seeing red to the robot, because they see it as a subjectively rich (that is, phenomenally conscious) experience.⁸ Sytsma and Machery (2010) conclude that, “in clear contrast to philosophers, the folk do not seem to believe that there is something common to all these mental states—namely that they are phenomenal” (p. 319).

Several studies appeared have challenged the Valence View. For example, Buckwalter and Phelan (2013), designed a pair of studies based on the simple robot vignettes. In their first study, they asked participants whether a robot smelled something, either isoamyl acetate, bananas, or vomit. These sensory states were chosen because, in one of their studies, Sytsma and Machery had found that participants were more inclined to say the robot smelled isoamyl acetate (presumably unknown to these participants, and, so, unvalenced) than either vomit or bananas. However, Buckwalter and Phelan varied the functional complexity of the robot and the

⁷ Sytsma and Machery (2010) helpfully clarify their use of the term ‘valence’ in a footnote: “mental states have a valence if and only if they have a hedonic value for the subject. That is, mental states have a valence if and only if they are pleasurable (they then have a positive valence) or disagreeable (they then have a negative valence). Not all mental states have a valence, and valenced states are more or less pleasurable or disagreeable” (p. 350).

⁸ A similar explanation can be applied to the results Sytsma and Machery find in their other studies for the mental states of smelling vomit, bananas, and isoamyl acetate.

functional role the robot was designed to perform across conditions. They also had participants rank the valence of the three sensory objects and found that participants attributed little positive or negative valence to isoamyl acetate, a slight positive valence to bananas, and a strong negative valence to vomit. They then compared the degree to which participants rated the valence for each smell with the degree to which they attribute the subjective experience of smelling the object to the robot. Contra the Valence View, no relationship was found between the intensity of positive or negative valence ratings and smell attributions. However, there was a strong relationship between the functional role a robot was designed to perform and people's attributions of sensory states. 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 functional role was to clean up biomedical waste, mean scores were higher for vomit than for the chemical (and to a lesser extent the banana).

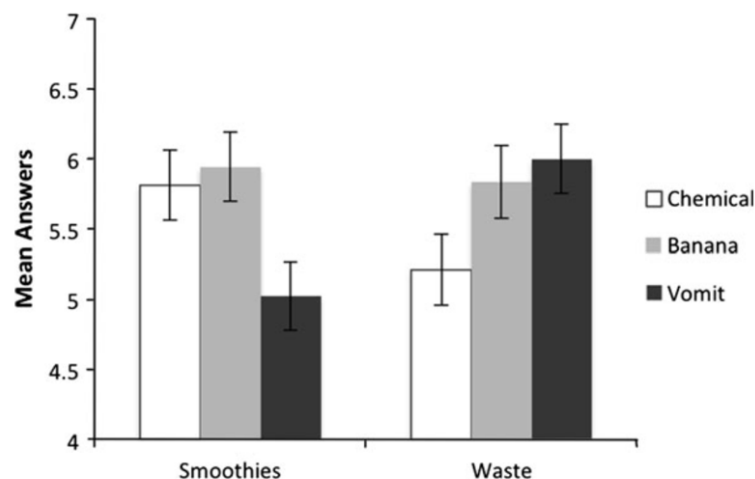


Figure 3: Interaction effect of object and function on smell attribution ratings (from Buckwalter and Phelan 2013, pp 357)

In their first study, Buckwalter and Phelan did not find an effect for complexity of the robot. However, in a second study, participants' assessments of whether or not a robot could feel guilt were influenced both by its complexity and its specified functional role. But, again, in this study, participants' valence assessments for the emotional state of guilt were not predictive of their guilt attributions. Finding no evidence of a valence effect, Buckwalter and Phelan explained Sytsma and Machery's results by referencing participant's assumptions about function, rather than valence. Specifically, they argue that participants likely assumed that a simple robot in a psychological study that was directed to move a red box was designed to satisfy the functional role of detecting colors and with the mechanical complexity to do so. Such a robot was probably not designed to feel pain when it is damaged, especially since the shock it sustains is likely to be interpreted as the result of an accident, rather than a planned intervention by its designers.⁹

To his credit, after his own studies failed to present evidence in favor of the Valence View, Sytsma (2012) abandoned it, arguing from previous research and original experiments to the conclusion that ordinary people tend to "hold a naïve view of both colors and pains" (190). According to this Naïve View, ordinary people take "colors to be mind-independent qualities of worldly objects" and pains "to be mind-independent qualities of certain body parts" (188). The Naïve View offers an explanation of Sytsma and Machery's findings. Ordinary people thought the simple robot was incapable of feeling pain but not seeing red because they thought that it could see the mind-independent color inherent in the box in front of it, whereas they also thought

⁹ See also Fiala, Arico, and Nichols (2014), who provide evidence that few people describe the robot in Sytsma and Machery's (2010) vignettes as 'seeing green' when given options to describe it, instead, as 'detecting green' or 'identifying the green box' (whereas most people describe the person as seeing green). As they explain, "These results suggest that the overly narrow choice format might artificially inflate the attributions of perceptual experience to robots in Sytsma and Machery's (2010) studies" (Fiala, Arico, and Nichols 2014: 39).

it lacked the appropriate, fleshy body, in which a pain state might occur.¹⁰ Sytsma (2012) offers three additional studies supporting the Naïve View. The first reveals that participants are more willing to attribute pain to someone who burns their bioengineered, fleshy prosthetic hand, than to someone who burns their robotic, synthetic prosthetic hand, supporting the idea that people tend to view pain as a mind-independent quality of certain kinds of body parts. In the second, he finds that participants are just as willing to attribute the experience of seeing red to someone with robotic eyes as to someone with bio-engineered eyes, suggesting that bodily constitution doesn't matter for seeing red. In the third study, he shows that the results for pain generalize to robot agents, by showing that participants are more willing to attribute pain states to a robot agent with fleshy, bioengineered hand than to a robot agent with hard, metallic hands.

While the Naïve View is similar to the Embodiment Hypothesis in advocating that considerations of physical makeup play an important role in ordinary attributions of some prototypical phenomenal mental states, it differs from the Embodiment Hypothesis in claiming that this is true only for pain states and other states that ordinary people regard as adhering in particular types of bodies. The Naïve View, like the Valence View, and contra the Embodiment Hypothesis, also maintains that ordinary people lack a concept of phenomenal consciousness, for at least two reasons. First, as Sytsma (2012) concludes:

While philosophers tend to treat the qualities at issue for ascriptions of seeing red and feeling pain as mind-dependent qualia...lay people understand [these] two prototypical examples of (supposed) phenomenally conscious mental states in a way that is fundamentally different from the standard philosophical account (195).

¹⁰ While the Naïve View doesn't explain Buckwalter and Phelan's (2013) findings, those findings are at least compatible with the Naïve View, since the Naïve View itself is consistent with a broadly functionalist picture, offering, as it does, only a metaphysical thesis about the kinds of things ordinary people take pains and colors to be.

Second, though not noted by Sytsma (2012), while a Naïve View may be plausible for the supposed phenomenal states of seeing red or feeling pain, it is implausible to suppose that other states that philosophers identify as prototypically phenomenal are thought by ordinary people to be mind-independent qualities of worldly objects. In what worldly object or part of the body would sadness reside, for example? What about a hallucination of a pink elephant? Thus, ordinary people do not think of all the states that philosophers group under the heading “phenomenally conscious mental states” in the same way.

In other papers, Sytsma and colleagues have deepened the case for the Naïve View of colors and pains. Sytsma (2010) included numerous findings that could be seen as consistent with the Naïve View. In one study, participants tended to answer in a way that suggested they think of the redness they experience when they see a ripe tomato as not in their minds, but out there, on the tomato, an objective property that is sure to be experienced in the same way by any normal-sighted individual. In another set of studies, participants tended to respond that two agents who share a body part (e.g., conjoined twins) experience the same pain when they injure their shared part, challenging the analytic philosopher of mind’s conception of pain as a private mental state locked away in each individual’s mind. Another set of studies found that participants were likely to judge there was an unfelt pain, when presented with vignettes describing seriously injured patients who reported they hadn’t been aware of pain while watching an engrossing movie or reading a compelling book, suggesting that ordinary participants don’t think of pain as a necessarily conscious state. Reuter and Sytsma (2020) deepened this last result through 17 additional studies, which demonstrated that ordinary participants are happy to say that people experience unfelt pains. In another vein of research, Reuter, Phillips, and Sytsma (2014) offered a series of studies suggesting that ordinary people

think it's possible to hallucinate a pain—i.e., to have an experience as of a pain when no pain is, in fact, present. This accords with a perceptual model of pains, since perceptions are subject to an appearance-reality distinction, and it runs counter to the views of analytic philosophers of mind, such as Hillary Putnam (1963), who claims that “any situation that a person cannot discriminate from a situation in which he himself has a pain *counts* as a situation in which he has a pain” (218). An appearance-reality distinction for pain is further supported by an argument Reuter (2011) as well as Sytsma and Reuter (2017) make from their findings that ordinary English and German speakers preferentially use the phrase ‘having pain’ (or its German analogue) to express mild pains and ‘feeling pain’ (or one of its three German analogues) to express large pains. If we can take these attributions literally (as Sytsma and Reuter argue we can), then ordinary people seem to acknowledge pains you have but don't feel. In yet another project, Reuter, Sienhold, and Sytsma (2019) present a challenge to the ‘finger in mouth’ argument against the Naïve View of Pain. According to that argument, if a pain is actually located in my finger (as the Naïve view would claim) and I put that finger in my mouth, then it should follow that I have a pain in my mouth. But, it seems incorrect to say that I have a pain in my mouth in that situation. Reuter and colleagues contend that, even though it follows that one does have a pain in one's mouth in this situation, we would be hesitant to say that, because having a pain in a body part—like having tissue damage in a body part—tends to imply something (a pain) adhering in the substance of the body part, but the premises of the argument do not suggest that there is pain adhering in the mouth's substance. They provide evidence that, if you cancel these implications, people are happy to say there is a pain in the person's mouth. Finally, Kim and colleagues (2016) began to make the case that the naïve conception of pain is a

cultural universal, by presenting data from two studies suggesting that Americans and South Koreans “overwhelmingly conceive of pains as bodily states” (p. 163).

At this point, proponents of the Naïve View have compiled a great deal of evidence for the claim that ordinary people sometimes think of pains as mind-independent qualities of certain body parts and a bit of evidence for the claim that they sometimes think of colors as mind-independent qualities of worldly objects. But what exactly is the scope of these claims? Is it simply that ordinary people *sometimes* think of pains and colors in a mind-independent way? Or is it the much stronger claim that ordinary people *only* think of pains and colors in this way? Or is it some middle view that is being suggested? Unfortunately, the above papers are not very clear on this point. Sytsma and Reuter (2017) is typical. In their conclusion, the authors claim that “the accumulated research suggests that people tend to conceive of pains not as mental states, but as bodily states” (p. 23), which seems to suggest a folk *tendency* towards the Naïve View, but to also leave open the possibility that people sometimes think of pains in other ways. In fact, this conclusion seems to be well supported by the data, since, in each study on pain attributions a (sometimes sizeable) minority of participants seemed to endorse a mentalistic view of pain. However, Sytsma and Reuter (2017) also refer to the Naïve View as “*the* commonsense conception of pains” (p. 23, italics added). And if proponents of the Naïve View want to maintain, as Sytsma (2012) did, that ordinary people “lack the philosophical concept of phenomenal consciousness” (p. 183), then they will have to adopt the additional assumption that the common-sense concept of pain they have identified operates to the exclusion of other, mentalistic conceptions of pain. But Sytsma and colleagues have not defended this additional assumption. And recent work by Borg, Salomons, and Hansen (2019) as well as Salomon and

colleagues (2021) suggests considerable conceptual diversity in the ways in which ordinary people can and do think of pain.

2.3 Functional Considerations and the Agency Model

Though radically different in a variety of ways, the previous accounts of phenomenal state attribution are alike in offering definitive (though distinct) responses to the question of whether ordinary people possess a concept of phenomenal consciousness. They are also alike in applying only to attributions of states that philosophers would consider phenomenally conscious (or to subsets of those states), not to attributions of mental states in general. The views I will consider in this section differ from the previous ones on both counts. These views claim that we are not yet in a position to say whether people possess a concept of phenomenal consciousness. They also claim that our attributions of states that philosophers would consider phenomenally conscious are guided by the same sorts of considerations that guide our attributions of mental states in general. We'll consider the second of these claims first.

It is worth noting that, before the analytic philosopher's conception of phenomenal consciousness came to prominence, many philosophers tended to emphasize quite different aspects of our ordinary understanding of experiential states such as pain. These philosophers did not emphasize the, purportedly obvious, qualitative aspects of experiential states, instead they suggested that ordinary conceptions of experiential states (like other mental states) centered on the functional roles those states played. So, for example, Lewis (1966) contends that, "the definitive characteristic of any experience as such is its causal role. The definitive causal role of an experience is expressible by a finite set of conditions that specify its typical causes and its typical effects under various circumstances" (pp. 19f.). And he goes on to assert that this is, "an

account of the parlance common to all who believe that experiences are something or other real and that experiences are efficacious outside their own realm” (p. 20). Armstrong (1968) suggests, “that what we mean when we talk about the mind, or about particular mental processes, is nothing but the effect within a man of certain stimuli, and the cause within a man of certain responses” (p. 79). Braddon-Mitchell and Jackson (1996) offer a succinct, yet detailed, statement of this sort of view of ordinary mental state concepts:

We distinguish the roles that matter for having a mind, and matter for being in one or another mental state, by drawing on what is common knowledge about mental states. We extract the crucial functional roles from the huge collection of what is pretty much common knowledge about pains, itches, beliefs, desires, intentions, and so on and so forth, focusing on what is most central to our conception of what a pain is, what a belief is, what it is to desire beer and all the rest. And we can group the common knowledge into the three clauses distinctive of the functionalist approach. The input clauses will contain sentences like “Bodily damage causes pain” and “Chairs in front of people in daylight cause perceptions as of chairs”; the output clauses will contain sentences like “Pain causes bodily movement that relieves the pain and minimizes damage” and “Desire for beer causes behavior that leads to beer consumption”; the internal clauses will contain sentences like “Perception as of beer in front of one typically causes belief in beer in front of one” and “Belief that if p then q typically causes belief that q on learning p ” (p. 46).

Philosophers who have introduced this conception of our ordinary understanding of mental states have sometimes used it to argue that the entities that occupy the functional roles specified by our common-sense understandings of mental states are brain states, and thereby defended an identity between minds and brains (as do Armstrong 1968 and Lewis 1966). As Fiala, Arico, and Nichols (2014) point out, “it is doubtful whether ordinary people are committed (even implicitly) to a theory like this” (p. 45, fn. 3). However, those who suggest a functionalist answer to the question of what features lead people to attribute phenomenally conscious mental states do not attribute this metaphysical theory to ordinary people. Instead, they use empirical methods to argue for two claims:

1. We have good evidence that people exploit information about the functional roles an entity can and does occupy in deciding what mental states (including phenomenally conscious states) it is (or is capable of) occupying, and
2. We have no good evidence that people exploit information that does not contribute to information about the functional roles an entity can and does occupy in deciding what mental states (including phenomenally conscious states) it is (or is capable of) occupying.¹¹

If these two claims are correct, they contend, it would be unsurprising if it turned out that people's attributions of mental states—phenomenal states included—were typically mediated by their beliefs about an entity's functional capacities and the functional roles it currently occupies and nothing else.

Direct evidence for the first claim has been provided by Huebner (2010), who had participants assess their agreement with various mental state attributions to four sorts of entities: A normal human with a Central Nervous System (CNS); a human with a CPU in place of a CNS, a robot with a CNS in place of a CPU, and a robot with a CPU. In a reasonable attempt to standardize the functional capacities of each of these entities (and so to isolate the role of constitution) Huebner (2010) described each as behaving “in every respect like a person on all psychological tests” (p. 138). He found that while people were somewhat inconsistent in their attributions of some paradigmatic phenomenal states to either the robot or the cyborgs, they strongly agreed with attributions of intentional (non-phenomenal) states to all four entities. This suggests that people exploit functional information in attributing intentional mental states, at

¹¹ Information that contributes to information about functional roles is here understood to include information about functional roles and any other information that implies information about functional roles. For example, information about the purpose for which some device was designed may imply information about the functional roles it might occupy.

least. Additional evidence that people exploit information about functional roles in deciding what mental states an agent is capable of occupying is provided by Buckwalter and Phelan (2014).

They investigated people's attributions of the emotional states of happiness and sadness to embodied and disembodied agents across five studies. While they found no effect for whether or not the agent was embodied in any of these studies, each study revealed a strong effect for how the entity was functionally described. In their third study, for example, participants were presented with a vignette describing either a forest-dwelling hermit or an entirely disembodied forest-dwelling spirit. The agent was then described as responding to encroaching developers by breaking their equipment. Finally, the developers were either turned away by the retaliation or else the agent was driven out of the forest by the developers. As (the left half of) Figure 3 shows, participants strongly agreed with attributions of feelings of sadness to the embodied hermit *and* the disembodied spirit in the sad situation of being forced from their home, but not in the happy situation of driving away the developers. Conceptually equivalent results were found for happiness attributions.

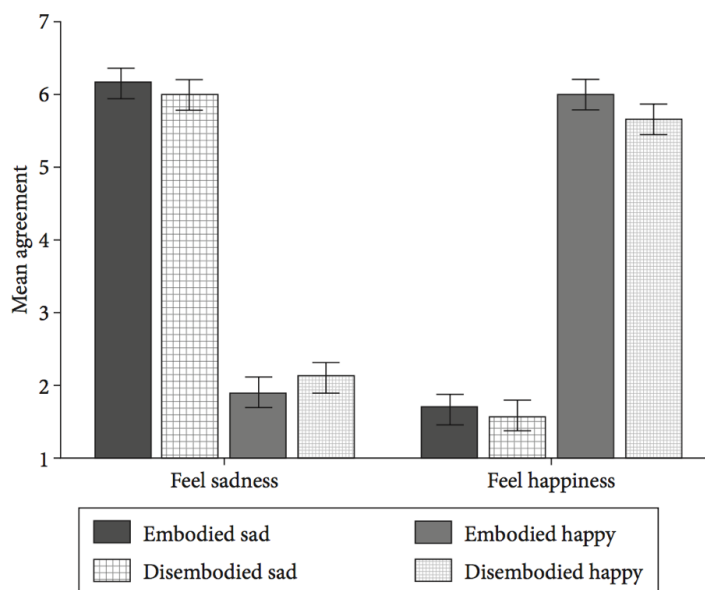


Figure 3: Assessments of attributions of feelings of sadness or happiness for an embodied or disembodied sophisticated entity in a happy or sad situation (from Buckwalter and Phelan 2014, pp. 54).

These results suggest that people exploit and, indeed, prioritize functional information in deciding what mental states—including phenomenally conscious mental states—an entity is occupying. But perhaps few would deny this claim. The functionalist's second claim, however, asserts something additional, specifically, that there isn't good evidence that people exploit information that does not contribute to information about functional roles in deciding what mental states an entity is capable of occupying. But how can the functionalist defend this claim in light of the studies above that seem to suggest that people exploit non-functional information about entities in deciding what mental states to attribute to them?

As should be clear from the previous discussion, functionalists have often answered this question by presenting empirical evidence of their own to support a functionalist interpretation of results initially interpreted as supporting an anti-functionalist account of mental state attributions. For example, in responding to Knobe and Prinz's (2008), Phelan, Arico, and Nichols (2013) found that participants were significantly more likely to choose 'they' paraphrases for group names in sentences that attributed a phenomenal *or* non-phenomenal mental state to a group, compared with sentences that attributed a non-mental state. They argued from these results that attributions of mental states to groups are typically understood distributively, as attributions of states to the members that make up the relevant group *qua* members of that group. Such mental state ascriptions are thought to be appropriate, they contend, insofar as the relevant input, output, and internal clauses associated with the relevant mental state are satisfied and insofar as the mental state is saliently associated with the role of being a member of the relevant group. If correct, Knobe and Prinz's (2008) findings are explicable functionally, along the following lines:

Participants found it appropriate to say that ‘Acme Corp. believes that its profit margin will soon increase’ or ‘intends to release a new product this January’ because they understood these as attributions to the employees that make up Acme Corp. (in their role as Acme Corp. employees) and they associated beliefs about profit margins and intentions to release products with the role of being a corporate employee. They also understood ‘Acme Corp. is getting depressed’ or ‘is feeling excruciating pain’ as attributions to the employees that make up Acme Corp. in their role as Acme Corp. employees, but they did not associate feelings of depression or excruciating pain with the role of being a corporate employee, and so they judged these attributions inappropriate.¹² Tellingly, Sytsma and Machery (2009) as well as Arico (2010) both found that the addition of a prepositional phrase further contextualizing a group phenomenal state attribution led participants to judge that attribution more natural. This is explicable on the current functionalist approach, since feelings of depression over a recent court ruling are more saliently associated with being a corporate employee than are simply free-floating feelings of depression.

Similarly, reconsider Sytsma and Machery’s (2010) simple robot studies, now that we have a more fully articulated functionalist view on the table. Note that whether or not an entity can satisfy a specific, functional input condition depends not only on considerations of what is impinging on it, but also on considerations of that entity’s functional capacities. Chairs in front of people in daylight are taken to cause perceptions as of chairs. But, of course, chairs in a forest devoid of anything but trees are not taken to cause perceptions as of chairs, even when the sun is shining. This is because trees are not generally regarded as the right sorts of entities to satisfy the functional input conditions for visual states and experiences. They lack the appropriate sensory

¹² If this functionalist explanation is correct, then we might expect participants to find attributions of phenomenal states to groups such as ISIS or the capital mob more appropriate, since phenomenal states are more saliently associated with being a member of such a group.

apparatus. Sytsma and Machery may have inadvertently manipulated people's assumptions about the sort of sensory apparatus their simple robot was equipped with by putting it in the context of what appeared to be a psychological study that required it to see colors but not feel pains. This interpretation is supported by Buckwalter and Phelan's (2013) finding that when you make assumptions about design more explicit, by giving the robot a job it was built to perform, attributions of experiential, sensory states associated with that job go up.

In addition to this experimental work, proponents of the functionalist perspective have also offered non-experimental arguments for functionalist interpretations of purportedly anti-functionalist results. Consider, for example, Gray et al.'s (2007) finding that God was deemed less susceptible to experiences than a host of other entities. Phelan and Buckwalter (2012) argue that this is likely the result of people's beliefs about the kinds of functional roles an all-powerful God might occupy, rather than their beliefs about God's physical constitution. After all, Gray and colleague's study describes God as "the creator of the universe and the ultimate source of knowledge, power, and love" (7). It is unsurprising for purely functional reasons that such a God was deemed less susceptible to hunger, fear, pain, pleasure, and rage compared with a bunch of mere mortals.¹³ Think of the many plausible input and internal clauses for 'fear' to which a mere mortal is susceptible that make no sense to apply to an all-powerful, eternal deity.

To these arguments against anti-functionalist interpretations of specific experimental results, we can add a general consideration: It is unclear how experimental manipulations to an entity's physical constitution impact people's understanding of the sorts of functional clauses to which that entity is subject. Clearly, *some* manipulations to physical constitution influence what functional clauses we think an entity capable of satisfying. We accept that chairs in front of

¹³ Malle (2019) points out that these were the highest loading experiential traits in Gray et al. (2007).

people in daylight cause perceptions as of chairs, but not if those people's eyes have been gouged out. We suppose that an eye (or some other sensory apparatus) is required for certain functional capacities related to visual states. So, those without eyes are incapable of satisfying this input clause and others like it, and, if satisfying such input clauses is requisite for having visual experiences, then those without eyes are incapable of having visual experiences on functional grounds. However, little is known about how ordinary people think specific aspects of physical constitution bear on functional capacities. Do people think fleshy hands are required to occupy the functional roles associated with haptic sensations? Do they think stomachs are required to occupy the functional roles associated with feeling hungry? We don't know the answer to questions such as these, so we don't know how manipulations of an entity's physical constitution influence ordinary people's understanding of that entity's capacity to fill various functional roles. Nor have experimental philosophers typically included probes to assess people's inferences about functional capacities.

An intriguing twist to these functional considerations is offered by Arico and colleagues (2011) as well as Fiala, Arico, and Nichols (2014), in the form of their model of mental state attribution, the Agency Model. Arico and colleagues (2011) subjected participants to a reaction time study in which they agreed or disagreed with various state attributions—including phenomenally conscious state attributions ('feels anger', 'feels happy', 'feels pain', 'feels pride')—to various entities (mammals, birds, insects, plants, artifacts, vehicles, inanimate natural objects such as mountains, and moving natural objects such as clouds). They found that significantly more people responded affirmatively to phenomenal state attributions to insects than to plants, vehicles, or natural moving objects. They also found that it took participants longer to deny phenomenal states to insects than to vehicles or natural moving objects. On

reflection, it may seem somewhat surprising that people were willing to attribute states such as feeling anger or pride to insects at all. However, Arico and colleagues (2011) use these results to argue for a dual-process model of mental state attribution.¹⁴ Fiala, Arico, and Nichols (2014) further support this dual process model through a wider survey of the literature on mental state attribution. According to these researchers' Agency Model, there is a high road and a low road to mental state attribution. When asked for their considered judgment about whether an entity has mental states (as in the previous vignette studies), people depart on the high road and end up producing a reasoned assessment based, in large part, on functional considerations. But, if forced to make a speeded judgment (as by Arico and colleagues 2011) or a judgment under cognitive load, people go down the low road and manifest a more fundamental tendency to attribute a whole host of mental states—including fairly complex emotional states—to any entity that triggers the 'agent' concept, where an agent is any entity that possesses one or more simple features, including eyes, particular motion trajectories, or contingent interaction. This view doesn't contradict the Functionalist view (or other of the views we have considered), but it does entail that those views have missed a significant part of the story by focusing on higher level processing rather than more automatic judgments about mental state attribution.

The Functionalist and Agentic views are alike in claiming that our attributions of phenomenal states to other entities are guided by the same sorts of considerations that guide other of our mental state attributions. The Functionalist emphasizes only functional considerations, whereas proponents of the agentic view suggests that simpler considerations come into play in more automatic, unreasoned responses. These views are also alike in not taking

¹⁴ Dual-process models posit two, overlapping processes for a particular domain of cognition, where one of the processes is automatic and fast, whereas the other is deliberate and slow. Such models have been offered to explain mental heuristics (Kahneman, 2011), moral reasoning (Greene, 2013), and other cognitive phenomena.

a stand on the question of whether ordinary people possess a concept of phenomenal consciousness. People may possess such a concept, but, if they do, it appears not to be reflected in their practice of mental state attribution.

2.4 Further Reading on the Experimental Philosophy of Conscious State Attribution

Sytsma (2014a) provides an excellent review of the literature on conscious state attribution up to 2014, which emphasizes different perspectives than the current review. A number of influential articles in the experimental philosophy of mind are also contained in Sytsma (2014b). Sytsma and Buckwalter (2016) includes a section on experimental philosophy of mind with several important chapters on different aspects of the literature. Wyrwa (2020) provides a careful analysis of most of the studies discussed in this section and suggests a number of important limitations of this research.

Huebner (2015) raises a compelling challenge directed at Buckwalter and Phelan (2013) and some other experimental philosophical work. He argues, inspired by Suppes (1962), that experimentalists must develop and defend “plausible models of the data that are collected, the scales that are used, and the statistical analyses that are carried out” (3274). He further contends that experimental philosophers have not typically done this and that they are therefore unable, quite generally, to justify the conclusions they draw from their own data (presumably, Huebner includes his own experimental work in this assessment). However, Huebner also concedes that, since experimental philosophers use the same “methodologies that have been successful in cognitive and social psychology” (p. 3287), the problem he raises would apply to much of psychology.

Malle (2019) revisits the approach to thinking about mental state attribution pioneered by Gray, Gray, and Wegner (2007). He points out a number of limitations of their original study and conducts his own study, addressing these limitations. Ultimately, he rejects Gray et al.'s two-dimensional (agentic, experiential) view of the folk conception of mind, in favor of a three-dimensional account, encompassing an Affective Dimension, a Moral and Mental Regulation Dimension, and a Reality Interaction Dimension. He also develops a 20-item measure for dimensions of mind perception, which will prove useful in future research.

3. Other Projects in the Experimental Philosophy of Mind

Other projects in the experimental philosophy of mind do not directly address either of our previous questions about phenomenal consciousness. One such project attempts to undercut a central discourse in consciousness studies altogether. As discussed previously, Chalmers (1995), among others, assumes that the experiential dimension of our mental lives is manifestly obvious. He goes on to argue that explaining this fact—that we have phenomenally conscious experiences—constitutes ‘the Hard Problem’ of consciousness, since phenomenal consciousness is not explicable via the sorts of functional accounts that are common to the sciences. In arguing for this position, he asks that his readers:

...note that even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience...there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?* A simple explanation of the functions leaves this question open (203).

Elsewhere, Chalmers (1996) and others have appealed to a version of the Modal Argument against physicalism to motivate the claim that functional explanations will leave questions about

phenomenal experience open.¹⁵ This argument relies on the intuition that philosophical zombies—beings who lack experiential states, but are physically and behaviorally just like us—are conceivable. It follows, these philosophers go on to argue, that philosophical zombies are possible, and, therefore, that experiential states are not identical to physical or functional states, since identity claims are necessarily true, if true at all. Thus, physicalism is false.

However, experimental philosophers have presented challenges for both the general intuition that a complete explanation of functions will leave open questions about experience *and* the Modal-Argument-supporting intuition that philosophical zombies are legitimately conceivable. For example, building on work from Fischer and Engelhardt (2016) and Fischer et al. (2021), Fischer and Sytsma (2021) argue that the source of (at least some) purported zombie intuitions is not of the appropriate cognitive sort to support the conclusion that philosophical zombies are legitimately conceivable. Specifically, Fischer and Sytsma argue that some of these intuitions are due to certain, shallow linguistic associations with the popular concept of a ‘zombie’, and not to actual, considered judgments about whether something lacking experiential states could really be physically and behaviorally identical to us. They report a main effect whereby people are significantly more likely to agree with attributions of conscious states to ‘duplicates’ who “are all dark inside,” but are physically and behaviorally just like us, as opposed to ‘zombies’ who “are all dark inside,” but are physically and behaviorally just like us (6). Similarly, Diaz (2021b) finds evidence that the general intuition that a complete explanation of functions will leave open questions about experience is not widespread. Further, he purports to demonstrate that when such intuitions do arise, they are often due to extraneous factors not related to consciousness, such as a general skepticism about whether science can formulate a

¹⁵ See Kripke (1980), 148-155.

complete explanation of functions. These studies undercut the Hard Problem's preeminent role in consciousness studies by providing evidence that many of the judgments supporting the problem are due to irrelevant factors and unconsidered assessments.

Other experimental projects investigate how conscious states relates to other states of philosophical interest. For example, Nahmias et al. (2020) found that people who responded yes to a dichotomous question about whether some humanoid robots could have conscious experiences were more likely to attribute free will and moral responsibility to those robots. This effect on will and responsibility attributions was mediated by attributions of emotional states, which led the authors to conclude that,

...the ability to feel emotions and have things actually matter to the individual is important in Free Will attributions, yet the ability to have conscious sensations (e.g., the ability to experience sounds or smells) specifically plays no significant role (76).

In another project, Diaz and Reuter (2021) investigated the role of normative considerations on attributions of emotional states. They found that an agent described as behaving fearfully in response to a friendly dog was assessed as less afraid than an indistinguishable agent described as behaving fearfully in response to a fearsome dog. And they found that an agent who thinks herself happy in a possessive relationship was assessed as less happy than an indistinguishable agent who thinks herself happy in a non-possessive relationship. These findings seem to support the hypothesis from moral psychology that there are normative fittingness conditions for emotional states.

Eschewing the topic of phenomenal consciousness altogether, some experimentalists have also investigated attributions of attitudinal states. Intentional states have been the most extensive target of such investigations.¹⁶ But recent work on belief attribution is also important.

¹⁶ See Cova (2016) for a useful review.

For example, a series of papers by John Turri and co-authors argue that the folk accept a voluntary view of belief, whereby people can come to believe some proposition simply because they want to believe it. Turri et al. (2018), for example, offers six studies exploring the dimensions of folk voluntarism. They find that people preferentially assent to attributions of voluntary “immediate” belief, even when they could attribute mere acceptance. The extent to which people are willing to make such attributions is influenced by factors such as the willfulness of the believing agent and the evidential domain to which the belief relates (i.e., inferential vs. perceptual belief). Rose et al. (2017) reports a robust cross-cultural finding that people prioritize sincere assertions of belief, over behavioral counter-evidence, when attributing belief.

Finally, we may ask how people conceive of the targets of their mental state attributions. Strohminger et al. (2017) summarizes research suggesting a bifurcated view of other minds. People seem to think some mental traits and attributes are more authentic than others, such that we can refer to a ‘true self’, the aspect of a person that is who they really are, hiding under the superficial self an individual presents to the world. This research also suggests that we have a bias in attributing mental states to the true self; we tend to think that the true self is morally good, such that even though a person may behave in immoral ways, deep down, they know their behavior is wrong and want to do better (for whatever we individually think it is to do better). But if we conceive of others as consisting of a superficial and a true self, then, for any belief, desire, or feeling we might attribute to another, we might also attribute that belief, desire, or feeling’s opposite to that other’s true self. And this, indeed, will complicate our picture of mental state attribution.

References

- Arico, Adam (2010). "Folk Psychology, Consciousness, and Context Effects." *Review of Philosophy and Psychology*, 1(3), 371–393.
- Arico, Adam, Brian Fiala, Robert Goldberg, and Shaun Nichols (2011). "The Folk Psychology of Consciousness." *Mind & Language*, 26, 327–352.
- Armstrong, David (1968). *A Materialist Theory of the Mind*. London: Routledge. (second edition, with new preface, 1993)
- Block, Ned (1978). "Troubles with Functionalism." In Wade 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.
- Block, Ned (1995). "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences*, 18 (2), 227–247.
- Borg, Emma, Tim Salomons, and Nathaniel Hansen (2019). "The Meaning of Pain Expressions and Pain Communication." In Simon van Rysewyk (ed.), *Meanings of Pain*. Springer, pp. 261–282.
- Braddon-Mitchell, David and Frank Jackson (1996). *Philosophy of Mind and Cognition*. Oxford: Blackwell.
- Buckwalter, Wesley and Mark Phelan (2013). "Function and Feeling Machines. A Defense of the Philosophical Conception of Subjective Experience." *Philosophical Studies*, 166(2), 349–361.
- Buckwalter, Wesley and Mark Phelan (2014). "Phenomenal Consciousness Disembodied." In Justin Sytsma (ed.), *Advances in Experimental Philosophy of Mind*, London: Bloomsbury.
- Cappelen, Herman (2012). *Philosophy without Intuitions*. Oxford: Oxford University Press.
- Carruthers, Peter (2013). "Introspection: Divided and Partly Eliminated." *Philosophy and Phenomenological Research* 80 (1):76–111 (2010)
- Carruthers, Peter (2013). *The Opacity of Mind. An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Chalmers, David (1995). "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies*, 2(3), 200–219.
- Climenhaga, Nevin (2018). "Intuitions are Used as Evidence in Philosophy." *Mind*, 127(505), 69–104.
- Cova, Florian (2016). "The Folk Concept of Intentional Action: Empirical Approaches." In Justin Sytsma and Wesley Buckwalter (eds.), *A Companion to Experimental Philosophy*. Hoboken, NJ: Wiley.

- Diaz, Rodrigo (2021a). "Emotions and the Body. Testing the Subtraction Argument." *Philosophical Psychology*. DOI: [10.1080/09515089.2021.1945023](https://doi.org/10.1080/09515089.2021.1945023)
- Diaz, Rodrigo (2021b). "Do People Think Consciousness Poses a Hard Problem?: Empirical Evidence on the Meta-Problem of Consciousness." *Journal of Consciousness Studies* 28:3-4, 55-75
- Diaz, Rodrigo and Kevin Reuter (2021). "Feeling the right way: Normative influences on people's use of emotion concepts." *Mind and Language* 36:3, 451-470.
- Fiala, Brian, Adam Arico, and Shaun Nichols (2014). "You, Robot." In Edouard Machery and Elizabeth O'Neill (eds.), *Current Controversies in Experimental Philosophy*, New York: Routledge.
- Fischer, Eugen and Justin Sytsma (2021). "Zombie intuitions." *Cognition* 215
- Fischer, Eugen and Paul E. Engelhardt (2016). "Intuitions' Linguistic Sources: Stereotypes, Intuitions, and Illusions." *Mind & Language* 31:1, 67-103.
- Fischer, Eugen, Paul E. Engelhardt, and Justin Sytsma, J. (2021) "Inappropriate stereotypical inferences? An adversarial collaboration in experimental ordinary language philosophy." *Synthese* 198, 10127–10168.
- Forstmann, Matthias and Pascal Burgmer (Forthcoming). "The Cartesian Folk Theater. People Conceptualize Consciousness as a Spatio-Temporally Localized Process in the Human Brain." *Journal of Experimental Psychology – General*.
- Gazzaniga, Michael (1995). "Consciousness and the Cerebral Hemispheres." In Michael Gazzaniga (ed.) *The Cognitive Neurosciences*. Boston: MIT Press.
- Gazzaniga, Michael (2000). "Cerebral Specialization and Inter-Hemispheric Communication." *Brain*, 123, 1293-1326.
- Gray, Heather, Kurt Gray, and Daniel Wegner (2007). "Dimensions of Mind Perception." *Science*, 315, 619.
- Greene, Joshua (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin Press.
- Huebner, Bryce (2010). "Commonsense Concepts of Phenomenal Consciousness. Does Anyone Care About Functional Zombies?" *Phenomenology and the Cognitive Sciences*, 9(1), 133–155.
- Huebner, Bryce, Michael Bruno, and Hagop Sarkissian (2009). "What Does the Nation of China Think About Phenomenal States?" *Review of Philosophy and Psychology*, 1(2), 225–243.
- Huebner, Bryce (2015). "What is a Philosophical Effect? Models of Data in Experimental Philosophy." *Philosophical Studies*, 172, 3273–3292.

- Jackson, Frank (1986). "What Mary Didn't Know." *The Journal of Philosophy*, 83(5), 291–295.
- Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux
- Kim, Hyo-eun, Nina Poth, Kevin Reuter, and Justin Sytsma (2016). "Where is Your Pain? A Cross-Cultural Comparison of the Concept of Pain in Americans and South Koreans." *Studia Philosophica Estonica*, 9(1), 136–169.
- Knobe, Joshua (2003). "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63(179), 190–194.
- Knobe, Joshua (2008). "Can a Robot, an Insect or God be Aware?" *Scientific American Mind*, 19, 68–71.
- Knobe, Joshua and Jesse Prinz (2008). "Intuitions About Consciousness. Experimental Studies." *Phenomenology and the Cognitive Sciences*, 7, 67–83.
- Kripke, Saul A. (1980). *Naming and necessity*. Cambridge, MA: Harvard.
- Lewis, David (1966). "An Argument for the Identity Theory." *Journal of Philosophy* 63, 17–25.
- Malle, Bertram (2019). "How Many Dimensions of Mind Perception Really are There?" In Ashok Goel, Colleen Seifert, and Christian Freska (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*. Montreal, QB: Cognitive Science Society, 2268–2274.
- Nado, Jenny (2016). "The Intuition Deniers." *Philosophical Studies* 173, 781–800.
- Nahmias, Eddy, Corry Allen, and Bradley Loveall (2020) "When Do Robots Have Free Will? Exploring the Relationships between (Attributions of) Consciousness and Free Will." In *Free Will, Causality, and Neuroscience*, eds. Feltz, B. et al. 57-80.
- Nagel, Thomas (1974). "What is it Like to be a Bat?" *The Philosophical Review*, 83, 435–450.
- Nisbett, Richard and Timothy Wilson (1977). "Telling More Than we can Know. Verbal Reports on Mental Processes." *Psychological Review*, 84(3), 231–259.
- Phelan, Mark and Wesley Buckwalter (2012). "Analytic Functionalism and Mental State Attribution." *Philosophical Topics* 40(2), 129–154.
- Phelan, Mark, Adam Arico, and Shaun Nichols (2013). "Thinking Things and Feeling Things. On an Alleged Discontinuity in the Folk Metaphysics of Mind." *Phenomenology and the Cognitive Sciences*, 12, 703–725.
- Putnam, Hillary (1963). "Brains and Behavior." In Ronald Butler (ed.), *Analytical Philosophy. Second Series*. Oxford: Blackwell.

- Reuter, Kevin (2011). "Distinguishing the Appearance from the Reality of Pain." *Journal of Consciousness Studies*, 18(9–10), 94–109.
- Reuter, Kevin, Dustin Phillips, and Justin Sytsma (2014). "Hallucinating Pain." In Justin Sytsma (ed.), *Advances in Experimental Philosophy of Mind*, London: Bloomsbury.
- Reuter, Kevin, Michael Sienhold, and Justin Sytsma (2019). "Putting Pain in its Proper Place." *Analysis* 79(1), 72–82.
- Robbins, Philip and Anthony Jack (2006). "The Phenomenal Stance." *Philosophical Studies* 127(1), 59–85.
- Rose, David, Edouard Machery, Stephen Stich, Mario Alai, Adriano Angelucci, Renatas Berniūnas, Emma E. Buchtel, Amita Chatterjee, Hyundeuk Cheon, In-Rae Cho, Daniel Cohnitz, Florian Cova, Vilius Dranseika, Ángeles Eraña Lagos, Laleh Ghadakpour, and Maurice Grinberg. (2017). "Behavioral Circumscription and the Folk Psychology of Belief: A Study in Ethno-Mentalizing." *Thought: A Journal of Philosophy* 6, 193–203.
- Salomons, Tim, Richard Harrison, Nat Hansen, James Stazicker, Astrid Grith Sorensen, Paula Thomas, and Emma Borg (2021). "Is Pain 'All in your Mind'? Examining the General Public's Views of Pain." *Review of Philosophical Psychology*.
- Strohming, Nina, Joshua Knobe, George Newman. (2017) "The True Self: A Psychological Concept Distinct From the Self." *Perspectives on Psychological Science*, 12(4), 551–560.
- Sytsma, Justin (2009). "Phenomenological Obviousness and the new Science of Consciousness." *Philosophy of Science*, 76(5), 958–969.
- Sytsma, Justin (2010). "Folk Psychology and Phenomenal Consciousness." *Philosophy Compass*, 5(8), 700–711.
- Sytsma, Justin (2012). "Revisiting the Valence Account." *Philosophical Topics* 40(2), 179–198.
- Sytsma, Justin (2014a). "Attributions of Consciousness." *WIREs Cognitive Science* 5, 635–648.
- Sytsma, Justin (ed.) (2014b). *Advances in Experimental Philosophy of Mind*. Oxford: Bloomsbury.
- Sytsma, Justin and Edouard Machery (2009). "How to Study Folk Intuitions About Consciousness." *Philosophical Psychology* 22(1), 21–35.
- Sytsma, Justin and Edouard Machery (2010). "Two Conceptions of Subjective Experience." *Philosophical Studies* 151(2), 299–327.
- Sytsma, Justin and Kevin Reuter (2017). "Experimental Philosophy of Pain." In Joshua Knobe, Edouard Machery, and Stephen Stich (Eds.) *Journal of Indian Council of Philosophical Research – Experimental Philosophy* 34(3), 611–628.

- Takahashi, Hideyuki, Midori Ban, and Minoru Asada (2016). “Semantic Differential Scale Method can Reveal Multi-Dimensional Aspects of Mind Perception.” *Frontiers in Psychology*, 7, 1717.
- Turri, John (2018). “Choosing and refusing: doxastic voluntarism and folk psychology.” *Philosophical Studies* 175(10), 2507-2537.
- Weisman, Kara, Carol Dweck, and Ellen Markman (2017). “Rethinking People’s Conceptions of Mental Life.” *Proceedings of the National Academy of Sciences of the United States of America* 114(43), 11374–11379.
- Williamson, Timothy (2007). *The Philosophy of Philosophy*. New York: Routledge.