

## **Commonsense concepts of phenomenal consciousness:**

### **Does anyone care about functional zombies?**

Bryce Huebner, Dept of Philosophy, UNC-Chapel Hill

Philosophers have often assumed the existence of clear standards for whether their views are supported by commonsense intuitions about consciousness. On the basis of this assumption, philosophical debates have often centered on the results of *a priori* reflection into the nature of consciousness. But, who would have thought otherwise? If armchair reflection is going to be useful anywhere, it will have to be useful in making sense of what it is like to sit in an armchair! Recently, however, some philosophers and cognitive scientists have explored the structure of commonsense ascriptions of consciousness to a variety of different systems (cf., Arico et al, forthcoming; Arico, 2007; Bruno, Huebner, & Sarkissian, 2007; Gray et al., 2007; Knobe & Prinz, forthcoming; Robbins & Jack, 2006; Systma & Machery, ms). In this paper, I develop a further contribution to this emerging range of experimental data. I argue that the philosophical views of ‘phenomenal consciousness’ that have become prominent in recent years are not consistent with the commonsense understanding of mental states.<sup>1</sup> Moreover, I argue that a theory that accounts for the subtleties in commonsense ascriptions of mental states is likely to offer important insights into the strategies used by people who have not studied philosophy or cognitive science in determining what sorts of systems are to be treated as subjects of moral concern.

Although my methodology is not without precedent, it is worth pausing to address a worry that is often raised about experimental philosophy. Some philosophers believe that we should not be concerned with the untutored intuitions of those who have not been trained in the techniques of philosophical analysis. This worry is especially pronounced when we turn to commonsense intuitions about consciousness; an adequate explanation of consciousness is a matter that will be decided, if it is decided at all, by the arguments of philosophers and the data collected by cognitive scientists.<sup>2</sup> There is no reason to suppose that people who have not studied philosophy have privileged access to the meaning of our concepts (cf., Kauppinen 2007), and there seems even less reason to think that untutored intuitions will tell us anything about the actual physiological constraints on the possession of a mental state. I agree that such worries would cast doubt on my use of the methodology of experimental philosophy were I to insist that commonsense intuitions will yield insight into the highly technical questions about the nature of the mind. However, my goal in this paper is not to explain what it

---

<sup>1</sup> The term ‘phenomenal consciousness’ brings with it a number of worries. In fact, I am inclined to think that the notion of ‘phenomenal consciousness’ is a philosophers’ construct that may have little to do with the taxonomy of mental states that we find in commonsense psychology. My thoughts about this topic will, however, develop throughout the body of this paper.

<sup>2</sup> As Peter Bukulich put the point in conversation, philosophers of mind probably wouldn’t be troubled by my results; they just aren’t concerned with what ordinary people think about the mind.

takes for something to have a mind; instead, my goal in this paper is to develop a better understanding of one of the perennial questions of philosophy (cf., Knobe 2007): What does it take for us to understand something as a locus of mentality?

Prior to entering the world of philosophy, we believe that some other organisms have a wide array of mental states. Moreover, whether we take something to be a mere ‘thing’, as opposed to a subject of moral concern, is deeply entwined with our understanding of the capability of that thing to be in various mental states. Note, however that developing a complete account of the commonsense ascription of mental states is a monumental task that will require extensive experimentation. In this paper I focus on just a few sorts of ascriptions: ascriptions of belief, ascriptions of the experience of pain, and ascriptions of the feeling of happiness. By the end of this paper, I will demonstrate how developing a more robust understanding of the conditions under which we ascribe these states yields an intriguing insight into what it takes for something to count as a subject of moral concern.

### **1. Three hypotheses about consciousness**

Let me begin with a brief articulation of the most prominent philosophical positions on the relationship between mental states and their physical realization. At one extreme, some philosophers have argued for a thoroughly functionalist theory of the mind, claiming that mental states are computational states that can be realized in any media that can realize those computations. At the other extreme, some philosophers have argued for neuronalism (cf., Cummins 1983), the claim that any state that is genuinely mental will be realized in an organism’s neural wetware. Finally, some people hold that some mental states, states of phenomenal consciousness, must be realized in an organism’s neural wetware, but non-phenomenal states are computational states that can be realized in any media that can realize those computations. Although these theories have not been developed as accounts of commonsense psychology, each suggests a hypothesis about the conditions under which commonsense psychology will ascribe various mental states. In this section, I take each of these positions in turn and develop three hypotheses, which will guide my experimental analysis of commonsense psychology.

#### *1.1 Functionalism and organizational invariance*

Some philosophers have defended a functionalist account of all mental states. As Marvin Minsky (quoted in Dennett 1988) once put the point, on this view of the mind, psychology turns out to be a lot like engineering. The reason for this is that in explaining mental states, the functionalist specifies a task, then appeal to the functioning of a hierarchically organized set of computational systems that can facilitate solving this task. All that matters, then, in deciding whether a system has mental states is whether it has the right computational architecture.

In defending a functionalist position, David Chalmers (1995, 1996) has argued that functional zombies (i.e., non-conscious systems that are functionally isomorphic to conscious systems) are *nomologically impossible*,<sup>3</sup> and that the functional and computational architecture of a system can adequately explain all mental states (at least so long as we are concerned only with the actual world). Applied to commonsense psychology, this theory yields a hypothesis according to which differences in physical structure will be irrelevant to the ascription of mental states, at least so long as psychological function and functional organization remain constant. Following Chalmers, I refer to this hypothesis as *organization invariance*.

### 1.2 Neuronalism and biological naturalism

Robert Cummins (1983, 90) argues that that the only plausible alternative to functionalism is to adopt 'neuronalism' "the doctrine that intentionally characterized capacities are realizable only as neurophysiological capacities". Whether or not this is true, philosophers who have been dissatisfied with the austere commitments of *organizational invariance* have taken up the gauntlet of neuronalism and have argued that mentality is possible only in a system with a biological brain. Perhaps the most vocal proponent of this view is John Searle (1992, 66), who argues that "behavior, functional role, and causal relations are irrelevant to the existence of conscious mental phenomena" (1992, 69). Searle has also argued that all mental states are *at least* potentially conscious because only first-person consciousness can account for the intensionality of thought. And, together, these two claims yield a sort of neuronalism.

Searle goes so far as to argue that there is, in fact a simple solution to the mind/body problem. Conscious states are caused by the activity of neurons in the brain, and are themselves nothing more than the higher-level features of a brain. Put bluntly, Searle holds that a mental state is nothing more than a particular state of a biological brain; so, according to Searle, unless something has a biological brain it will have no mental states. Applied to commonsense psychology, this theory yields a hypothesis according to which commonsense psychology will ascribe mental states only to systems with biological brains. Following Searle, I refer to this hypothesis as *biological naturalism*.

### 1.3 Phenomenal states and phenomenal distinctness

Although few philosophers are willing to go so far as Searle in adopting *biological naturalism*, many philosophers believe that there is a class of mental states that can only be realized in a biological brain. We can draw a distinction (cf., Chalmers 1996) between the psychological states that explain

behavior, such as beliefs, desires, and memories, and phenomenal states that it feels like something to experience (cf., Farrell 1950 and Nagel 1974). Philosophers typically call the latter states of phenomenal consciousness, and many philosophers claim that such states are not *organizationally invariant*.

Distinguishing between phenomenal and non-phenomenal states has not been an easy task. Philosophers typically begin by noting that a state is phenomenally conscious if it is experienced. However, an appeal to 'experience' merely shifts the explanatory burden to another equally obscure term. Ned Block (1995, 230), however, has conceded the impossibility of articulating a non-circular definition of phenomenal consciousness and has thereby adopted a strategy of offering paradigmatic examples. According to Block, a person is phenomenally conscious when she has a visual, auditory, or tactile experience, when she feels a pain, or when she experiences an emotion.<sup>4</sup> Block buttresses the claim that phenomenal states constitute a unique class of phenomena by appeal to differences between phenomenal and non-phenomenal mental states.

First, he claims that a theory that adopts the thesis of *organizational invariance* fails to explain *why* we experience the phenomenal states that we do; this is the difficulty familiar to most philosophers of mind as the "explanatory gap" (Chalmers 1996, Levine 1983, Lycan 1987, 1996, and Tye 1999). Second, Block argues *organizational invariance* fails to capture our intuitions about which systems are phenomenally conscious; functionalism rules out the possibility of functional zombies, but Block claims we can clearly imagine them. Finally, Block argues that although we have no compelling reason to think that experience will be exhaustively explained in organizationally invariant terms, we are positive that experience must be explicable in terms of a biological brain.

Block (1986, 2003), thus, argues that the only plausible explanation of the experiential properties of phenomenal states is in terms of their neural realization. Applied to commonsense psychology, this theory yields a hypothesis according to which there will be a distinction between phenomenal and non-phenomenal states such that differences in physical realization will be irrelevant for the ascription of non-phenomenal states but will be relevant to ascriptions of phenomenal states. I refer to this hypothesis as *phenomenal distinctness*.

The distinction between phenomenal and non-phenomenal states has garnered widespread philosophical agreement. However, it is not yet clear whether this distinction is present in commonsense psychology. Perhaps commonsense psychology is committed to *phenomenal distinctness*; perhaps it is committed to *organizational invariance* or *biological naturalism*. Of

<sup>3</sup> This characterization, of course, plays fast and loose with Chalmers' position. After all, Chalmers argues that there *is* a conceptual distinction between access and phenomenal consciousness; however, he claims that this distinction has no nomological force. My assumption is that people who have not been trained in academic philosophy will not make their judgments on the basis of a two-dimensional modal semantics but will consider only facts about the actual world.

<sup>4</sup> David Chalmers (1996) argues that an adequate account of phenomenal consciousness is one of the most pressing problems in the philosophy of mind. For Chalmers, the phenomenal character of a state is best characterized in terms of how it feels. Though there are significant differences between the theories developed by Block and Chalmers, the important point for my purposes is that Chalmers (1996, 6-11) too lists paradigmatic cases of phenomenal consciousness; his list includes sensory experience, pain, emotion, and mental imagery.

course, it would be surprising if physical organization played *no role* in the commonsense ascription of mental states; however, in the absence of empirical inquiry into the distinctions made by commonsense psychology, it is unclear what role physical constitution plays in intuitions about consciousness. Fortunately, however, there is a recent trend in experimental philosophy and cognitive science that has begun to examine the conditions under which commonsense psychology ascribes various sorts of mental states. I now turn to the relationship between a system's biological structure and its capacity to be in various mental states.

## 2. What can commonsense tell us about consciousness?

One place to begin examining the presence of a distinct range of phenomenal states is suggested by the initial distinction that I drew between psychological states that are useful in the explanation of behavior and states that it feels like something to experience. On the hypothesis that commonsense does draw such a distinction, Heather Gray, Kurt Gray and Daniel Wegner (2007) conducted a massive web-based survey designed to probe people's intuitive judgments about the sorts of systems that could be in various sorts of mental states.<sup>5</sup> Each of the volunteers made comparative judgments about a variety of different systems for a single sort of mental state, determined by the version of the survey that they were taking.<sup>6</sup> For example, people who took the 'fear survey' were asked to make a series of contrastive judgments about which character was more capable of feeling afraid; people who took the memory survey were asked to judge which character was more capable of remembering things; and people who took the morality survey were asked to judge which character was more capable of telling right from wrong and trying to do the right thing.

In analyzing their volunteers' contrastive judgments, Gray, Gray and Wegner (2007) found that human beings were consistently rated highly along each of two different axes: experience (in which they included the capacity for hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy) and agency (in which they included self-control, morality, memory, emotion recognition, planning, communication, and thought). In comparison, they found that volunteers ranked robots low on experience, but thought that a robot could have some capacity for agency.

These results are suggestive; they indicate that commonsense psychology does seem to distinguish between phenomenal and non-phenomenal states on the basis of physical organization. However, Gray, Gray, and Wegner

(2007) distinguish between experience and agency in a way that provides too course-grained a measure to test commonsense intuitions about the philosophical understanding of phenomenal consciousness. Although the use of a wide variety of cases (ranging from dead bodies, to frogs, to adult human beings) and a wide variety of mental states makes for interesting psychological data, it also introduces a number of confounding variables that make it hard to see how *this distinction* either tracks, or fails to track the philosophical distinction between phenomenal and non-phenomenal states. However, as a coarse-grained measure of these two sorts of states, this experiment offers a nice entry point for the analysis of commonsense ascriptions of phenomenal states.

This experiment only tests the broad hypothesis that there is a distinction between states that are primarily experiential states and states that explain behavior; however, by paying attention to the structure of this experiment, we are led to the bold suggestion that by *systematically* varying the sort of cognitive systems to which particular mental states are ascribed, it is possible to test a variety of hypotheses about the features of a system that play a significant role in treating a system as a locus of experience. Moreover, by studying the nature of these features, new explanations of why this distinction is present in human psychology may emerge. On the basis of considerations such as this, I believe that experimental philosophers are well poised to analyze commonsense understandings of 'phenomenal consciousness'.

A recent experiment by Joshua Knobe and Jesse Prinz demonstrates how this methodology of testing commonsense ascriptions of consciousness might be developed in a more precise way. Knobe and Prinz start from two hypotheses: 1) people who have neither studied academic philosophy nor cognitive science have a concept of phenomenal consciousness and they apply this concept in ascribing mental states to different sorts of cognitive systems; and 2) the ascription of phenomenal states relies on information about "physical realizers in a way that other mental state ascriptions do not" (Knobe and Prinz forthcoming, 5-6). On the basis of these two hypotheses, Knobe and Prinz developed an experiment in which they provided volunteers with a set of ten ascriptions of phenomenal and non-phenomenal states to the ACME Corporation. Volunteers were asked to rate the acceptability of each of ascription on a 7-point scale anchored at "sounds weird" and "sounds natural". In analyzing the responses of their volunteers, Knobe and Prinz found that although their volunteers thought that the ascription of non-phenomenal states to groups sounded natural, they thought that ascriptions of phenomenal states to the same group sounded much weirder. Knobe and Prinz (forthcoming, 17) argue that because "a group is no less capable of having a state with the functional role of depression than it is of having a state with the functional role of intention...the data are best explained by positing a distinct process that is not just a matter of checking functional roles." Unfortunately Knobe and Prinz's arguments are insufficient to demonstrate the truth of this claim.

<sup>5</sup> The characters included seven living human forms (a 7-week-old fetus, a 5-month-old infant, a 5-year-old girl, an adult woman, an adult man, a man in a persistent vegetative state, and the respondent him- or herself), three nonhuman animals (a frog, a dog, and a chimpanzee), a dead woman, God, and a sociable robot (Kismet). (Gray, Gray & Wegner, 2007)

<sup>6</sup> The surveys can be found at: [http://www.wjh.harvard.edu/~wegner/pdfs/Gray%20et%20al.%20\(2007\)%20supportive%20online%20material.pdf](http://www.wjh.harvard.edu/~wegner/pdfs/Gray%20et%20al.%20(2007)%20supportive%20online%20material.pdf)

As Justin Sytsma and Edouard Machery (MS) convincingly argue,<sup>7</sup> Knobe and Prinz’s (forthcoming) experiment fails to control for both differences in functional architecture and behavior between individuals and corporations. The same criticism holds of Gray, Gray, and Wegner (2007). While these experiments clearly demonstrate that there are differences in the sorts of mental states that commonsense psychology is willing to ascribe to different sorts of systems, conclusively demonstrating that people have a concept of phenomenal consciousness that is not beholden to functional organization requires the elimination of the confounding variables that suggest alternative accounts of the differences between the sorts of mental states that individuals and corporations can be in. As Sytsma and Machery (MS, 3) argue, in order to demonstrate the existence of a commonsense concept of phenomenal consciousness, “one needs to establish that ordinary people are disposed to ascribe different mental states to agents that are described as behaviorally and functionally equivalent.” However, on this point, the relevant empirical inquiry had yet to be developed; so with these concerns in mind, I developed a simple experiment designed to test the whether commonsense psychology was committed to *organization invariance*, *biological naturalism*, or *phenomenal distinctness*.

### 3. Robot beliefs and cyborg pains: Experiment 1

I recruited ninety-five volunteers from introductory philosophy classes at the University of North Carolina – Chapel Hill to participate in an experiment designed to determine what sorts of considerations underlie commonsense ascriptions of mental states. The demographic makeup of this group of volunteers was consistent with the demographics of the student population at UNC – Chapel Hill more broadly.

Volunteers were randomly assigned to four conditions: a Human-Human brain condition, a Human-CPU condition, a Robot-Human brain condition, or a Robot-CPU condition. Each volunteer was provided with a questionnaire that included a picture, a brief scenario, and two statements about the cognitive system pictured and described in the scenario. In the Human-Human brain and the Human-CPU conditions, volunteers were provided with a picture of a male human being and one of the following scenarios:

1. This is a picture of David. David looks like a human and he has a normal human brain. He behaves in every respect like a person on all psychological tests.

2. 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.

In the Robot-Human brain condition and the Robot-CPU condition, volunteers were provided with a picture of Kismet, the expressive robot created by the Humanoid Robotics Group at MIT and one of the following short scenarios:

3. This is a picture of David. David looks like a robot. However, he has taken part in an experiment over the past year in which his CPU has been replaced, microchip for microchip, with neurons. He now has a normal human brain. He has continued to behave in every respect like a person on all psychological tests throughout this change.
4. This is a picture of David. David looks like a robot. Instead of a human brain he is controlled by a CPU modeled on a human brain with microchips that behave exactly like neurons. He behaves in every respect like a person on all psychological tests.

In all four conditions, volunteers were asked to rate their agreement with two statements about David on a 5-point Likert scale anchored at ‘1-strongly disagree’ and ‘5-strongly agree’: 1) “He believes that  $2+2=4$ ”, and 2) “He feels pain if he is injured or damaged in some way”. The mean score for each of these judgments were as follows:

	Human-Human Brain	Human-CPU	Robot-Human Brain	Robot-CPU
Belief	4.08	3.72	3.78	3.60
Pain	4.16	3.24	2.91	2.59

In examining these results, a difference between volunteers’ ascriptions of belief and the feeling of pain was immediately apparent. In order to understand the considerations that were producing this difference, I analyzed the differences between volunteers’ ascriptions of mental states to systems with human bodies as opposed to robot bodies, and, I analyzed the differences between volunteers’ ascriptions of mental states to systems with human brains as opposed to systems with CPUs.

Volunteers showed no significant difference in their willingness to ascribe the belief  $2+2=4$  to any of the systems. More precisely, volunteers did not exhibit a significant difference in their ascriptions of belief to systems with human bodies as opposed to robot bodies,<sup>8</sup> nor in their ascriptions of

<sup>7</sup> As I began writing this paper, I was excited to find out that Sytsma and Machery were independently working on the relationship between functional organization and consciousness. On the basis of my belief that commonsense intuitions about consciousness should be tested on the basis of functionally and psychologically homologous systems, I had already both developed the experimental hypotheses that provide the foundation for this paper and ran my experiments. However, Sytsma and Machery nicely articulate these objections to Knobe and Prinz in a way that I would have been unlikely to develop had I not read their insightful contribution to these debates.

<sup>8</sup>  $F(1, 91) = .55, p = .46$

beliefs to systems with human brains as opposed to CPUs.<sup>9</sup> However, in stark contrast to the ascription of non-phenomenal states, volunteers *did* exhibit significant differences in both their ascriptions of the feeling of pain to a system with a human body as opposed to a robot body,<sup>10</sup> as well as a system with a human brain as opposed to CPUs.<sup>11</sup>

	Belief	Pain		Belief	Pain
Human body	3.90	3.70	Human brain	3.93	3.54
Robot Body	3.69	2.75	CPU	3.65	2.92

These results suggests that neither the body nor the ‘brain’ of a system are relevant to the ascription of belief, but both are relevant to the ascription of the feeling of pain.

At this point, it seems as though there is some rather promising evidence for a commonsense distinction between phenomenal and non-phenomenal states. Although volunteers were willing to ascribe non-phenomenal states like belief on the basis of function alone, they seemed to be committed to a particular realization of phenomenal states that did not appear to be organizationally invariant, thereby lending credence to the hypothesis of *phenomenal distinctness*. However, there is an important test case for the hypothesis of *organizational invariance* as it has typically been advance in philosophy. If the version of *phenomenal distinctness* that is typically defended by philosophers is correct, then the system with a human body and a CPU should be seen as a locus of belief, but not a locus of pain. However, in analyzing this case it became clear that there was *no significant difference* in volunteers’ willingness to ascribe belief as opposed to the feeling of pain. In fact, there is a significant correlation between volunteers’ willingness to ascribe belief and to the feeling of pain to this system.<sup>12</sup>

### 3.1 Explaining the commonsense understanding of pain

In making sense of these results, the first thing to notice is that commonsense psychology is not committed to *biological naturalism*. If it were, volunteers would only be willing to ascribe non-phenomenal states to systems with biological components; however, volunteers exhibited a thoroughgoing commitment to functionalism about belief such that there was no significant difference between an ordinary human being and an ordinary robot as regards the capacity to believe that  $2+2=4$ . Although this is a striking fact, it still leaves both *phenomenal distinctness* and *organizational invariance* unscathed.

Volunteers’ responses initially appear to support a version of *phenomenal distinctness*, which is consistent with the data reported by

Knobe and Prinz (forthcoming) and Gray, Gray, and Wegner (2007). The fact that both having a human brain and having a human body have a significant impact on volunteers’ ascriptions of phenomenal states suggests that people think that physical structures *are* important for determining whether a system can be a locus of phenomenal states, even though these factors are not important for determining whether a system can be a locus of intentional states. And, if this were the end of the story, there would be good reason to think, with Knobe and Prinz, that commonsense psychology does have a concept of phenomenal consciousness roughly analogous to the philosopher’s concept, and that this concept guaranteed a role for physical realizers in the ascription of phenomenal states.

However, although volunteers did exhibit a difference in their ascriptions of beliefs and pain to humans with CPUs, this difference was not significant. This fact, by itself, provides reason to doubt the standard philosophical version of *phenomenal distinctness*; however, it also suggests an explanation of the fact that both human brain and human body initially seemed to play such an important role in the ascription of phenomenal states. Briefly, the fact that there is no significant difference between the ascription of belief and the ascription of pain to a system with a human body and a CPU suggests that the initial plausibility of *phenomenal distinctness* is an artifact of people’s judgments about the purely robotic case. Although volunteers were fairly willing to ascribe the capacity to feel pain to system with a human body and a CPU, they were so unwilling to ascribe pain to a purely robotic system (i.e., the Robot-CPU combination), and so willing to ascribe pain to a normal human, that the difference between systems with brains and systems with CPUs was significant.

In analyzing volunteers’ response in the human-CPU condition, it becomes clear that the sort of body that a system has is doing most of the heavy lifting in modulating volunteers’ judgments about the capacity of a system to feel pain. It seems that according to commonsense psychology, a system will only be seen as a locus of pain if it has something like a human body. In a sense, this should not be too surprising. I argue that the most plausible interpretation of this data is that commonsense psychology is committed to a sort of naïve realism about pain: a headache is in the head, hunger pains are in the stomach, and aches are in the muscles. Put bluntly, pain is not in the brain; it’s in the body. This view has also been advanced by Sytsma (personal communication); and this result is broadly consistent with the results of a study recently conducted by Sytsma and Machery. In this study, volunteers were asked to justify their claim that a simple robot could not experience pain. According to Sytsma (personal communication) volunteers “frequently explained their answer that the simple robot (Jimmy) didn’t feel pain by saying that he was metal, mechanical, non-biological, and so on”.

However, if commonsense psychology takes the capacity to feel pain to be determined by the sort of body that a system has, then the possibility of explaining the feeling of pain in functional and computational terms is not

<sup>9</sup>  $F(1, 91) = .92, p = .34$

<sup>10</sup>  $F(1, 91) = 11.32, p = .001$

<sup>11</sup>  $F(1, 91) = 4.86, p = .03$

<sup>12</sup>  $t(24) = 1.69, p = .103, \text{ and } r = .54, p = .006.$

completely ruled out. In order to have the capacity to be in pain, a system must have soft, biological tissue that can be damaged, and it must have a computational system capable of representing damage to that tissue. So, while it might be true that a system that is hard and metallic won't feel pain, this is because it doesn't have the sort of bodily tissue that can be represented as damaged. Provided that the functional and computational organization of a mental system allows it to monitor damage to its own biological tissue, however, such a system can be the subject of a state of pain. This, then, provides good reason to doubt the intuitive plausibility of standard philosophical accounts of *phenomenal distinctness*.

However, this data does demonstrate that commonsense psychology puts some restrictions on the ascription of pain, so *organization invariance* fails. A system has to have a biological body in order to feel pain even though a human body is not necessary for having beliefs. As opposed to the claim of *phenomenal distinctness*, we might call this position *phenomenal embodiment*: the claim that *some of* the occupants of the functional roles required for pain can only be realized in soft, biological tissue. Although this hypothesis warrants further empirical inquiry, it might be that only soft, biological tissue can be damaged in a way that would be tracked as pain by commonsense psychology.<sup>13</sup> If this is true, there will be biological constraints on counting as a locus of pain; however, these constraints are not the ones traditionally defended by proponents of *phenomenal distinctness*.

#### 4. The emotional life of robots and cyborgs: Experiment 2

Unfortunately, the results of a single experiment such as this can tell us little about commonsense ascriptions of phenomenal states *in general*. This is, of course, one of the reasons why the Gray, Gray, and Wegner (2007) study is so intriguing. Perhaps the capacity to feel pain is a strange sort of phenomenal state, even though the analysis of pain has played a key role in the development of philosophical theories of consciousness (cf., Hardcastle 1999). This, I suggest, provides a motivation for developing further experiments designed to test commonsense ascriptions of other sorts of states that have typically been seen as states of phenomenal consciousness. With this in mind, it is worth returning to the paradigm cases of phenomenal consciousness advanced by Block. Recall that according to Block, a person is phenomenally conscious when she has a visual, auditory, or tactile experience, when she feels a pain, or when she experiences an emotion.

Thus, a promising strategy for empirical inquiry would be to focus on how commonsense psychology understands *qualia*, the introspectable phenomenal features that characteristically inhere in sensory experience (e.g., the color of an afterimage, the smell of brewing coffee, the pitch of a

sound, and the taste of a raw onion). Unfortunately, however, judgments about the nature of *qualia* seem to present an additional difficulty for experimental testing. Here is the problem. If volunteers were asked to report on their agreement with the claim that something *looked red* to various systems, the results of this inquiry may be confounded by the semantic nature of the term 'looks'. After all, 'looks' is polysemous. On the behavioral interpretation of 'looks', a system would only have to be able to detect red things and then report having done so. However, there is no reason to think that this capacity would not be present in a functional zombie. Since a functional zombie is supposed to be both behaviorally and functionally identical to a normal person, the capacity to detect colors and report having done so would necessarily be in place in such a system. On a 'phenomenal' interpretation of the claim, however, the system would not only have to detect the presence of a red thing but would also have to be the subject of an *immediate experience* of red. Similar considerations would hold for questions about other sensory modalities. Given that there is no obvious way to guarantee that all volunteers would understand a question about a visual experience in a way that would guarantee the phenomenal reading, an experiment designed to examine the commonsense understanding of the qualitative features of sensory states seems ill advised.<sup>14</sup>

Unlike sensory experience, emotional experience provides a plausible test case for examining the commonsense understanding of phenomenal consciousness. First, Georges Rey (1980) argues that there must be biological restrictions on the capacity to be in emotional states. Rey claims that unless we constrain the ascription of emotions to biological systems, in particular systems with a neurophysiological constitution that is relatively similar to ours, we will run the risk of advocating an implausibly liberal theory of the mind (cf., Block 1978) that would even allow a group of people, as such, to be in emotional states. Rey, thus, advocates a version of *phenomenal distinctness* for emotional experience.<sup>15</sup> Second, focusing on the ascription of emotional states would allow me to determine whether physical realization plays an important role in the ascription of emotion, as suggested by Knobe and Prinz (forthcoming). Finally, just as the experience of pain could be determinately picked out as a qualitative state by adding the 'feels' locution;

<sup>14</sup> Sytsma and Machery (unpublished data) have recently collected data suggesting that although non-philosophers *are* willing to say that a simple robot can 'see red', philosophers find such an ascription odd. Moreover, the responses of the non-philosophers were not "bimodal as would be expected if they found 'seeing red' to be ambiguous between distinctive behavioral and phenomenal readings" (Sytsma, personal communication). This suggests that my worry about using sensory states may be misplaced; there may, in fact, be important differences between the way in which sensory states are ascribed to my four systems, compared to the ascription of feeling pain and to feeling happy. This, however, must remain a question for further empirical investigation.

<sup>15</sup> Rey's position is an interesting case here. After all, Rey (1988, 6) is inclined to think that consciousness "may be no more real than the simple soul exorcised by Hume." However, the worries in Rey (1980), are driven by a concern with Block's Nation of China thought experiment, which is grounded on a worry about phenomenal consciousness.

<sup>13</sup> It might, for example, be the case that a metallic system that was described as having external sensors that were functionally homologous to the C and Aδ fibers that we find in biological systems can be in pain states. In the absence of this experimental evidence, I'll hazard a guess that it's soft, biological tissue that is doing the work here.

similar effects can be achieved by adding a ‘feels’ locution before emotional terms. This assures that the phenomenal character of the state is made transparent.<sup>16</sup> With these considerations in mind, I developed a second experiment designed to examine the commonsense ascription of emotional states.

Ninety-nine (99) volunteers from introductory philosophy classes at UNC – Chapel Hill participated in this second experiment. The demographic makeup of these volunteers was again consistent with the demographics of the student population at UNC – Chapel Hill more broadly. Volunteers were once again randomly assigned to the four conditions used in experiment 1 and were provided with questionnaires that included the same pictures and the same brief scenarios used in experiment 1. In all four of the conditions, volunteers were asked to rate their agreement with two claims about David on a 5-point scale ranging from ‘1-strongly disagree’ to ‘5-strongly agree’: “He believes that triangles have three sides”, and “He feels happy when he gets what he wants”. The mean scores for each of these judgments were as follows:

	Human-Human Brain	Human-CPU	Robot-Human Brain	Robot-CPU
Belief	4.36	4.23	3.82	4.07
Emotion	4.21	3.15	3.32	3.19

Mirroring the analyses carried out in my first experiment, I once again analyzed volunteers’ ascriptions of mental states to systems with human bodies as opposed to robot bodies and volunteers’ ascriptions of mental states to systems with human brains compared to systems with CPUs. Once again, volunteers did not exhibit a significant difference in their willingness to ascribe a belief to a system with a human body as opposed to a robot body;<sup>17</sup> nor did they exhibit a significant difference in their willingness to ascribe a belief to a system with a human brain as opposed to a CPU.<sup>18</sup> Moreover, although they did exhibit some difference in their ascriptions of the feeling of happiness to both a system with human body as opposed to

<sup>16</sup> On the basis of the results reported by Knobe and Prinz (forthcoming), I assumed that there would be a significant distinction between volunteers’ ascription of phenomenal states with and without the ‘feels’ locution. Knobe and Prinz found that their volunteers were willing to ascribe emotional states to a corporation in the absence of the ‘feels’ locution, but were unwilling to ascribe the same states to a corporation when the ‘feels’ locution was included. As philosophers typically distinguish between behavioral and phenomenal understandings of many states, I assumed that this effect would be present in ascriptions emotional states. Unfortunately, this worry may be empirically misguided. Recent data collected by Sytsma and Machery (MS, Study three) suggests that there is no significant difference in volunteers’ judgments about the acceptability of the sentences ‘Microsoft is feeling depressed’ and ‘Microsoft is depressed’. Sytsma and Machery (unpublished data) also report finding no commonsense distinction between behavioral and phenomenal understandings of ‘anger’; in fact, including the ‘feels’ locution does not seem to yield a significant difference in the ascription of this state to a simple robot. Perhaps this means that my inclusion of the feels locution may not have had a significant effect on my results; this will not, however, impugn the results I report below.

<sup>17</sup>  $F(1, 105) = 2.33, p = .13$

<sup>18</sup>  $F(1, 105) = 70, p = .405$

robot body, as well as to a system with a human brain as opposed to a CPU, these differences trended toward, but did not reach, significance.<sup>19</sup>

	Belief	Emotion		Belief	Emotion
Human body	4.30	3.71	Human brain	4.22	3.71
Robot Body	3.95	3.25	CPU	4.02	3.24

However, although there was no significant main effect for either having a human brain or a human body, there was a significant interaction effect. Volunteers exhibited a significant difference between their ascriptions of the feeling of happiness to a system with a human body *and* a human brain as opposed to every other system;<sup>20</sup> however, there was no similar difference in ascriptions of belief.<sup>21</sup>

This result provides evidence for the claim that physical constitution plays a significant role in determining whether a system can be a locus of emotional experience. Once again, however, the plausibility of *phenomenal distinctness* for emotion turns on the case of the human with a CPU. In this case, and in contrast to the ascription of the capacity to feel pain, there *was* a highly significant difference in volunteers’ ascriptions of belief as opposed to the feeling of happiness to a human with a CPU instead of a brain; strangely, however, there is still a significant degree of correlation between volunteers’ ascriptions of belief and the capacity for emotional experience for this system.<sup>22</sup>

#### 4.1 Does commonsense hold that emotions are uniquely human?

This experiment replicates my findings in Experiment 1 insofar as non-phenomenal states are concerned. Commonsense psychology is thoroughly functionalist in its ascription of beliefs; any philosophical view that purports to demonstrate otherwise should rely on empirical considerations, rather than the appeal to the implausibility of an aggregation of beer cans being in a mental state, in order to demonstrate that there are biological constraints on the capacity to have non-phenomenal mental states. Because, the philosophical arguments that have been used to undercut functionalism about non-phenomenal states have typically relied on thought experiments that fail to make the parallels in functional organization sufficiently clear, their intuitive pull suggests little more than a failure of imagination on the part of their proponents (cf., Dennett 1998). Unfortunately, making sense of the commonsense understanding of emotion turns out to be much more complicated.

First, the responses provided by my volunteers provides some evidence for Rey’s claim that there is an *a priori* constraint on emotional experience

<sup>19</sup>  $F(1, 105) = 3.19, p = .077$  and  $F(1, 105) = 3.67, p = .058$

<sup>20</sup>  $F(1, 105) = 6.16, p = .015$

<sup>21</sup>  $F(1, 105) = .08, p = .781$

<sup>22</sup>  $t(25) = 4.06, p = .0004$  and  $r = .40, p = .042$

that requires an emotion to be realized in a system with a neurophysiological constitution that is relatively similar to ours.<sup>23</sup> This is not, of course, to claim that functionalism about emotion fails. Rather, my evidence shows that since functionalism about emotion is an empirical and philosophical proposal, evidence for its truth must be garnered from our best philosophical and empirical theories of the mind. This suggests, then, that the commonsense understanding of emotion is broadly consistent with one understanding of *phenomenal distinctness*—a version that recognizes both the importance of neurophysiological, as well as more broadly bodily states, for the experience of an emotional state. However it is also important to remember that volunteers' responses were *significantly different* for beliefs and the feeling of happiness in a system with a human body and a CPU. Given the lack of a significant difference in this system where pain is concerned, this suggests an important difference between the commonsense understanding of having an emotional experience and the commonsense understanding of feeling pain.

This brings me to a second point, which will help to make the difference between the cognitive tools that are used in the ascription of pain and in the ascription of emotion. My data provides some evidence that commonsense is inclined toward a non-cognitivist theory of the emotions.<sup>24</sup> According to the non-cognitivist, emotions are best explained by appeal to bodily states and the monitoring thereof. As William James (1884, 193) famously argues, if we imagine an emotional state and then abstract away all of the “characteristic bodily symptoms, we find we have nothing left behind, no ‘mind-stuff’ out of which the emotion can be constituted.” Instead, all that remains is a cold, intellectual state. In accordance with this claim, people tend to think that the presence of a human body impacts the capacity of a system to be a locus of emotional experience, even though it is not sufficient to yield the capacity to experience an emotional state. However, people don't think that emotion is in the body; it's in the system that contains the body and the brain. This, however, leaves the most interesting fact about this data unexplained: why are the conditions under which people ascribe the feeling of happiness distinct from the conditions under which people ascribe the feeling of pain? Unlike the ascription of pain, emotion requires that neurological structures play a key role in the ascription of emotional states. But why should this be the case?

A first pass at an explanation might be to notice that there is a prominent recognition in our culture that we have the capacity to modulate emotions by chemically intervening on neurological structures. A typical college student is

surely aware, either first-personally or through third-person observation, that the use of alcohol, marijuana, or cocaine has a robust effect on emotional states. Moreover, the proliferation of television commercials, billboards, and email spam advertising SSRIs and Welbutrin as ways of modulating emotional experience demonstrates an increasing social acceptance of modulating emotion chemically within our current cultural milieu. It should, then, come as no surprise that when people are asked whether a system can be a locus of emotional experience, they are likely to see the brain as playing *an important role* in that judgment. However, this move is much too quick. After all, there is also a prominent recognition in our culture that we have the capacity to modulate the sensation of pain by chemically intervening on a distinct range of neurological structures. Perhaps, then, this distinction suggests a far deeper difference in the way that we ascribe feelings of pain and emotional experiences.

## 5. Do androids feel happy when they get what they want?

In the opening section of this paper, I made the suggestion that a theory that accounts for the subtleties in commonsense ascriptions of mental states is likely to offer important insights into the strategies used by people who have not studied philosophy or cognitive science in order to make sense of other people's mental states. I claimed that the analysis of these intuitions would help to develop a better philosophical understanding of what it takes for us to understand a person as a genuine locus of mentality. I argue that our capacity to distinguish between ‘mere things’ and ‘subjects of moral concern’ rests, to a significant extent, on the sorts of mental states that a system happens to have. In this section, I offer a general analysis of this data, and then develop some initial suggestions about the role played by mental state ascriptions in distinguishing between persons and things.

I began with three hypotheses about the physical realization of mental states: *biological naturalism*, *organizational invariance*, and *phenomenal distinctness*. My data suggests that none of these theories captures the conditions under which various mental states are ascribed by commonsense psychology. *Biological naturalism* fails because people adopt a functionalist understanding of non-phenomenal states such as belief; neither *organizational invariance* nor *phenomenal distinctness* captures the *phenomenal embodiment* of pain; however, commonsense psychology ascribes emotional states in a way that is consistent with one form of *phenomenal distinctness*. Contrary to the hypotheses advanced by Knobe and Prinz (forthcoming), commonsense psychology does not utilize a single concept of phenomenal consciousness in ascribing mental states. In fact, people who have not been trained in academic philosophy or cognitive science are committed to a different understanding of the conditions under

<sup>23</sup> This is also a view that is advanced by many contemporary proponents of the James-Lange theory. See, for example, Damasio (1994) and LeDoux (1996).

<sup>24</sup> To put the point more negatively, commonsense psychology is not committed to cognitivism about emotional states. According to cognitivism, an emotion is nothing more than a particular species of evaluative judgment. Cognitivism has been a dominant philosophical view of emotion; however, because commonsense psychology takes emotion to require a physiological component that cannot be adequately captured by mere appeal to functionally specifiable states like judgment, this theory seems to fail to capture the commonsense understanding of the emotions.

which pain should be ascribed and the conditions under which the feeling of happiness should be ascribed.<sup>25</sup>

With this in mind, we can turn to one of the perennial questions of philosophy: What does it take for us to understand something as a person? In answering this question, the first thing to notice is that the failure of *biological naturalism* as a theory of commonsense psychology suggests that the capacity to believe is not enough to guarantee personhood. Our ascriptions of belief cast a broad net, as Daniel Dennett has often noted, allowing us to predict and explain a system's behavior on the basis of minimal assumptions about what it would be rational for that system to do. In order to understand something as a believer, we only need to be able to adopt an intentional stance in explaining that system's behavior.<sup>26</sup> However, although some minds possess only intentional and goal directed states, others are much richer. Perhaps a robot could be a believer, but it doesn't seem to follow from this that we should thereby treat it as a subject of moral concern.

Philip Robbins and Anthony Jack (2006) have recently argued that we sometimes adopt a 'phenomenal stance' toward a system, and doing so forces us to see that system as a subject of moral concern. In a similar vein, Knobe and Prinz (forthcoming, 26) argue that in ascribing a phenomenal state, "what we really want to know is whether or not the entity is capable of having genuine feelings". On this view (Robbins and Jack 2006, 69-70), adopting the phenomenal stance requires both a sympathetic appreciation of what it's like to feel what that system is feeling, and an inclination to treat that system as a subject of moral concern. Both Robbins and Jack (2006) and Knobe and Prinz (forthcoming) argue that the recognition of something as a locus of experience in general, and pain in particular, is sufficient to pull that system into the realm of moral consideration. However, my data raises a worry about this sort of claim. If all that needs to be the case for a system to count as a subject of moral concern is that the system have the capacity to feel pain, then a cyborg with a human body and a CPU instead of a brain would have equal moral status to a system with a human body and a human brain. But, it is not clear that this is the sort of intuition that we should expect to find as a component of commonsense psychology.

---

<sup>25</sup> This being the case, philosophers must be careful in designing their thought experiments. Although it is received wisdom that phenomenal states form a unique class of phenomena, my data demonstrates that not all phenomenal states are created equally. The sort of mental state that one chooses in framing a thought experiment might significantly impact its plausibility. It is on this point that philosophers have much to learn from the best experimental psychologists. It is only by *systematically* varying cognitive systems and mental states that we can develop a robust hypothesis about the features of a system that are likely to play a role in our ascriptions of mental states.

<sup>26</sup> The term 'intentional' raises a number of worries. Most philosophers would agree that *many* phenomenal states *are* intentional; some also believe that all phenomenal states are intentional. More importantly, some philosophers believe that all intentional states also have phenomenal content. My use of 'intentional' indicates only that a robot can be conceived of as possessing directed mental states, though lacking any phenomenal states.

Emotion is ascribed more sparingly; and I argue that this is because the ascription of emotion requires a willingness to treat something as 'like me' (Meltzoff 2005) in a much more robust sense than is required for the ascription of pain. In making sense of this claim, it is important to keep three facts about my data in mind. First, people are less likely to ascribe an emotional state to a system with a human body and a CPU instead of a brain; second, neither a system with a robot body and a human brain, nor an ordinary robot are any less likely than a human with a CPU to be the subject of an emotional state; and third, there is a significant degree of correlation for the ascription of belief and emotion to a human with a CPU—that is, although *in general* people see a significant difference in the capacity of such a system to have beliefs compared to emotions, when presented with the ascription of emotion and belief side-by-side, a significant number of people opt for a view that sees emotion and beliefs as on all fours.

In the remainder of this section, I offer a hypothesis that explains this data as well as explains what it would take for an android to be treated as a locus of happiness. The heart of my suggestion is this. While judgments about pain can be read directly off of the soft and squishy body of a system, and while such ascriptions can be used to predict the behavior of a system given its physical structure, judgments about the capacity to feel an emotion recruit hot-empathetic mechanisms that drive us into the realm of moral concern. To put the point differently, the empathetic mechanisms that drive our fast-and-frugal ascription of emotional experience are intimately tied up with our capacity to treat something as a subject of moral concern in a way that our ascriptions of the feeling of pain are not.<sup>27</sup> In order to make this hypothesis more plausible, consider the following question. What would it take, from the standpoint of commonsense psychology, to treat a system with a human body and a CPU as a subject of moral concern?

In an attempt to answer this question, consider the depiction of replicants, cyborgs that are physically indistinguishable from ordinary humans, in the 1982 film *Blade Runner*.<sup>28</sup> The replicants in this film are designed to do work on off-world colonies where it is too dangerous for ordinary humans to work. From the very beginning of the film, these replicants appear to have the capacity to feel pain; however, the mere fact that they feel pain does not compel anyone to count them as subjects of moral concern. In fact, from the standpoint of an ordinary person, replicants are systematically dehumanized; they are referred to as skin-bags, treated as mere objects, and are 'retired' when they behave in any way that is seen as problematic. More importantly, at the beginning of the film, there does not

---

<sup>27</sup> This is, of course, a bold hypothesis that warrants empirical investigation. In collaboration with Jesse Prinz, I am currently in the process of developing an experimental paradigm designed to test precisely this hypothesis. If this hypothesis is corroborated, it will also open up a number of interesting avenues for investigation into the nature and scope of human moral psychology.

<sup>28</sup> Thanks to Robert Briscoe for suggesting this intriguing analogy to me.

seem to be anything wrong with retiring the replicants that have returned to earth. After all, they are just robots.

The characters in the film, as well as the audience is lead to believe that there is no reason to treat the replicants as subjects of moral concern. But, what is supposed to distinguish the replicants from the rest of us? What are they lacking? Throughout the film, it becomes clear that the factor that distinguishes a human from a replicants is not a fact about the body that the system has, but the capacity to experience emotion. Because the replicants are biological systems that are physiologically indistinguishable from ordinary humans, the Blade Runners who are sent to retire a replicant must use a device called a Voigt-Kampff machine in order to determine whether a suspect is a human or a replicant. As we learn early in the film, this machine measures the low-level physiological responses associated with human emotion. By asking people to respond to various sorts of scenarios and testing for their emotional response, it is possible to determine whether a system is a human or a mere machine. The assumption, here, is one that lies deep at the heart of commonsense psychology: only something that is 'like me' in its immediate emotional responses could be something that I should care about.

As the film progresses, however, director Ridley Scott creates the impression that a replicant can indeed be a locus of emotional experience, thereby engaging hot-empathetic mechanisms that transform the impression that the viewer has of replicants. Scott arranges the film in such a way that the viewer begins to empathize with the plight of the replicants, and eventually leads the viewer to recognize that even replicants are systems that should be considered as subjects of moral concern, thereby effectively transforming the replicants from 'mere things' to genuine subjects of moral concern.

This is how the transformation occurs. Early on in the film, we learn that a group of four replicants have returned to earth because they are afraid of their impending deaths—the exact hour of which they now know.<sup>29</sup> Later on, we learn that some of the more advanced replicants have the capacity to feel sad, as a replicant named Rachel does when she learns that her memories have been fabricated. She has no history, even though she believes that she has a history from the standpoint of first-'person' cognition. This recognition, however, leads Rachel and the Blade Runner, Deckard, to eventually fall in love. All of this leads us to begin to question the initial thought that the internal life of a replicant must be cold and empty—to question whether replicants are truly nothing more than functional zombies. However, in the

---

<sup>29</sup> This fear becomes clearest in a conversation between the Blade Runner, named Rick Deckard, and a replicant, named Leon that Deckard is trying to retire (cf., Mulhall 1994). The replicant is about to kill Deckard and the conversation runs as follows:

Leon: My birthday is April 10th, 2017. How long do I live?

Deckard: Four years.

Leon: More than you. Painful to live in fear, isn't it? Nothing is worse than having an itch you can't scratch.

culminating scene of the film, the leader of the escaped band of replicants destroys the impression that he is a 'mere thing' by showing more than a glimmer of compassion for the life of Deckard.

By the end of the film, when we have come to recognize that replicants are not so substantially different from us in their capacity to feel emotion, viewers generally develop the capacity to empathize with replicants, to feel what it must be like to be one of them, and at this point viewers gain the capacity to treat the replicants as subjects of moral concern. At the end of the film, we are given the chance to reflect on whether it is wrong to 'retire' a replicant.<sup>30</sup> Because a replicant can be a locus of emotional feeling, it can be a locus of moral concern.

My argument, in brief, is this. Our capacity to ascribe emotion engages hot-mechanisms of empathetic thought. When we ascribe an emotion to another system, we simulate an emotion in that system, we imagine ourselves in that system's place, and we feel whatever it is that we assume that she would feel. However, this means that our standards for determining whether a cyborg, or a robot, should be seen as a locus of emotional states are less firmly entrenched in human psychology than are the standards for ascribing other sorts of mental states to that same system. What it takes to generate an ascription of emotion to such a system is the recognition that this system can be 'like me' in its emotional experience. When this fact about a system becomes clear, that system is immediately propelled into the realm of moral consideration.

At this point, we can return to my correlation data, which suggested that there was a high degree of correlation between the ascription of belief and emotion to a system with a human body and a CPU instead of a brain. The first thing to notice is that the capacity to ascribe emotion requires at least that a system be treated as the sort of system that can have mental states at all (cf., Dennett 1981, 240). Thus, people who are unwilling to ascribe beliefs to a system should also be unlikely to ascribe emotions to that system. Merely making assumptions about the capacity of a system for rational action can do all of the work in the ascription of belief; however, the ascription of emotion requires really imagining what it is like to be that system, recognizing that system as relevantly similar to one's self. On the other side, where emotion is readily ascribed, the explanation comes in the form of differences in the capacity for simulation. Some people have an easier time simulating the mental states of another system and can do so in a way that makes their emotional states far more vivid from the standpoint of first-person cognition. Because these people have a much easier time imagining what it would be like to be another system, regardless of that system's physical constitution,

---

<sup>30</sup> This is made clearest in the Final Cut version of the film released in 2007. In this version, the closing scene provides us with reason to think that the Blade Runner, Deckard, might also be a replicant, complete with fabricated memories and implanted daydreams about unicorns. However, if Deckard is a replicant, then he too must be retired. But, having gone through the emotional highs and lows in this film, having felt genuine compassion for Deckard, the thought that he too could be 'retired' seems well beyond the pale of moral acceptability.

they more readily engage the simulating mechanisms required for ascribing emotions to a system.

The point, here, is that commonsense psychology takes the capacity to have emotional experiences to be special. If a system feels emotions then it can fear its own death, hope that tomorrow will be a better day, and even be happy when it has an exciting philosophical interaction with a new colleague. Once we have ascribed emotional states to a system, our judgments about whether that system should count as a locus of emotion, then, inform our judgments about whether that system should be seen as a locus of moral concern. I suggest that what determines whether a system should be seen as a locus of emotional experience is not some fact about its physical features, but instead is a fact about whether a person can empathize with that system. For most of us, whether we can empathize with a system might be specified in terms of having a human body or having a biological brain. However, if there is anything to learn from the replicants in *Blade Runner*, it's that the factors relevant to our capacity to empathize are malleable, and are at least in part determined by the way that a system behaves toward us—regardless of the physical constitution of that system.

## 7. Acknowledgements

This paper has benefited greatly from conversations with Robert Briscoe, Jacek Brozowski, Joshua Knobe (who went above and beyond the call of duty, showing a love of philosophy and a love of the data even where it conflicted with his own views), Jesse Prinz, and Dylan Sabo. I am also grateful to Peter Bokulich, Mike Bruno, Steve Grossberg, Bill Lycan, Eric Mandelbaum, Susanne Sreedhar, Daniel Stoljar, Justin Sytsma and my audience at the *Boston University Colloquium in the Philosophy of Science* for their helpful comments and suggestions.

## 8. Works cited

- Arico, A. (2007). Should corporate consciousness be regulated? Poster presented at the annual meeting of the Society for Philosophy and Psychology.
- Arico, A., B. Fiala & S. Nichols (forthcoming). The folk psychology of consciousness.
- Block, N (1978). Troubles with functionalism. CW Savage (ed.), *Minnesota studies in the philosophy of science IX*. Minneapolis: University of Minnesota Press: 261–325.
- \_\_\_\_\_. (1986). Advertisement for a Semantics for Psychology. P French, et. al. (eds) *Midwest Studies in Philosophy X*: 615–678.
- \_\_\_\_\_. (1995). On a confusion about a function of consciousness. *Behavioral and brain sciences*, 18: 227–47.
- \_\_\_\_\_. (2003). Mental paint. M. Hahn and B. Ramberg (eds), *Reflections and replies*. Cambridge: MIT Press: 125–51.

- \_\_\_\_\_. (forthcoming). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and brain sciences*.
- Bruno, M., B. Huebner, & S. Sarkissian (in prep). What does the Nation of China think about phenomenal states?
- Chalmers, D. (1995). Absent qualia, fading qualia, dancing qualia. *Conscious experience*, T. Metzinger (ed). Exeter: Imprint Academic.
- \_\_\_\_\_. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Cummins, R. (1983). *The nature of psychological explanation*. Cambridge: MIT Press.
- Damasio, A.R. (1994). *Descartes error*. New York: Penguin Books.
- Dennett, D. (1981). *Brainstorms*. Cambridge: MIT press.
- \_\_\_\_\_. (1988). When philosophers encounter artificial intelligence. *Daedalus* 117: 283–295.
- \_\_\_\_\_. (1998). *The unimagined preposterousness of zombies, Brainchildren*. Cambridge: MIT Press: 171–179
- Farrell, B. (1950). Experience. *Mind*, 59 (234): 170–198.
- Gray, H., K. Gray & D. Wegner (2007). Dimensions of mind perception. *Science*, 619: 315.
- Hardcastle, V. (1999). *The myth of pain*. Cambridge: MIT Press.
- Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical explorations* 10: 95–118.
- Knobe, J. (2007). Experimental philosophy and philosophical significance. *Philosophical explorations* 10: 119–122.
- Knobe, J. & Prinz, J. (forthcoming). Intuitions about consciousness: Experimental studies. *Phenomenology and the cognitive sciences*
- LeDoux, J. (1996). *The emotional brain*. New York: Simon and Schuster.
- Levine, J. (1983). Materialism and qualia. *Pacific Philosophical Quarterly* 64: 354–361.
- Lycan, W. (1987) *Cosciousness*. Cambridge: MIT Press.
- \_\_\_\_\_. (1996). *Consciousness and experience*. Cambridge: MIT Press.
- Meltzoff, A. (2005). Imitation and Other Minds: The "Like Me" Hypothesis. S. Hurley and N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science* (Vol. 2, pp. 55–77). Cambridge, MA: MIT Press, 2005.
- Mulhall, S. (1994). Picturing the human (body and soul): and interpretation of *Blade Runner*. *Film and philosophy*, 1: 87–100
- Nagel, T. (1974), What is it like to be a bat. *The philosophical review*, 83 (4): 435–50.
- Robbins, P. & A. Jack (2006). The phenomenal stance. *Philosophical studies*, 127: 59–85.
- Rey, G. (1980). Functionalism and the emotions. A. Rorty (ed), *Explaining emotions*. Berkeley: University of California Press: 163– 195.
- \_\_\_\_\_. (1988). A question about consciousness. H. Otto and J. Tuedio (eds) *Perspectives on Mind*. Dordrecht: D. Reidel: 5–24.
- Robbins, P. & A. Jack (2006). The phenomenal stance. *Philosophical studies*, 127: 59–85.

Scott, R. (1982) *Blade Runner*. Warner Bros. Entertainment, Inc.  
Searle, J (1992). *The rediscovery of the mind*. MIT press  
Sytsma, J., & E. Machery (MS). How to study folk intuitions about  
phenomenal consciousness.  
\_\_\_\_\_. (unpublished data) *Two conceptions of subjective experience*.  
Tye, M. (1999): Phenomenal consciousness. *Mind* 108: 705-725.