ORIGINAL ARTICLE

Attribution functionalism

Mark Phelan

Correspondence

Mark Phelan, Department of Philosophy, Lawrence University, 711 East Boldt Way, Appleton, WI 54911

Email: mark.phelan@lawrence.edu

Funding information

This work was supported by an internal grant from Lawrence University.

Abstract: People rely on functional information when deciding which mental states to attribute to an entity. Some researchers claim that non-functional bodily cues also and independently shape ordinary attributions of phenomenally conscious mental states—such as seeing red or feeling pain. Across two studies I show that the key embodiment experiments systematically confound bodily cues with unbalanced inferences about function: their "embodied" conditions embed richer functional affordances than their controls. Once this error is identified, little compelling evidence remains against extreme *attribution functionalism*, the view that functional information alone proximally guides attributions of both non-phenomenal and phenomenal mental states.

KEYWORDS

philosophy of mind, theory of mind, mental state attribution, functionalism, embodiment

1. INTRODUCTION

We often think in functional terms about other persons, artifacts, and entities. We do so when we think about the role that some feature or behavior plays in a complex system or process (e.g., "the tiger is using its paw as a cup, to bring water to its mouth"). And we do so when we think about the proper function that some feature or behavior of an entity exists to produce; that is, when we think about the function that explains why the entity under consideration exhibits that feature or behavior (e.g., "that tiger's stripes function as camouflage").[1] But we cannot directly perceive an object's function. So, we must infer the functions we attribute to other entities from other things we can perceive about them, or from the situations they occupy, or from things we have previously learned about the kinds of entities they are, or from other sorts of evidence. I know that many animals use stripes as camouflage, and so I infer that my tabby cat's tawny stripes exist because they provided camouflage to his ancestors, if not to him directly. I see a kid jump out of the way of a bicyclist that is careening towards her and I infer that the purpose of her movement was avoidance, in part because I suppose she saw the bike and did not want to get hit. We infer functions from things we accept about an entity and its larger context, but it is unclear what specific suppositions constitute the base of inference to function attributions or how such inferences proceed. This is one topic that this paper explores.

Attributing functions and functional capacities to the entities we encounter often entails attributing certain causal relations involving psychological states to those entities. For example, inferring that a kid jumped out of the way to avoid a careening bicyclist entails attributing certain causal relations involving her psychological states, such as *she saw the bicyclist*, *she thought the bicyclist was going to collide with her*, and *she decided to jump onto the grass*. If people readily infer functional capacities for the entities they encounter and if such inferences entail the attribution of causally implicated psychological states, then, presumably, this is a primary route by which people attribute psychological states to others. People attribute those psychological states they take to be entailed by the functional capacities they attribute to relevant entities.[2] Call this position on ordinary psychological state attribution *attribution functionalism*. The extreme version of this view would claim that functional information is the only information relevant to folk psychological state attribution. Here, I argue that the extreme version of attribution functionalism has not been ruled out, despite some seeming evidence to the contrary from the experimental philosophy of mind. I argue that this purported evidence is plausibly explained by experimental participants' inferences to functional capacity, so that the studies in question do nothing more than reenforce the extreme Attribution Functionalist's claim that people simply attribute those psychological states they take to be entailed by the functional states and capacities of a given entity.

[1] These examples of functional thinking correspond to ahistorical (e.g., Cummins, 1975) and historical accounts of function (e.g., Millikan, 1989). Other accounts exist within the literature, these are used only for illustrative purposes.
[2] The inference may also go in the opposite direction. For example, one may presuppose that someone feels a certain way or thinks a certain thing and, on this basis, interpret their action as having a certain purpose.

2. EMPIRICAL EVIDENCE FOR AND AGAINST ATTRIBUTION FUNCTIONALISM

Many philosophers of mind accept something approaching the extreme version of attribution functionalism for the class of psychological states they delineate as intentional states, which are characteristically regarded as non-phenomenal—for example, willing, having beliefs, desires, and intentions, tokening concepts, and so on.[3] What is more, several studies in the experimental philosophy of mind literature emphasize the importance of functional considerations to the attribution of the class of psychological states that philosophers delineate as phenomenal states—for example, having the experience of seeing red, feeling pain, feeling happy, and so on.[4] These philosophical experiments were prefigured by studies from psychology that identify behavior as a major cue for mental state attribution.[5] However, some other studies suggest that the extreme version of attribution functionalism is incorrect for at least some phenomenally conscious mental states. According to the authors of these studies, some information other than and which does not contribute to functional information is relevant to folk phenomenal state attributions, even if it is not relevant to folk non-phenomenal, intentional state attributions. The debate over extreme attribution functionalism thus focuses on folk attributions of phenomenal states, which will also be my focus. In this section, I will review the relevant literature.

The last fifteen years of the experimental philosophy of mind show that people exploit functional information in attributing phenomenally conscious states to other entities. A partial survey of results is sufficient to establish this conclusion:

- Sytsma and Machery (2010) found that a simple robot, which lacks typical sense organs and a brain, is thought by participants as capable of seeing red and smelling a chemical when it discriminates a red box or a chemical-scented box from other boxes in response to instructions.

[3] This assumption follows classic considerations from analytic metaphysics of mind, where Ned Block's (1978) Blockhead thought experiment—among others—dissuaded many from accepting a functionalist account of phenomenal states, without shaking their faith that intentional states might admit of such a reductive explanation.
[4] Note that while I am relying on a dominant philosophical line that delineates phenomenal from intentional states, and while I will continue to refer to phenomenal and intentional states without qualification in what follows, nothing I will say here presupposes that this delineation is correct.
[5] See Johnson (2003) for a review.

- Sytsma (2010) found that a pair of conjoined twins who “grimaced and shouted ‘Ouch!’ after they “forcefully kicked a large rock” with their shared foot, were assessed by experimental participants as clearly sharing the same pain.
- Buckwalter and Phelan (2013) found that a simple robot is assessed as more capable of occupying a range of experiential states (e.g., smelling bananas, feeling guilt) when it is described as having been designed for a state-relevant functional purpose (e.g., to make smoothies, to be an emotional support to the elderly) or as functioning in state-relevant ways (e.g., discriminating bananas when exposed only to their smell, avoiding a target of guilt). A simple robot who lacks the relevant functional design or behaves in contra-functionally appropriate ways was assessed as less capable of these experiential states.
- Buckwalter and Phelan (2014) found that embodied humans and disembodied spirits were attributed emotional states (e.g., feeling sadness, feeling happiness) at the same high rates when described as behaving in ways indicative of certain functional capacities (e.g., behaving in a vindictive way towards an ex, protecting a home from encroachers), but not when they behaved in contra-functionally appropriate ways.
- Salomons et al. (2022) found that participants agreed with an attribution of pain to a hospital patient who “moans, grimaces, cries out, clutches his stomach and, if asked, says that he is in severe pain” even in the absence of identifiable tissue damage. Furthermore, they found that people disagreed with an attribution of pain to a figure skater who damaged her wrist but did not behave in ways associated with pain.

I will refer to studies like these collectively as *function studies*. Function studies demonstrate a connection between perceived function and the attribution of phenomenally conscious mental states. In all the function studies just mentioned, phenomenal states are attributed (or attributed at higher rates) when participants are provided with information that the entity behaves (or can behave) in ways that are taken to be causally related to those phenomenal states. In the last three experiments just mentioned, phenomenal states are further shown not to be attributed (or to be attributed at lower rates) when the sorts of things the entity is portrayed as doing are contrary to the sorts of actions saliently associated with those phenomenal states. Both sorts of function studies demonstrate that people exploit functional information in deciding what phenomenally conscious states to attribute to other entities. But this is unsurprising, since we have first-hand experience of attributing phenomenal states to other people and animals based on our observations of their behavior (e.g., she is shouting, he is crying) and our assumptions about their functional design (e.g., she is normally sighted, dogs are unhappy when they are hungry).

While it is clear, then, that functional information plays a role in phenomenal state attribution, some experimental philosophers suggest—contrary to extreme attribution functionalism—that exhibiting the appropriate functional profile is not a decisive condition for lay-attributions of the relevant phenomenal states. These theorists maintain that at least some phenomenally conscious mental states are attributed by lay people only to (or more enthusiastically to) entities that are embodied in a certain way. Specifically, these theorists maintain that lay people attribute phenomenal states more readily to entities that possess a unified, biological body than to entities that lack such a body, even if the two entities are assessed as functionally equivalent. It is a corollary of this position that, all else being equal, even if an entity is assessed as capable of the associated functional profile, attributions of phenomenal states will be withheld (or less enthusiastically endorsed) if the entity lacks a unified, biological body. Entities that might instantiate some relevant functional profile or other, but lack a unified, biological body, include disembodied spirits (which lack a body), robots (which lack a biological body), and groups of people or swarms of insects (which lack a unified body). In fact, several studies have shown that endorsement rates for phenomenal state attributions to such entities were significantly lower than endorsement rates for non-phenomenal, intentional state attributions to those entities, or than attributions of phenomenal states to biologically embodied, purportedly functionally equivalent entities. These studies include:

- Gray, Gray, and Wegner (2007) found that a robot and God were assessed as less capable of experiential states (e.g., feeling pain, experiencing pride) than a host of other, biologically embodied entities (e.g., a frog, an infant, an adult woman).
- Knobe and Prinz (2008) showed that attributions of phenomenal states (e.g., experiencing joy, feeling pain) to a fictional Acme Corporation were judged much "weirder" than attributions of non-phenomenal, intentional states (e.g., believing, intending) to that corporation.

- Huebner (2010) found that participants agreed with attributions of non-phenomenal, belief states to a human, a robot, a robot with a human brain, or a human with a cpu instead of a brain. However, these participants were significantly more likely to agree with attributions of pain and happiness to the human than to the robot or either cyborg. Even though all four agents were said in Huebner's vignettes "to behave in every respect like a person on all psychological tests" (p. 138).
- Sytsma (2012) found that while participants agreed with attributions of pain to an agent who burns her bioengineered, fleshy prosthetic hand, they somewhat disagreed with attributions of pain to an agent who burns her robotic prosthetic hand, despite the fact that, in either case, the agent is said to be "able to do all of the things with her…[prosthetic]…hand that she did with her old hand" (p. 190).
- McLaughlin and Rose (2018) showed that participants rejected attributions of feelings of pain, itches, warmth, and pressure to a lifelike robot neonate, even after they removed any participant from the sample who rejected the claim that Jane-R (the robot) and Jane (a human neonate) are indistinguishable in outward appearance and outward behavior.

I will refer to these studies and studies like them collectively as *embodiment studies*. In early embodiment studies, such as those by Gray et al. and Knobe and Prinz, phenomenal state attributions to non-biologically embodied entities are shown to get significantly lower endorsement rates than non-phenomenal, intentional state attributions to those same entities. In the more recent embodiment studies that will be my focus here, some phenomenal state (e.g., experiencing pain, feeling an emotion) is attributed (or attributed at a higher rate) to an entity described as being biologically embodied, whereas it is not attributed (or is attributed at a lower rate) to an entity described as lacking a unified, biologically body, even though participants are provided with information suggesting that the two entities function in the same ways. Each embodiment study has been taken to support the claim that the body of an entity factors into assessments of that entity's capacity for at least some phenomenal states, and that it does so independently of functional information.

These embodiment studies constitute the best evidence from the experimental philosophy of mind against an extreme version of attribution functionalism. They have been used to support various versions of what I will call the *embodiment thesis*, that considerations of an entity's body play a role in lay-attributions of phenomenal states, independent of functional considerations. For example, on the basis of their finding that God (and other non-biological entities, such as a robot) were assessed as less capable of 'experiential' states than were entities such as a little girl and the family dog, Gray et al. (2007) argued that we have "not one dimension of mind perception, but two," (p. 619). Knobe and Prinz (2008) went further, arguing their experiments show that if people classify a mental state ascription as one that requires phenomenal consciousness, they "will regard it as acceptable only if they feel that the mental state in question has the right sorts of physical constitution" (p. 75). Other authors of embodiment studies, such as Huebner (2010), do not explicitly argue for a version of the embodiment thesis, but their studies are taken as supporting the view by others. Knobe et al. (2012), for example, count these studies among a diverse range of data that suggest, "complex behavior is the main cue for attributions of nonphenomenal states whereas embodiment is the main cue for attributions of phenomenal consciousness" (p. 94).

But do the embodiment studies actually support the embodiment thesis? I will argue that they suffer from a common trait that undermines their ability to support embodiment. Embodiment theories of phenomenal state attribution all essentially claim that information about an entity's body factors into ordinary, phenomenal state attributions in a way that is independent of functional information. However, none of the relevant studies establish that their results arise independently of manipulations of functional information. We lack a systematic understanding of what sorts of manipulations lead people to infer information about the things an entity can do. In the absence of such information, those who would establish embodiment must control for participants' inferences about functional information, or else they must demonstrate that effects related to manipulations of body are not mediated by participants' inferences to function. Many of the embodiment studies attempt to control for inferences to functional information. However, as I will demonstrate, we have good reason to suppose that these attempts are unsuccessful. Proponents of embodiment have not ruled out an alternate explanation for the embodiment studies, that their results are due to subtle manipulations of participants' inferences to function, where information about functions is the sole, proximal base of information people use to infer the mental states of other entities.

3. INFERRING FUNCTIONAL INFORMATION

The extreme version of attribution functionalism offers something like the following story for folk mental state attribution: People think of the different traits and behaviors of other entities as satisfying various functional roles and making possible different functional capacities.[6] They associate various psychological states with these functional roles and capacities, and they attribute to the entities with which they interact those psychological states that they associate with the functional roles and capacities they think the entities instantiate. Functional information is the only proximal source of information relevant to the attribution of psychological states of either the non-phenomenal or phenomenal variety. However, other information may serve a distal role in mental state attribution, by cueing inferences to functional information. The stimuli in the embodiment studies cue inferences to functional information and it is the differences in inferred functional information between test condition, not differences in embodiment, that directly explain the results of these studies. I will offer an inductive argument from original studies in favor of the last component of this story, that uncontrolled inferences to functional information explain the results of the embodiment studies.

But first let us briefly consider the sorts of information that may cue inferences to function. The function studies suggest some cues. By depicting an entity as behaving in ways stereotypically associated with discriminating a particular color or avoiding a particular target, experimental philosophers seem to have been able to get participants to infer that the entity is functionally capable of discriminating colors or avoiding a target, since, in their studies, participants assented to mental state attributions more wholeheartedly in those conditions in which the entity behaved in more functionally appropriate ways, while the physical embodiment of the entity made no difference. Additionally, by describing an entity as being designed to achieve a certain phenomenal-state-involving purpose experimental philosophers were able to get participants to ascent more wholeheartedly to the relevant phenomenal state attributions. So, it seems that people infer functions from information about the complex behaviors an entity performs and from information about its design history. But what other sorts of cues lead people to infer function?

The psychological literature on how people represent function is instructive. This literature has developed with the primary aim of understanding object categorization. It has tended, therefore, to emphasize a unique, proper function for an object and to investigate the role this

[6] I am non-committal as to why people think in this way. Plausible explanations include general learning from experience or specification of an innate mechanism.

function plays in categorizations of the object. Following Gibson (1950, 1979) in understanding an affordance as a way an object can be used given its physical structure and the physical structure of the perceiving agent, some researchers have shown that young children prioritize functional affordances in object categorization (e.g., Kemler-Nelson et al., 2000; Oakes and Madole, 2008; Baumgartner and Oakes, 2011). Other researchers have shown that children emphasize a creator's original design intentions in determining an object's categorization-relevant function (e.g., Bloom, 1996; Gelman and Bloom, 2000). Other researchers, proponents of the 'HIPE' Theory, have promoted a maximalist view, according to which people's representations of an object's function are representations of a complex, relational structure comprising (potentially) facts about the object's history, its physical environment, and various of its behaviors and their outcomes (e.g., Chaigneau et al., 2004; Barsalou et al., 2005). This last view explicitly construes function as perspective relative, thus countenancing distinct functions for a single object, depending on the observer's point of view and meta-cognitive needs at the time. Additionally, Chaigneau et al. (2004) demonstrates how distant features, such as an object's design history and an agent's goal in using it, can lead to inferences about an object's function in the absence of more immediate features, such as details about its physical structure. According to the HIPE Theory, inferences to an object's function may arise from information about features as diverse as the object's life history, its environment, its creator's intentions, and the outcomes it produces during use.

There are, then, many potential cues for inferences to function. In the experiments that follow I will primarily be investigating the role played by an entity's history and its environment, since these are the most relevant functional cues for explaining the results of the embodiment studies that I will be investigating. I will argue that uncontrolled inferences to functional information offer the best explanation for the embodiment studies. While I cannot examine each of the many embodiment studies in detail, I will examine two classic embodiment studies to establish two claims, which together provide strong support for my conclusion. First, I will show, for one study (Huebner, 2010) how, once we control for different inferences to function across conditions, the results that purportedly support embodiment disappear. Then I will show how another study (Sytsma, 2012) also prompts unintended inferences to function relevant information, suggesting that the functional explanation generalizes to other embodiment studies.

4. UNCONTROLLED INFERENCES TO FUNCTION: EMPIRICAL EVIDENCE

Of the embodiment studies, most critical attention has been paid to Gray, Gray, and Wegner (2007) and Knobe and Prinz (2008).[7] However, Huebner (2010) and Sytsma (2012) also present influential studies that can be seen as supporting embodiment, but which have not faced the critical scrutiny of earlier studies. Huebner and Sytsma are also salient for our purposes because they attempt to control for inferences to functional differences across conditions in a way that deserves to be called *the orthodox method*, given its wide adoption in subsequent embodiment studies. My studies suggest that inferences to function are not adequately controlled across conditions using the orthodox method. Thus, uncontrolled inferences to functional information offer the best explanation for the embodiment studies, or so I will argue.

4.1 Study #1: Functional Differences in Huebner's Cyborg Vignettes

Huebner's studies illustrate how the orthodox way of controlling for functional information falls short and how, once we control for inferences to function, the purported embodiment effect disappears.

Huebner asked people about the psychological states of four types of entities: a normal human, a human whose brain has been replaced with a CPU, a normal robot with a CPU, and a robot whose CPU has been replaced by a brain. This is the vignette for a human whose brain has been replaced by a CPU:

> *CPU Replacement:* This is a picture of David. David looks like a human. However, he has taken part in an experiment over the past year in which his brain has been replaced, neuron for neuron, with microchips that behave exactly like neurons. He now has a CPU instead of a brain. He has continued to behave in every respect like a person on all psychological tests throughout this change.

The vignettes were paired with a picture of a human or a humanoid robot, depending on condition. Participants then reported their level of agreement with the following mental state attributions to the agent they read about:

[7] See Phelan (2023) for an overview of these two works and responses to them. There has been recent, better experimental work using the dimensions of mind perception paradigm introduced in Gray, Gray, and Wegner (2007). Sytsma & Snater (2023) provide a useful overview of this work specifically.

Study #1: He believes 2+2=4.

He feels pain if he is injured or damaged in some way.

Study #2: He believes that triangles have three sides.

He feels happy when he gets what he wants.

Participants recorded their responses on 5-point scales anchored at 1-strongly disagree and 5-strongly agree. Huebner performed analyses of variance (ANOVAs) on these results. He also converted participant responses into three-categories: disagree (comprising ‘strongly disagree’ and ‘disagree’ responses), agree (comprising ‘strongly agree’ and ‘agree’ responses), and ambivalent (comprising ‘neither agree nor disagree’ responses) and performed Chi-square tests on this nominal data. Each method of analysis uncovered significant effects across both studies. Univariate ANOVAs revealed significant differences between assessments of pain attributions ($p<0.001$) for Study 1 and happiness attributions ($p=0.005$) for Study 2 to Huebner’s four entities. Bonferroni post-hoc tests revealed significant differences between the normal human and every other entity in Study 1, such that participants agreed with attributions of pain upon injury to the normal human more than to any other entity. Bonferroni post-hoc tests also revealed significant differences between the normal human and both the robot with a CPU and the human with a CPU in Study 2. Participants agreed with attributions of happiness upon satisfaction of desires to the normal human more than to either other entity. However, no significant differences emerged for attributions of belief in either study. Instead, participants registered high levels of agreement with belief attributions to each entity in both studies.

It is easy to see why these studies have been taken to corroborate the embodiment hypothesis for phenomenal states. After all, participants were more likely to attribute pain and happiness to the normal human compared with the cybernetic human or the robot, both of whom lack an ordinary biological body. However, these findings constitute evidence against extreme attribution functionalism only if people really see Huebner’s entities as functionally equivalent. In his studies, Huebner attempts to ensure that this is the case by concluding each of his vignettes by specifying that the entity behaves “in every respect like a person on all psychological tests.” But this standardization of function presupposes a uniform and quite robust understanding of psychological testing on behalf of lay participants. Contra that presupposition, people might suppose it is possible to possess some functional capacities not captured by psychological testing. My first study shows this is not an adequate control for inferences to function.

4.1.1 Participants

Using G*Power, and assuming a medium effect size (d = 0.5) similar to those Huebner reported and power of 0.8, I calculated that at least 273 participants were needed. Because Huebner's original sample sizes had been quite small, I aimed to recruit at least 400 participants. In total, 422 participants (59% men, age M = 36.69, SD = 10.27) from Amazon's Mechanical Turk completed the study before it stopped receiving new HITs. Participants were paid $0.40 for participating. I preregistered to exclude participants who did not complete the study questions (though not the demographic probes), as well as the later participating of any duplicate IP addresses.[8] In total, 15 participants were excluded for these reasons, leaving a sample size of 407.

4.1.2 Procedure

Buckwalter and Phelan (2013) showed that participants were willing to attribute various olfactory experiences and emotional states to a simple robot, so long as it was described as performing behaviors stereotypically related to those phenomenal states or as being designed for certain phenomenal state involving purposes. So, I focused on the comparison between the normal human and the cyborg human who had his brain replaced with a CPU. Again, Huebner found that the cyborg was assessed as less capable of both pain and happiness, but not simple belief. Attempting to replicate this result, I gave participant's Huebner's descriptions of the two entities and asked them to assess their agreement on a five-point scale with the following claims:

1) He believes that 2+2=4.
2) He feels pain if he is injured or damaged in some way.
3) He feels happy when he gets what he wants.

Additionally, to assess whether lay people who are told that an entity behaves "like a person on all psychological tests" regard that entity as functionally equivalent to a normal person, as Huebner (2010) presupposes, I added a probe which asked participants if:

4) He behaves like a normal person in every way.

[8] Project details and data for this and the next study are available here: https://osf.io/kp6t9/.

This probe seems to capture what is at issue in the debate over embodiment, since the embodiment theorist wants to find evidence that information about bodies plays some role in phenomenal state attribution that is independent of the role played by functional information. Additionally, including this probe allows us to select those participants who *actually do* see the person whose brain has been replaced by a CPU as behaviorally equivalent to a normal person and to assess the degree to which biological embodiment plays a role in their phenomenal state attributions.[9]

I also added an additional probe to investigate subtle differences between Huebner's belief probes and his phenomenal state probes. On the one hand, believing that 2+2=4 and that triangles have three sides are quite basic intentional states. It would be trivial to determine if someone held these beliefs. And it is possible to hold these beliefs non-occurrently—that is, someone can be said to believe something even if they are not thinking that thought in the moment. On the other hand, feeling pain if one is injured or damaged in some way and feeling happy when one gets what one wants are more complicated states. It would require repeated observation in different sorts of situations to determine if someone was subject to these dispositions. And feeling pain and happiness are occurrent states—that is, someone is said to feel pain or happiness only if they are feeling those states right now. So, whereas believing 2+2=4 is a non-occurrent mental state, feeling happy when one gets what one wants is a disposition to an occurrent mental state. To find out if the effects Huebner uncovered are due to the relevant complexity of the phenomenal states he selected, rather than their phenomenal nature, I asked participants if:

5) He forms beliefs about items in his immediate vicinity when he looks around a room.

If responses to this probe patterned with responses to the happiness and pain probes, then those responses are presumably due to the relative complexity of dispositional states, not to the phenomenal nature of hapiness and pain. In that case the effects are good candidates for explanation in functional terms. More complex states may be thought to require more functionally complex systems for their instantiation and neuronal systems may be thought to be more functionally complex than computational ones.

[9] I am presupposing that lay participants have a concept of a 'normal person' and I take it that if someone is functionally equivalent to a normal person, then they will be regarded as behaving like a normal person in every way. (Though the converse is not true; even if an entity behaves like a normal person in every way, it is not necessarily functionally equivalent to a normal person, since it might instantiate all the relationships between sensory inputs and behavioral outputs that a normal person does via radically different interrelations between its mental states.)

Finally, I added a third condition to this study, involving a new vignette not originally investigated by Huebner. Phelan and Buckwalter (2012) discussed how there is a large difference in the personal history of normal human David and the David whose brain has been replaced. Cyborg David is and normal David is not described as having undergone a time intensive and invasive experimental procedure. Is it possible that, as the HIPE model would suggest, participants are inferring that cyborg David's functional capacities have been shaped—perhaps inhibited—by this procedure, and that this aspect of his personal history is why they judge him less capable of happiness and pain? To investigate this question, I asked a subset of participants about this third vignette:

> *Brain Replacement:* This is a picture of David. David looks like a human. However, he has taken part in an experiment over the past year in which his brain has been replaced, neuron for neuron, with biological, lab-grown neurons that behave exactly like his original neurons. He now has a fully regenerated brain. He has continued to behave in every respect like a person on all psychological tests throughout this change.

By including this third condition, we can isolate the distinction between biological embodiment and personal history. If the effects Huebner uncovered are traceable to biological embodiment, then responses to this new condition should pattern more closely with the normal human probe than with the CPU replacement probe. If they are traceable instead to David's personal history, then responses to this new condition should pattern more closely with the CPU replacement probe than the normal human probe.

To summarize, participants in this study were randomly presented with one of three conditions involving three sorts of entities: a normal human, a human whose brain has been replaced with a CPU, and a human whose brain has been replaced with a new brain. Participants assessed their agreement with the five probes described above on 5-point scales anchored at 1-strongly disagree and 5-strongly agree. The order of the five probes was randomized. The presentation order produced no meaningful results.

4.1.3 Results and Discussion

*

Analyses of all participants. I initially analyzed results for the 407 participants who completed all five probes, regardless of responses to the *behaves* probe. One-way ANOVAs revealed no effect for entity on the *believes*, *forms beliefs*, or *feels happy* probes. However, univariate ANOVAs revealed small significant effects for entity on the *pain* and *behaves* probes. Following Huebner, I conducted post-hoc Bonferroni tests, which revealed significant differences for normal human vs. human whose brain has been replaced by a CPU for *feels pain* ($p = 0.039$) and *behaves* ($p = 0.017$). No other significant results were found for Bonferroni tests and only two other results had a p-value less than 0.2. See Table 1 for details.

Probe	**Normal Human *M***	**Human Brain Replace *M***	**Human CPU Replace *M***	**F-score, η2**
Believes: He believes that 2+2=4.	4.82	4.77	4.79	F (2, 404)=0.38, p=0.684, ηp2=0.002
Forms beliefs: He forms beliefs about items in his immediate vicinity when he looks around a room.	4.1	4.05	4.22	F (2, 404)=1.6, p=0.204, ηp2=0.008
Feels happy: He feels happy when he gets what he wants.	4.35	4.23	4.16	F (2, 404)=2.24, p=0.108, ηp2=0.011
Feels pain: He feels pain if he is injured or damaged in some way.	4.24	* 4.1	3.96	F (2, 404)=3.11, p=0.046, ηp2=0.015
Behaves: He behaves like a normal person in every way.	4.35	* 4.05	4.08	F (2, 404)=4.22, p=0.015, ηp2=0.020

Table 1: Results of one-way ANOVAs including all participants for five State probes and three Entities, Study 1.

Like Huebner, I converted participant responses into the three-categories of disagree (no), agree (yes), and ambivalent (not sure) and performed Chi-square tests on this nominal data to compare the different entities. Participants were significantly more likely to agree with attributions of *feels happy*, $X^2(2, N=275) = 13.318$, $p = 0.001$, and *behaves*, $X^2(2, N=275) = 9.062$, $p = 0.011$, to the normal person compared with the person whose brain had been replaced by a CPU. They were also significantly more likely to agree with attributions of *feels happy*, $X^2(2, N=266) = 6.542$, $p = 0.038$, to the person whose brain had been replaced with other neurons than to the person whose brain had been replaced by a CPU.

These results reflect Huebner's original findings in important respects, providing solace to the proponent of embodiment. As with Huebner, an ANOVA failed to reveal any significant difference between ascriptions of belief to the various entities. Indeed, this was also the case with the *forms beliefs* probe; though, interestingly, there was a non-significant trend such that people agreed with attributions of belief formation *more* for the person whose brain had been replaced by a CPU than for any other entity. Also replicating Huebner, there was a significant difference for attributions of pain to the various entities and post-hoc comparisons showed that participants were significantly less likely to attribute pain to the person with a CPU than to the normal person. Chi-square tests on the categorical data also showed that people were significantly less likely to agree with attributions of happiness to the human with a CPU than to either the normal human or the human whose brain had been replaced with other neurons.[10]

However, these results also raise difficulties for embodiment. Lower rates of agreement with phenomenal state attributions to *both* surgical replacement patients may have more to do with personal histories of surgical intervention than with physical composition.[11] What is more, participants rated the capacity to form beliefs much lower for all three entities ($M = 4.12$, $SD = 0.772$) than the attribution of the simple belief that 2+2=4 ($M = 4.79$, $SD = 0.545$). A paired samples T-test revealed this difference to be significant, $t(406) = 15.905$, $p < 0.001$. Thus, contra the hypothesis that people are hesitant to attribute phenomenal states to non-biological entities, it

[10] An anonymous reviewer writes that the results thus far reported "fail to replicate Huebner's…large effects for pain and medium-sized effects for happiness" and concludes that "Amazon Mechanical Turk is broken beyond repair." But an alternative explanation for the smaller effect sizes I found for pain and happiness compared with Huebner, 2010, is that the earlier studies were underpowered. Across his two studies (on which my first study is based), Huebner had an average of 25.5 participants per cell. My study includes an average of 135.7 participants per cell. It is well known that small sample sizes can produce inflated effect sizes due to random variability. Because of regression to the mean, extreme effects are less common in larger samples. While further research is warranted, appropriate power is a reasonable explanation of the smaller, directional replications that I found.

[11] Marginal differences suggest participants agreed more with attributions of *feels happy*, $X^2(2, N=273) = 5.75$, $p = 0.056$, *feels pain,* $X^2(2, N=273) = 4.936$, $p = 0.085$, and *behaves,* $X^2(2, N=273) = 4.709$, $p = 0.095$, to the normal person than to the person whose brain had been replaced by other neurons.

appears people attributed the belief that 2+2=4 because of its utter triviality, whereas the dispositions to feel pain, happiness, or to form beliefs are complex dispositions.

Finally, the person whose brain has been replaced by a CPU and the person whose brain has been rebuilt with other neurons were regarded as behaviorally—and, therefore, functionally—distinct from a normal person. Is it, then, just inferences to functional differences from non-functional factors that explain Huebner's results? To find out, I repeated the above analyses on just those participants who agreed that the entity they assessed behaved like a normal person in every way.

Analyses of participants who assess the entity as behaving like a normal person: Eliminating participants who chose a response other than "agree" or "strongly agree" for the *Behaves* probe left 349 participants. Unsurprisingly, a proportionally greater number of participants agreed that the normal human behaved like a normal person (129/141) than agreed with that assessment for either the human whose brain had been rebuilt with microchips (110/132) or the human whose brain had been rebuilt with other neurons (110/134). One-way ANOVAs conducted on this population revealed no effect for entity on any of the remaining four probes *believes*, *forms beliefs*, *feels happy*, or *feels pain*. See Table 2 for results:

Probe	**Normal Human *M***	**Human Brain Replace *M***	**Human CPU Replace *M***	**F-score, $\eta 2$**
Believes: He believes that 2+2=4.	4.87	4.82	4.81	$F(2, 346)=0.735$, $p=0.480$, $\eta p2=0.004$
Forms beliefs: He forms beliefs about items in his immediate vicinity when he looks around a room.	4.14	4.18	4.29	$F(2, 346)=1.373$, $p=0.255$, $\eta p2=0.008$
Feels happy: He feels happy when he gets what he wants.	4.40	4.32	4.36	$F(2, 346)=0.464$, $p=0.629$, $\eta p2=0.003$
Feels pain: He feels pain if he is injured or damaged in some way.	4.32	4.18	4.07	$F(2, 346)=2.331$, $p=0.099$, $\eta p2=0.013$

Table 2: Results of one-way ANOVAs including all participants who indicated behavioral equivalence to a normal person for four State probes and three Entities, Study 1.

Following Huebner, I conducted post-hoc Bonferroni tests, which revealed no significant differences.

Again, I converted participant responses into categorical data and performed Chi-square tests for each of the states. No significant differences emerged between participants categorical responses for the person whose brain had been replaced by a CPU and the person whose brain had been rebuilt with other neurons. Participants were significantly more likely to agree with attributions of *feels happy* to the normal person compared with the person whose brain had been replaced by a CPU, $X^2(2, N=234) = 6.225$, $p = 0.044$.[12] However, this result conforms to the finding that participants were significantly more likely to agree with attributions of *feels happy* to the normal person compared with the person whose brain had been replaced by another brain, $X^2(2, N=239) = 4.27$, $p = 0.039$. But this result is inexplicable through embodiment. Together, the two results suggest that a personal history of significant surgical intervention may be an important driver of inferences to the functional capacity for happiness.

Once the population of people who regarded the brain-replacement patients as functionally equivalent to the normal human was isolated, all effects for the scalar data disappeared. Effects for the categorical data also disappeared, or else were not explicable according to embodiment.

4.2 Study #2: Functional Differences in Sytsma's Prosthetic Hand Cases

We have seen how inferences to function account for differences in mental state attributions in one influential embodiment study. Does such an account generalize to other embodiment studies? In this section, I argue that it does by demonstrating that another influential embodiment study (Sytsma, 2012) leads participants to infer unintended functional information.

Across a series of papers, Justin Sytsma and co-authors have argued for a Naïve Account of folk concepts for some phenomenal states. Sytsma (2012) claims that ordinary people take "colors to be mind-independent qualities of worldly objects" and pains "to be mind-independent qualities of certain body parts" (p. 188) and the bulk of subsequent work has focused on defending these two specific claims.[13] While certainly not a doctrinaire embodiment view—and Sytsma

[12] Though, for this result, chi-square's assumption that the expected frequency for each cell should be greater than five was violated.

[13] Arguments and experiments in support of the Naïve Account are found in Sytsma (2010); Sytsma (2012); Sytsma and Reuter (2017); Reuter (2011); Reuter, Phillips, and Sytsma (2014); Reuter, Sienhold, and Sytsma (2019); and Kim, Poth, Reuter, Sytsma (2016), among others.

himself would likely reject the label, since he rejects, on empirical grounds, the folk-reality of a distinction between phenomenal and non-phenomenal states—Sytsma's Naïve Account of pain-attributions does constitute a version of the embodiment hypothesis as I am construing it. After all, it maintains that information about the body factors into folk assessments of an entity's capacity for the phenomenal state of pain and that it does so independently of how functional information about the entity factors into such assessments.

In support of his Naïve Account, Sytsma has conducted numerous studies assessing attributions of pain to various sorts of entities, including normal humans, robots, people with various types of prosthetics, conjoined twins, and people with phantom limbs. Across these studies, Sytsma has often attempted to standardize participants' assessments of the functional capacities of the relevant entities in the orthodox way, by including a statement in his vignettes indicating behavioral equivalence. For example, in Sytsma (2012), participants received one of the following paired vignettes involving a prosthetic hand:

> Susan is a normal adult female except that she lost her right hand in an automobile accident when she was 21. She has recently had an experimental procedure done that replaced her missing hand with a [robotic/bioengineered] one. The new hand was [built in a lab out of metal-alloys, sophisticated pneumatics, and complex computer circuits/grown in a lab out of bone, muscle tissue, and skin]. It is [hard and metallic/soft and fleshy] and looks [different from/similar to] her old hand. [Nonetheless/What's more], Susan is able to do all of the things with her [robotic/bioengineered] hand that she did with her old one—she's even resumed playing the piano!

The last sentence is the instrument of standardization across the two vignettes. Sytsma includes it to isolate how the distinct material compositions of the hands factor into pain attributions apart from functional differences. Finding that participants agree more with attributions of pain to Susan when she grabs a hot pan with a fleshy hand than with a metallic one, he concludes that a certain kind of embodiment is crucial to lay attributions of pain. But this follows only if Sytsma's statement successfully controls for disparate inferences to function. The current study suggests that it does not. Despite the attempt at standardization, people still assess the bioengineered hand as a future technology that is more advanced than the robotic hand.

4.2.1 Participants:

Using G*Power, I calculated the necessary sample-size for the independent T-Tests I planned to perform. Assuming a medium effect size (d = 0.5) and power of 0.8, G*Power returned a suggested sample size of 102, or 51 participants per cell. In total, 133 participants (Gender: 57.9% men, Age: M = 40.89 SD = 10.1) from Amazon's Mechanical Turk completed the study. Recruitment was limited to Master Turkers for this study, and participants were paid $0.30 for participating. Once again, I planned to eliminate participants who did not complete the study questions (though not the demographic probes), as well as the later participating of any duplicate IP addresses. However, no participants met these conditions, so all were included in the analyses.

4.2.2 Procedure:

Participants were presented, in one of two conditions, with only the first paragraph of either Sytsma's robotic hand or his bioengineered hand vignette (as included above). Each participant was then asked to generate two sorts of responses:

- They were asked to indicate their best guess for the year during which Susan's experimental procedure is likely to take place on a scale running from the year 2000 to 2500, gradated in yearly increments.
- They were asked to rate how technologically advanced Susan's prosthetic hand was on a scale running from 1 to 10, where 1 represented "fairly advanced, contemporary technology (for example, an iPhone)" and 10 represented "very complex, possible future technology (for example, a teleportation machine)."

I predicted that, despite attempts at standardizing functional capacities in these vignettes, participants would infer that the bioengineered hand was more functionally complex, as revealed through significantly higher assessments of its technological complexity and significantly later estimates of the year in which the vignette occurred.

4.2.3 Results and Discussion:

A two-tailed independent samples T-test revealed a significant difference between participants' assessments of the year in which the vignette involving the bioengineered hand (M = 2108) and the vignette involving the robotic hand (M = 2065) occurred, $t(131)=2.607$, $p = 0.01$, $d = 0.45$. A two-tailed independent samples T-test also revealed a significant difference between participants' assessments of how technologically advanced the bioengineered hand was ($M = 8.37$), compared with the robotic hand ($M = 6.46$), $t(131)=5.814$, $p < 0.001$, $d = 1$. While the temporal setting probe returned only a medium effect, there was a very large effect for degree of technological advance.[14] On the assumption that technological advance is conceived of, at least in part, in terms of functional capacity, this suggests that the functional confounds we saw for Huebner's embodiment study arise for Sytsma as well. The robotic prosthetic hand was thought of as less advanced technology, likely available within the next several decades. The bioengineered prosthetic hand, on the other hand, was thought of as highly advanced technology available beyond the lifetime of the average study participant.[15]

4.3 General discussion

My thesis has been that strong attribution functionalism remains a plausible, general account of mental state attribution, because we have good reasons to conclude that the embodiment studies—which purport to show that information about an entity's body factors into phenomenal state attributions independently of functional information—do not adequately control for inferences to function across their conditions. The previous two studies are intended to supply our good reasons. Among other things, they show lay tendencies to:

- conclude that a cyborg does not behave like a normal person;
- assess capacities for pain and happiness relative to assessments of functional capacity;
- regard biological prosthetics capable of feeling pain as future technologies, more advanced than less-sensate robotic prosthetics.

And they show that people make function-relevant inferences such as these despite the inclusion of explicit statements within vignettes that are supposed to control for inferences to function

[14] Indeed, the two groups differed by a standard deviation—the difference between the average responses for these groups is equal to within-group variability between responses.

[15] An anonymous referee points out that another study in Sytsma (2012) asks participants to assess cases in which Susan's eyes have been replaced by prosthetics, either robotic or biological. No difference emerged in participants' assessments of Susan's ability to see red between these two vignettes. On the assumption that the year and advancement measures would show a comparable difference for bio/robotic eyes as they do for bio/robotic hands, the reviewer suggests that such functional differences may not explain the observed difference in pain assessments after all. This is an interesting suggestion that deserves further attention, but it is insufficient to undermine the present account. Even if the year and advancement measures showed a comparable difference for bio/robotic eyes, it could be that the state of seeing red is deemed less complex than the state of feeling pain, and thus achievable through less functionally complex, robotic systems.

To be sure, there are shortcomings with the current studies. For example, the use of the *behaves* probe as a measure of function-inferences in study one means that participants' responses to it and the *happiness*, *pain*, and *forms beliefs* probes had to be collected within-subjects, raising the possibility that mental state ratings influenced behavior ratings. This could not be avoided in the current paradigm, but in future research I hope to investigate lay assessments of the behavioral capacities of robots and cyborgs without including mental state probes, in order to develop more successful controls on inferences to function. Another potential shortcoming relates to my use of indirect measures to investigate participants' assessments of functional sophistication in study two. Participants regarded bioengineered prosthetic hands as coming about later and being more technologically advanced than robotic prosthetic hands, but this does not strictly entail that they thought bioengineered hands were more functional. Subsequent research should investigate how assessments of technological advance reflect inferences to function.[16]

Despite these shortcomings, the current research raises serious questions for the embodiment studies. Their vignettes engender function-relevant inferences, which the orthodox approach is ill-equipped to control. Yet, the orthodox approach continues to be favored by embodiment theorists.[17] So long as it does, their studies will provide little reason to reject extreme attribution functionalism.

5. Functional considerations and nothing else?

Much of the experimental philosophy of mind sets out to investigate "ordinary" practices of mental state attribution. But clearly people may be motivated in ordinary contexts to attribute or to refrain from attributing various mental states for any number of idiosyncratic reasons. It is said that Descartes, on the strength of his theory, concluded that the whimpers of the dogs he vivisected

[16] In a brief pilot study, I had 248 participants imagine they were buying a new phone when a salesclerk told them phone A was more technologically advanced than phone B. Asked what the salesclerk might have meant, 96 participants (38.7%) selected an explicitly functional paraphrase—A can do more things than B can—out of seven available options. (The most popular paraphrase—A is faster than B is—was selected by only 103 participants, and is of particular relevance for phones.)

[17] For example, McLaughlin and Rose (2018) support embodiment by demonstrating differences in phenomenal state attributions to a normal human and a robotic duplicate, assuming inferences to function are controlled through inclusion of the explicit statement that "Jane-R and Jane are…so similar in their behavior that one would be unable to tell by observing their behavior on any given day which is Jane-R and which is Jane." And Johnson and Knowles (2023) take similar behavioral descriptions as sufficient to establish functional equivalency, attributing differences in phenomenal state attribution to embodiment differences while ignoring the possibility that body structure may inform disparate inferences to functional capacities.

amounted to nothing more than the rasp of a tension spring wound too tight. Had everyone become Cartesians, then ordinary practice might have withheld phenomenal states from those organisms lower on the Great Chain of Being. (Arguably, it did in various places at various times.) An ordinary person under the sway of some neural identity theory might, similarly, reject pain attributions to non-biological entities lacking a nervous system. Strictly speaking, then, extreme attribution functionalism is false. Functional information is surely not the only information always relevant to phenomenal state attribution. But then, neither is functional information plus information about embodiment. I doubt that there is *an* ordinary practice of mental state attribution to be investigated at this level of analysis.[18]

It is clear, however, that we are not the only organisms that attribute phenomenal states to one another, as the case of empathy shows. There is compelling evidence that empathy—defined by de Waal and Preston (2017) as "emotional and mental sensitivity to another's state, from being affected by and sharing in this state to assessing the reasons for it and adopting the other's point of view" (p. 1)—is present in many species. More to our point, it seems that other primates, elephants, rodents, corvids, and canines engage in reconciliation and consolation behavior, a variety of cognitive empathetic behavior which requires imagining how the target feels. Chimpanzees, for example, reconcile after a fight with "embracing, gentle touching, and mouth-to-mouth kissing" (p. 89); and by-stander chimpanzees console the victim of an aggression through physical touch (de Waal & Aureli, 1996). Consolatory behaviors are exhibited in the absence of bystander distress, suggesting that such behaviors are not motivated by the bystander's desire to relieve their own negatively valanced state, but are responsive, instead, to the state of the victim. Furthermore, across mammal species and in humans, empathetic behavior depends on activation of similar brain regions, including the amygdala and cingulate cortices (Paradiso et al. 2021). When I console a friend, it is, in part, because I have assessed that they are sad, angry, or agitated. It seems plausible that something similar is going on in the mind of the chimpanzee when it engages in consolatory behavior. My capacity to attribute emotional and other phenomenal states

[18] Recently, more critical attention has been paid to the use of "ordinary" in experimental philosophy of mind. Borg et al. (2020), for example, argue against the assumption that the ordinary concept of *pain* must be either bodily or mental. They put forward a (kind of) multidimensional concept of pain with prominent bodily and mental dimensions. (See also Salomons et al. 2022.) Liu (2021), on the other hand, argues that ordinary uses of the word "pain" are context sensitive and have both a bodily and a mental sense. I have just been suggesting that uses of the word "pain" and its cognates are probably too numerous and varied to form a limited class, since they are often guided by idiosyncratic folk theories (and the same goes for other mental state words). However, as I will explain directly, there is an important domain of cognition for which extreme Attribution Functionalism may be true.

is realized—at least in those cases in which I am not under the sway of some explicit theory—by neural mechanisms that had their evolutionary origin in my and the chimpanzee's common ancestors. What base of evidence do these mechanisms (some of which are present in human infants as young as eight months-old)[19] use to attribute phenomenal states to others? It is more reasonable to ask after the domain of these mechanisms than to expect some domain of information that is relevant to every phenomenal state assessment humans make, and it is of these mechanisms that extreme attribution functionalism may be true.[20]

The chimpanzee, of course, attributes phenomenal states—or does whatever is its analogue of attributing phenomenal state—without the benefit of explicit theory. What it relies on, the studies on reconciliation, consolation, targeted helping, and other forms of cognitive empathy suggest, is the sort of information long recognized as central to the functionalist project—information about bodily impingements and behavioral outputs. How does a chimp know its conspecific is experiencing anxiety? Because he was just attacked by a dominate male. Because he is obsessively self-grooming. It is easy to think of ways in which the capacity to attribute phenomenal states to other entities would have been adaptively beneficial to our highly social, hominin ancestors. It is much more difficult to think of ways in which—in an environment devoid of robots, corporations, and disembodied spirits—it would have been adaptively beneficial in making these attributions to have consulted biological constitution. This does not disprove the claim that embodiment is a cue for attributions of phenomenal consciousness. But in the absence of positive evidence in favor of that position, it should be a nudge towards more extreme versions of attribution functionalism.

Acknowledgements: Ori Friedman provided excellent criticism of an earlier draft of this paper (in the process of rejecting it for *Cognition*). I am grateful for his feedback, which substantially improved the final product. Two anonymous reviewers for *Mind and Language* also provided helpful comments. Many other people made important suggestions along the way, whether they knew it or not.

[19] See Liddle et al. (2015).

[20] An attribution functionalist account of such pre-theoretic mechanisms accords with the findings of Arico et al. (2011), who uncovered a tendency for people to attribute a variety of mental states, including complex emotional states, to any entity that has eyes, particular motion trajectories, or contingent interactions when such judgments were speeded or made under cognitive load.

References

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

Barsalou, L. W., Sloman, S., & Chaigneau, S. (2005). The HIPE theory of function. In L. Carlson & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization, and development* (pp. 131–148). Oxford University Press.

Baumgartner, H. A., & Oakes, L. M. (2011). Infants' developing sensitivity to object function: Attention to features and feature correlations. *Journal of Cognition and Development*, 12(3), 275–298.

Block, N. (1978). Troubles with functionalism. In W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology* (pp. 261–326). University of Minnesota Press.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, 60, 1–29.

Borg, E., Harrison, R., Staziker, J., & Salomons, T. (2020). Is the folk concept of pain polyeidic? *Mind and Language*, 35(1), 29–47.

Buckwalter, W., & Phelan, M. (2013). Function and feeling machines: A defense of the philosophical conception of subjective experience. *Philosophical Studies*, 166(2), 349–361.

Buckwalter, W., & Phelan, M. (2014). Phenomenal consciousness disembodied. In J. Sytsma (Ed.), *Advances in experimental philosophy of mind* (pp. 45–74). Bloomsbury.

Chaigneau, S. E., Barsalou, L. W., & Sloman, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology: General*, 133, 601–625.

Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, 72(20), 741–765.

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, 18(8), 498–509.

de Waal, F. B. M., & Aureli, F. (1996). Consolation, reconciliation, and a possible cognitive difference between macaques and chimpanzees. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 80–110). Cambridge University Press.

Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition*, 76(2), 91–103.

Gibson, J. J. (1950). The perception of the visual world. Houghton Mifflin.

Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin.

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

Huebner, B. (2010). Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences*, 9(1), 133–155.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358, 549–559.

Johnson, G., & Knowles, A. (2023). Do you believe that aliens feel pain? An empirical investigation of mental state attributions. *Cognition, Brain, and Behavior: An Interdisciplinary Journal*, 27(2), 199–213.

Kemler Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition*, 77, 133–168.

Kim, H., Poth, N., Reuter, K., & Sytsma, J. (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, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, 7, 67–83.

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

Liddle, M. J., Bradley, B. S., & McGrath, A. (2015). Baby empathy: Infant distress and peer prosocial responses. *Infant Mental Health Journal*, 36(4), 446–458.

Liu, M. (2021). The polysemy view of pain. *Mind and Language*, 38(1), 198–217.

McLaughlin, B. P., & Rose, D. (2018). On the matter of robot minds. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy* (Vol. 2, pp. 270–312). Oxford University Press.

Millikan, R. (1989). Biosemantics. *The Journal of Philosophy*, 86(6), 281–297.

Oakes, L. M., & Madole, K. L. (2008). Function revisited: How infants construe functional features in their representation of objects. *Advances in Child Development and Behavior*, 36, 135–185.

Paradiso, E., Gazzola, V., & Keysers, C. (2021). Neural mechanisms necessary for empathy-related phenomena across species. *Current Opinion in Neurobiology*, 68, 107–115.

Phelan, M., & Buckwalter, W. (2012). Analytic functionalism and mental state attribution. *Philosophical Topics*, 40(2), 129–154.

Phelan, M. (2023). Experimental philosophy of mind: Conscious state attribution. In A. Bauer & S. Kornmesser (Eds.), *The Compact Compendium of Experimental Philosophy* (pp. 263–288). De Gruyter.

Reuter, K. (2011). Distinguishing the appearance from the reality of pain. *Journal of Consciousness Studies*, 18(9–10), 94–109.

Reuter, K., Phillips, D., & Sytsma, J. (2014). Hallucinating pain. In J. Sytsma (Ed.), *Advances in experimental philosophy of mind* (pp. 75–100). Bloomsbury.

Reuter, K., Sienhold, M., & Sytsma, J. (2019). Putting pain in its proper place. *Analysis*, 79(1), 72–82.

Salomons, T., Harrison, R., Hansen, N., Stazicker, J., Sorensen, A. G., Thomas, P., & Borg, E. (2022). Is pain 'all in your mind'? Examining the general public's views of pain. *Review of Philosophy and Psychology*, 13(3), 683–698.

Sytsma, J. (2010). Folk psychology and phenomenal consciousness. *Philosophy Compass*, 5(8), 700–711.

Sytsma, J. (2012). Revisiting the valence account. *Philosophical Topics*, 40(2), 179–198.

Sytsma, J. (2014). Attributions of consciousness. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(6), 635–648.

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

Sytsma, J., & Reuter, K. (2017). Experimental philosophy of pain. *Journal of Indian Council of Philosophical Research*, 34(3), 611–628.

Sytsma, J., & Snater, M. (2023). Consciousness, phenomenal consciousness, and free will. In P. Henne & S. Murray (Eds.) *Advances in experimental philosophy of action* (pp. 13–32). Bloomsbury.