Mandik, P. (in press). An epistemological theory of consciousness? In Alessio Plebe, ed. *Philosophy in the Neuroscience Era, Special issue of the Journal of the Department of Cognitive Science*, Univ. of Messina.

An epistemological theory of consciousness?

PETE MANDIK, Department of Philosophy William Paterson University of New Jersey.

ABSTRACT: This article tackles problems concerning the reduction of phenomenal consciousness to brain processes that arise in consideration of specifically epistemological properties that have been attributed to conscious experiences. In particular, various defenders of dualism and epiphenomenalism have argued for their positions by assuming special epistemic access to phenomenal consciousness. Many physicalists have reacted to such arguments by denying the epistemological premises. My aim in this paper is to take a different approach in opposing dualism and argue that when we correctly examine both the phenomenology and neural correlates of phenomenal consciousness we will see that granting the epistemological premises of special access are the best hope for a scientific study of consciousness. I argue that essential features of consciousness involve both their knowability by the subject of experience as well as their egocentricity, that is, their knowability by the subject as belonging to the subject. I articulate a neuroscientifically informed theory of phenomenal consciousness—the Allocentric-Egocentric Interface theory of consciousness—whereby states of recurrent cortical networks satisfy criteria for an epistemological theory of consciousness. The resultant theory shows both how the epistemological assumptions made by dualists are sound but lead to a reductive account of phenomenal consciousness.

"…if I were a reductionist, I would be this sort of reductionist…"

David Chalmers 1996, p. 189

§1. Slipping Into the Zombie-free Zone

This article takes up the issue of the plausibility of *epistemological theories of consciousness*: solutions to the so-called "hard problem" of phenomenal consciousness (Chalmers 1996) that are rooted in physicalistic explanations of what we know and how we know it. Such accounts elaborate how physical systems come to (perceptually) know their physical environments and show how perceivers may come to find themselves positing and puzzling over the phenomenal aspects of experience: the qualia that have given physicalists so many headaches over the years. The main point of an epistemological theory of consciousness is to solve (or dissolve) apparent *metaphysical* issues concerning qualia in favor of *epistemological* explanations of why minds in the business of knowing about the physical world would ever come to think that there were qualia in the first place. In its most extreme form, for instance, as articulated by Daniel Dennett (1991), the point of an epistemological theory of consciousness is to show that once it is explained why we *think* that there are qualia, there is no further *metaphysical* work to be done in explaining qualia themselves. In other words, once the structure of our justified (and not so justified) beliefs concerning qualia is laid bare, no further

work—only damage and confusion—is done by supposing that these beliefs are true. A (perhaps derisive) description of such a project is as forgoing any explanation of qualia in favor of explaining them away. In less extreme forms of epistemological theories, as in the case of recent work by Andy Clark (2000a, 2000b), it is not denied that there are metaphysical facts about qualia. Instead, the point of less extreme epistemological theories is that certain facts about our epistemologies entail certain metaphysical facts about qualia. Even Chalmers himself flirts with such an epistemological account (1996, pp. 287-292) but as Clark notes, Chalmers ultimately does not pursue it (2000a, p. 31). My aim in this paper is to pursue such a theory, and further, to flesh out how both neuroscience and philosophy converge to support such a view. I begin by focusing on Clark's work. I argue that Clark's account fails, but that its failure instructively illuminates the path that any successful epistemological theory of conscious must follow. In later sections I flesh out what sort of beast one finds when one follows the path to its end.

Clark's argument hinges on an oft-cited distinction in work on consciousness: Block's distinction between *phenomenal* consciousness and *access* consciousness (Block, 1995). Phenomenal consciousness is what you have when there is *something it is like to be* you. You are subject to phenomenal consciousness when you are subject to mental states that have qualia. Access consciousness is a more boring form of consciousness. You are subject to access consciousness only insofar as you are sensitive to information that you may subsequently use in various ways, including verbal report. Access and phenomenal consciousness are supposed to be distinct notions. The most notorious cases in which these notions come apart are in hypothesized *zombie* cases—cases involving subjects devoid of phenomenal consciousness. Zombies, in spite of being phenomenally vacant, may exhibit similar outward behaviors and internal physiologies to us and thus are able to pick up and report on all the information about their physical environments and bodies' interiors that we are. However, one need not buy into the preposterous imaginability of zombies to appreciate the possibility of access without phenomenology. Human blindsight patients are offered as real world examples in which access to visual information (e.g. that a red chair is present) is had without the phenomenal raw feels that make it the case that there is something it is like to see a red chair (Weiskrantz, 1996). Whether phenomenal and access consciousness are entirely distinct is questionable and oft questioned (see, for example, Dennett, 1995). Recently Clark has argued for a case in which a certain kind of access consciousness entails a certain kind of phenomenal consciousness. More specifically, according to Clark, if one has access to information regarding which of several sensory modalities in virtue of which one is responsive to the world, then there must be some phenomenal feel that distinguishes one modality from the next—there must be something it is like to sense the world through one modality rather than the other. Clark's claim is especially interesting in the way he unpacks it. The reason that this entailment is supposed to hold is supposed to provide an entering wedge for a solution to the notorious "Hard Problem" of consciousness (Clark 2000a, p. 32).

Before considering Clark's argument that there is at least one case in which access consciousness entails phenomenal consciousness, it will be useful to consider a case in which phenomenal consciousness follows trivially from access consciousness. It will be useful to have such a case in mind to evaluate whether Clark's argument is just a version of this trivial inference. Consider first the following minimal characterization of

what qualia are: they are the introspectible similarities and differences of sensory experiences (Mandik, 1999, p. 47; Lycan, 1996, p. 69-70). What it is like to see a banana is more like what it is like to see a plantain than it is to smell a lime, and I know this by introspection. I am in cognitive contact, then, with my own qualia. This epistemological intimacy we share with our own qualia is an important plank in many qualiophilic platforms. It is the basis of the prevalent attitude that our qualia are obvious to us, as expressed by Block's borrowing of Louis Armstrong's remark on jazz "If you got to ask, you ain't never gonna get to know" (1978, p. 281). Thus, on many accounts of what qualia are, they are among the things that a subject has cognitive access to. Among the things I have access consciousness of are the introspectible similarities and differences between my experiences. I have access consciousness, then, of facts about my phenomenal consciousness. There is thus at least one (trivial) inference from access to phenomena: Access to phenomenal consciousness entails the existence of phenomenal consciousness. (Note that the entailment goes the other way too, since it is built into the notion of phenomenal consciousness that one have accesses consciousness concerning facts about one's own qualitative states.)

Is Clark's argument a version of this trivial one? Let us begin by examining his argument. Clark invites us to imagine an organism or perhaps even a robot capable of both perceptually discriminating items in its environment and introspectively discriminating which of several sensory modalities it employs to achieve its perceptual discriminations. Suppose, then, that the system in question perceptually judges that bananas are more similar to plantains than to pomegranates and that the sensory modality by which it comes upon this information is its visual modality. Suppose further that the system is able to answer questions about how it made the judgment concerning its sensory modality. According to Clark, such a system. . .
. . . must say *either:*

> (a) I have no access to the act by means of which I detect the difference. The answer just comes to me. I perceive nothing when I make my judgments - I simply find myself saying that there are two objects, one red and one yellow, and so on.
> Or:
> (b) I have access not just to the products of my sensory activity, but also to certain aspects of the sensory activity itself. For example, I am non-inferentially aware that I am using a visual rather than a tactile modality. I am aware that I see, rather than hear or feel, the difference. (2000a, p. 30)

Clark urges us to conclude that if the system gives the latter answer, then there must be something it is like for the system to obtain the information in question. Clark concludes that having access consciousness of which sensory modalities it used to make the perceptual judgment entails having a mental state with full-blown phenomenal consciousness. Note that the notion of entailment operative here is supposed to be stronger than, say, nomological entailment. For Clark, phenomenal consciousness is supposed to follow *conceptually* from the kind of access under consideration (2000a, p. 31).

I'll argue that Clark has not supplied the resources to adequately distinguish his inference from the trivial entailment mentioned above. Showing this will involve

showing how Clark's position occupies an unstable territory between two opposing camps: qualia liberals and qualia conservatives. The liberal sees qualia in lots of places that Clark does not. In contrast, the conservative sees qualia in far fewer places than Clark does.

For an extreme example of a qualia liberal, consider someone who holds that the mere fact that a system is able to make perceptual judgments about features of its environment entails the existence of phenomenal consciousness. On such a view, if a system is capable of perceptually discerning bananas from pomegranates, then there must be something it is like to perceive bananas such that what it is like is different from what it is like to perceive pomegranates. For this sort of liberal, in order to have qualia, the creatures need not access information about which modality they employed, they need only employ the modality. (See, for example, Stone, 2001 and, in particular, Lycan, 1996, pp. 76-77.) For the qualia liberal, the entailment Clark describes follows trivially. If a system must have phenomenal consciousness when it introspectively discerns that the perceptual basis of its judgments about the fruits was visual as opposed to tactile, then this is because phenomenal consciousness—what-it-is-like-ness—was *already there in the first place*. It was already there to be introspected, independently of the actual act of introspecting. It does not, as Clark would have it, arise *because* of the introspecting.

In contrast to the qualia liberal, the conservative holds that adding all the modality-specific introspective access in the world will be (logically, conceptually) insufficient to give rise to phenomenal consciousness. For a qualia conservative like Chalmers (1996), the phenomenal does not logically supervene on the physical and adding the ability to discriminate one's own sensory modalities is merely adding more physical stuff that does not logically entail the addition of qualia. While the liberal says that Clark's inference from access to phenomena follows trivially, the conservative says that the inference doesn't follow at all. For the qualia conservative, a system no more needs qualia to detect the difference between grapefruits and bananas than a thermostat needs qualia to detect the difference between one temperature and the other. The only things required are mechanisms sensitive to certain causally efficacious properties, but these mechanisms need not give rise to any qualia. All may be dark inside, and there is no more anything it is like to be a thermostat than to be a rock. Similarly, taking a system with mechanisms capable of detecting external states and adding to it mechanisms capable of detecting internal states of the previous mechanisms does nothing to change whether all is dark inside the system. If external states may be detected without qualia, so may internal states.

Clark's position constitutes an inadequately defended (by Clark) middle ground between qualia liberalism and qualia conservatism. For Clark, the mere presence of the first-order discriminatory states is insufficient to bestow qualia, but adding certain kinds of higher-order states does the trick. For the qualia liberal, the system with higher-order states has qualia, but only because the qualia were already present in the first order states. For the qualia conservative, there are no qualia necessitated by the first order states nor are any necessitated by the introduction of any higher-order states.

§ 2. Can there be a Transcendental Argument for a Reductive Theory of Consciousness?

Despite the flaws pointed out above, Clark's project sheds important light on the future prospects of epistemological theories of consciousness. A key step in arguing for an epistemological theory of consciousness will be the following: whatever qualia are, they are such that we can and do have knowledge of them. Recall the key fact upon which the trivial inference from access to phenomena turns: that qualia are such that we know them. Call this the "epistemological criterion". Might we use something like the epistemological criterion to construct a transcendental argument for physicalism?

The first steps of the argument will have as a premise the epistemological criterion—the premise that makes such an argument transcendental. The links to physicalistic conclusions will be given in portions of the argument whereby the conditions on being known are linked to physical conditions, through links relating knowledge to the causal aspects of perception and introspection. Such an argument involves the following desideratum for any theory of consciousness, namely, that it be impossible that, for all you know, you are a zombie.

Many qualiophilic arguments, starting with strong realist intuitions about qualia, end up with theses about qualia—in particular, qualia epiphenomenalism (see, for example, Jackson, 1982, Robinson, 1982, and Chalmers, 1996)—that leave the reader wondering how we can possibly ever know qualia. If qualia cause nothing, then how can we have justified beliefs about even our own qualia? The worry arises that qualia, which were supposed to be essentially phenomenal, thus become quintessentially noumenal. This should not come as a terrible surprise given the way it recapitulates the history of so much of classical metaphysics and epistemology. Strong external-world realism leads quickly to skeptical hand-wringing—how can we know a world that is so far away? It may seem odd that such a problem can arise for the furniture of the inner phenomenal world as well, but what is crucial is the way that a strong, metaphysical, realism cuts the ties between justification and truth (Putnam, 1981). The historical analogy between external-world realism and qualia realism serves as well to point out the natural place a transcendental turn might take on the way to bridge the explanatory gap between qualia and physiology: the turn depends on the twin suppositions that qualia are knowable and the knowledge must be physical through-and-through.

Clark argues that certain facts about access, facts I have construed as epistemological, force subjects into "phenomenal space" and "mark out a *necessarily* zombie-free zone" (2000a, p. 37). I have argued, contra Clark, that the zombie-free zone has not yet been adequately circumscribed, for it is not clear where and how Clark can stake out neutral territory between the camps of the qualia liberals and qualia conservatives. Though Clark's arguments are flawed, they do point in the promising direction of an epistemological theory of consciousness palatable to those unable to stomach Dennett's qualia nihilism. Clark's suggestions—and, I hope, mine—serve to light the way to a zombie free zone where qualia live, and, importantly, qualia are known.

Before further developing the argument, it is worth noting the following concerning reductionism and consciousness. Let us call a theory of consciousness reductive if it entails that my physical doppelganger cannot be a zombie. A reductive

theory of consciousness must have outlines that are discernible from both the first and third person points of view. The demand on first-person discernability arises from its being a theory of consciousness. The demand on third-person discernability arises from its being a reductive theory. A problem that plagues reductive theories of consciousness is the question of what, if anything, attaches the outlines discerned from the one point of view to the other. After the reductive theory has been described the worry arises that the separate portions may be implemented separately. This worry is oft expressed in terms zombies: creatures that constitute implementations of the aspects of the third person portions of a theory without simultaneously constituting implementations of the first person portions of the theory. Epistemological reductive theories attempt to bridge the gap from the first-person to the third-person by discerning epistemic features accessible from the first-person point of view that necessitate certain elements in the third-person point of view. What third-person facts must obtain for there to be first person knowledge of consciousness?

The crucial facts concern the knowability of one's own conscious states. My conscious states are necessarily knowable as such: knowable as my conscious states. In order for this to be true, conscious states must be conceptualized and egocentric. Being conceptualized accounts for knowability and being egocentric accounts for the knowability of a state *as my own*. These features of consciousness are discernable from the first-person point of view. Alleged counterexamples depend on attributions from the third-person point of view. This is why whenever one imagines the existence of a zombie, one imagines someone else being a zombie. Imagining that one is currently mistaken about whether one is a zombie is exceedingly difficult .

Now we are in a position to further spell out the transcendental argument. Spelling this out further, the argument in outline is:

1. I know that I am not a zombie
2. My knowing that I am not a zombie is constituted by the satisfaction of condition K
3. My physical doppelganger satisfies condition K

.: My physical doppelganger is not a zombie

Most of the heavy lifting will involve spelling out condition K such that 2 and 3 turn out true. I've already given some indication of what condition K will consist in. My knowing that I'm not a zombie entails that my conscious states are necessarily knowable by me as such: knowable by me as my conscious states. In order for this to be true, conscious states must be conceptualized and egocentric. In the next two sections I spell out the conceptual and egocentric components of condition K respectively.

§ 3. Knowledge and the Conceptual Content of Conscious Experience

Conscious states involve conceptualizable contents. There are two related lines of thought to consider here. The first concerns how the application of a concept in experience

influences what it is like to have that experience. So, for example, wine tastes different to me now that I have and am able to apply the concept of tannic acid. The night sky looks a lot different to me now that I have and can apply the concept of the heavenly bodies being different distances from me. Squeaks in my car engine sound a lot different to me now that I have and can apply the concept of a broken valve lifter.

The second consideration takes a bit more time to elaborate. The elaboration will involve arguing that since phenomenal experience is necessarily knowable, phenomenal experience is exhaustively constituted by conceptual content. Putting consciousness aside for a moment, let us consider some features of knowledge and concepts. Suppose that there is a rock that is heavy, lumpy and igneous. Suppose that George has the concepts of lumpiness and heaviness, but no concept of being igneous. Suppose further that at no point does George acquire the concept of being igneous. What, then, can George know about the rock? He may know that it is lumpy and heavy, but barring acquisition of the concept of being igneous bars George from knowing that the rock is igneous. That the rock is igneous is, relative to George, un-conceptualized residue. Since idealism about rocks is false, rocks are the sorts of things that can have lots and lots of un-conceptualized residue. In worlds with rocks but no knowers, rocks are 100% un-conceptualized residue.

Let's turn from rocks to phenomenal experiences and ask whether they can be, in whole or part, unconceptualized residue. One important thing to note about phenomenal experience is its first-person necessary knowability. This means that if a person has phenomenal experience then that person is necessarily able to know that they have phenomenal experience. I take it that not only am I not a zombie, but I know that I am not a zombie. I may not be able to know whether or not you are a zombie, but that would simply be a failure of third-person knowability. If phenomenal experiences are the sorts of things that might even be beyond the knowability of the persons that have them, then for all that person knows, they are a zombie, which I take to be absurd. If a phenomenal experience has any phenomenal quality, Q, that is beyond the knowability of the person having the experience, then for all that person knows, they lack experiences with Q. They would be a Q-zombie for all they know. Again, I take that to be absurd. Since non-zombies can know of themselves that they are non-zombies, phenomenal experiences can have no phenomenal qualities that are necessarily unknowable from the first-person point of view. If something is necessarily knowable by me in that every aspect of it is necessarily knowable by me, then it can have no aspect that outstrips my concepts. If there were such an aspect it would inaccessible from the first person point of view.

So far this seems to show only that there must be a correlation between phenomenal characteristics and phenomenal concepts. Why make the further step of identifying phenomenal character with the contents of phenomenal concepts? Here's why: If, with regards to the phenomenal, character is distinct from conceptual content, then it would be possible for me to be in two different phenomenal states even though I had the same doxastic state. That is, I could believe that I had an experience with quality Q on both occasions but in one case the belief would be true and the other case it would be false. However, if this is possible, then there would be states that are phenomenally distinct but subjectively indiscernible. I would be unable to know, from the first person point of view, whether I was in a state with quality Q or not. I could, for all I know, be a

Q-zombie. I take this to be absurd. It follows, then, that with respect to the phenomenal, character is not distinct from conceptual content.

We need to consider whether the conditions for conceptual content would be satisfied by my physical doppelganger. For the purposes of the current argument I can be relatively open about what counts as a physical doppelganger. Nothing in the current discussion necessarily hinges on whether my physical doppelganger has only the same intrinsic physical properties as me or has, in addition, a similar history of causal interactions with a similar environment. Suppose we are maximally broad about what physical similarity entails. We need only be careful that we do not assume at the outset that this includes phenomenal similarity. One might question, however, why physical similarity would entail conceptual similarity. A good response to this question would be another question, namely "why not?" My physical doppelganger is equally good at sorting objects and having conversations about them as I am. My physical doppelganger can solve various puzzles you might pose for him, at least, he will solve all the puzzles that I do. He will get himself out of many sticky situations, or at least, as many as I will be able to. I'm aware of no convincing account of concepts that would ascribe concepts only to me but not to a system capable of all the same sorting, solving, and conversing that I am. For these reasons my physical doppelganger is my conceptual doppelganger. It is time now to consider the second component of condition K.

§4 Egocentricity

If my conscious experience is such that I can know it as my own, it must not only be knowable, but knowable *as my own*. This latter feature of consciousness is what I shall call the egocentricity or subjectivity of conscious experience. Being knowable as my own is only one side of the subjectivity of consciousness, the other side is that beings insufficiently similar to me have a difficult time knowing what my conscious experience is like. Discussions of this aspect of the subjectivity of consciousness have been closely associated with the famous knowledge argument of Jackson (1982) which is itself built on certain features of conscious experience remarked upon by Nagel (1974), who famously posed the question of what it is like to be a bat. The answer that Nagel urged is that bat experience must be insufficiently similar to our own for us to know what it is like to be a bat. Jackson's knowledge argument attempts to use the subjectivity of conscious experience as a premise in an argument that certain aspects of conscious experience must be non-physical. In brief, the knowledge argument begins by supposing it possible for someone to know all of the physical facts without ever having had a conscious experience of seeing red objects. It is further supposed that such a person would not, then, know what it is like to see red. Assuming further that such knowledge—knowing what it is like to see red—is factual knowledge, it follows then (from this assumption and the previous suppositions) that this must be knowledge of a non-physical fact. There have been various physicalist responses to the knowledge argument and space does not permit a review of them here. In Mandik (2001) I argue that the requisite notions of subjectivity can be explicated in terms of egocentric representation in certain kinds of neural networks. Below I develop a somewhat different approach, but the notion of egocentric representation will nonetheless play a central role.

Most of the discussion of consciousness in the literature involves visual consciousness and this paper is no different in this regard. If we are interested in thinking of visual consciousness in terms of the physical properties that are responsible for it, it is useful to look at the portions of the human nervous system responsible for vision in terms of a processing hierarchy. This processing hierarchy has at its lowest levels the transduction of information by cells in the eye and at its highest levels neural processing in the cerebral cortex. There is much else going on between eye and cortex and we can spell this out in further anatomical detail. The flow of information begins at the rods and cones in the retina and the retinal ganglia. From there, information is sent along the optic nerve to the bilateral sub-cortical structures known as the lateral geniculate nuclei (LGN). Next, information is sent to the first stage of cortical processing in the back of the brain in the occipital lobes (in area V1). Next information is sent to further areas in the cortex along two different routs (Milner and Goodale, 1995). One route is from the occipital lobes to the parietal lobes. The second is from occipital lobes to temporal lobes. From there information is sent as far as the frontal lobes (Olson, Gettner, and Tremblay,1999) and the hippocampus (Milner and Goodale, 1995). It should be further noted that not only does information flow from the lowest levels to the highest, but there are also "back projections" along which information flows from highest levels back down to lowest (Pascual-Leone and Walsh, 2001).

One particularly important thing to note about the different levels of the visual processing hierarchy is the different kinds of representations that occur at the different levels. In brief, the key feature is that the very lowest levels are highly specific and egocentric (such as representations in LGN and V1) whereas higher levels are increasingly abstract and conceptual (such as representations in hippocampus and frontal cortex). To appreciate this difference with a relatively quick example, consider the following. Suppose there was a neuron that responded to the presence of a coffee mug, but only if that coffee mug was presented with a particular orientation and location in the visual field. This neural activation would be a representation relatively more egocentric and less abstract than the activation of a neuron in response to the presence of a coffee mug regardless of its orientation and position in the visual field. Among representations that we can regard as egocentric, there are varying degrees of specificity or egocentricity. A very egocentric representation of a stimulus would respond only if the stimulus was present in a particular location defined relative to the retina. This would be a representation of a location in retinocentric space. Less specific, and thus less egocentric, would be a representation of a stimulus that doesn't discriminate between stimuli that are presented to different parts of the retina, but does respond when the stimulus is present in a location defined relative to the head. While the coffee mug example is perhaps fanciful, there is little controversy over the claims that (1) stimuli are represented by cells with retinocentric receptive fields in retinal ganglia and LGN (Hubel and Wiesel, 2001), (2) there are body-centered representations that abstract away from retinal locations in areas such as posterior parietal area 7a (Andersen, 1995), and (3) there are representations of spatial locations in hippocampus that abstract away from the orientation of the organism with respect to the larger environment (Taube, Muller, and Ranck, 1990).

In Mandik (2001) I describe how all kinds of egocentric representation involve the representation of properties defined relative to the representing subject and how the different kinds of egocentric representation (retinocentric, body-centered, etc.) may be

distinguished in terms of how much they leave out or abstract away from the details of the representing subject. I also spell out how the resultant account of egocentricity need not apply only to spatial representation but may be generalized to, for example, representations of times and temperatures. This sketch of the processing hierarchy gives us several kinds of egocentricity, but which kind is the egocentricity of consciousness? To answer this, we must say more about consciousness.

§5 The Neural Accomplishment of Condition K

The question naturally arises of how to relate consciousness to the visual processing hierarchy and before we address that topic head on, it will be useful to look at a particular kind of example of conscious visual states. One particularly useful kind of example to look at concerns the phenomenon of motion induced blindness (Bonneh, Cooperman, and Sagi, 2001). Motion induced blindness doesn't involve any kind of brain damage or malfunction and is a relatively easy phenomenon to induce in normally sighted subjects. In a typical experiment, a subject stares at a computer screen that has stationary yellow dots on a black background. Behind the stationary yellow dots is a collection of moving blue dots. After a short while, it appears as if one or more yellow dots disappear. However, this is merely an appearance—in reality the yellow dots remain on the screen for the entire interval. One hypothesis that is relatively easily ruled out is that during these moments of blindness the information about the yellow dots isn't getting into the nervous system at all. To the contrary, however, the information seems to be making it up pretty high in the visual processing hierarchy. It gets up as high as cortical areas including parietal cortex (for further discussion, see Mandik 2005). The key here is that motion-induced blindness isn't due to a failure of the nervous system to represent the presence of yellow dots. It is instead due to a failure of the nervous system to represent the presence of yellow dots *consciously*. Soon I will say some more about what the relevant kinds of representation are in the processing hierarchy, but first I must note three points about how motion-induced blindness relates to relevant notions of consciousness. First, it is clear in these cases is that at one moment the subject is conscious of the yellow dot and at another moment she is not. Second, in conditions in which the subject is conscious of the yellow dot, she has a representation of a yellow dot that is a conscious representation, and in conditions in which the subject is not conscious of a yellow dot, the presence of a yellow dot is still represented in the nervous system, albeit unconsciously. Third, we must address the most interesting (to philosophers) aspect of these conscious mental states, namely their qualia, phenomenal character, or properties in virtue of which there is something it is like to have them. While much more can be said about this, for now I will simply suggest that what it is like is like seeing a yellow dot and this can be accounted for by the representation of the presence of a yellow dot.

Given this quick sketch of the visual processing hierarchy I want to turn now to where in the processing hierarchy to locate conscious representations. We can arrive at the view that conscious representations are to be found at an intermediate level by considering reasons for rejecting locating conscious representations at either the highest or lowest extremes. To see why egocentric representations at the lowest levels don't suffice for consciousness, we might do well enough to consider that no one takes seriously the suggestion that retinal ganglion activity suffices for conscious experiences.

But we can add that even egocentric representations slightly higher up in the hierarchy are insufficient for conscious visual states. So, for example, Milner and Goodale (1995) describe a patient, DF, who suffered bilateral damage to lateral occipital cortex resulting in visual form agnosia—an inability to consciously perceive the shapes of objects. DF can nonetheless respond to visual information about form and orient her hand appropriately to put a card in a slot despite reporting that she cannot see the slot or its orientation. Further, she cannot report the slot's orientation. It seem that in spite of this failure to consciously perceive form, DF is utilizing egocentric representations to orient her hand appropriately. Just as it is inappropriate to locate consciousness at the lowest end of the visual processing hierarchy, so is it inappropriate to locate it at the highest end. For examples of representations at the highest part of the hierarchy we can consider instances of relatively abstract categorical knowledge, like your knowledge that all mammals are warm-blooded. This is a piece of information that you have known and thus represented for a long time, but it is highly unlikely that you were doing so consciously the entire time. It is likely that until reading the previous sentence this is the first time in a while that the thought occurred to you consciously. What do these various examples establish? At best, they show that states at the far end of the hierarchy can sometimes be unconscious, not that they never can be. Nonetheless, when we examine conscious states we notice something important, namely that they plausibly exist in the middle of the hierarchy and thus combine aspects of higher- and lower-level representations. Consider, as a typical example of a conscious visual state, the visual perception of a coffee mug. You see the coffee mug from a particular point of view with a particular orientation in your visual field and thus does your conscious experience involve egocentric representations. However, you further are able to bring to bear conceptual knowledge in your experience: you see it as a coffee mug and as a concave item that holds beverages. Seeing it as such involves representations from higher in the hierarchy.

These remarks about the conceptual and egocentric aspects of conscious states are observations made "from the inside," as it were. Viewed from the outside, conscious states consist in mutually interacting representations positioned at slightly different levels in the middle of the processing hierarchy. In other words, activations of representations in a solely feed-forward manner do not suffice for consciousness, but when there is feedback from the higher to the lower levels, then the representations involved constitute conscious states. One line of evidence for such a view comes from experiments involving trans-cranial magnetic stimulation conducted by Pascual-Leone and Walsh (2001). By utilizing precisely timed magnetic stimulation of certain cortical regions they were able to allow the flow of information up from V1 to V5 and prevent the normal feedback of information from V5 down to V1. V5 is an area associated with the conscious perception of motion, but results suggest that motion was consciously perceived on occasions of V5 activation only when feedback to V1 was allowed. Further evidence for the need of reciprocal interaction between higher and lower levels comes from Lamme et al. (1998), who suggest that the responses elicited by stimuli in anesthetized animals constitute merely feed-forward activation of representations in perceptual networks and lack feed-back activations from representations higher in the processing hierarchy. The resultant view of the neural basis of consciousness is what I have referred to elsewhere as the Allocentric Egocentric-Interface theory of consciousness:

> [C]onscious states are hybrid states that involve the reciprocal interaction between relatively allocentric and relatively egocentric states: a conscious state is composed of a pair of representations interacting at the Allocentric-Egocentric Interface. Unconscious mental states are states that are either too high up or too low down in the hierarchy or are not engaged in the requisite reciprocal interactions. What a person is conscious of is determined by what the contributing allocentric and egocentric representations are representations of. The phenomenal character of these states is identical to the representational content of the reciprocally interacting egocentric and allocentric representations. (pp. 463-464).

For the present purposes we can regard the allocentric representations as conceptual representations. The lower of the two states in the hybrid provides the egocentricity of consciousness. The higher of the two provides the conceptual content of consciousness. We thus have a sketch of theory that spells out how a consciousness can reduce to physiological processes in a way that answers to the demands of the transcendental argument.

§ 6 Putting it All Together: The Reality of Appearance.

It will be useful to take stock and summarize what has gone on so far. First we examined Clark's arguments that if a creature was able to non-inferentially discriminate its own sensory modalities, it must do so in virtue of there being something it is like for the creature to be in those sensory states. I criticized Clark for insufficiently building a bridge from what a subject knows of its own states to there being something it is like to be in those states. I praised Clark for lighting the way toward a potentially successful project, one in which one may argue transcendentally for a reductive theory of consciousness. I sketched how such an argument might go, starting with considerations that each of us knows that he or she is not a zombie and leading to the considerations that the structure of conscious experience must be both conceptual and egocentric. I then spelled out how the conceptual and egocentric criteria can be satisfied by a neurophysiological theory of consciousness whereby conscious states are hybrids of mutually interacting representations activated at different levels of a sensory processing hierarchy. What remains open is to solve the sort of problem that befell Clark's attempt at an epistemological theory of consciousness, namely to spell out a clear connection between the relevant epistemic notions and the notion of what it is like to be in a conscious state. It is to this last step that I turn. Crucial in what follows is the equating of what it is like with how things appear with how things are represented by states with conceptual contents.

To see how the relevant notions of appearance figure into the above account of consciousness, it will be useful to spell out how the above view of consciousness relates to perception and introspection, for perception and introspection are faculties by which things appear to us. There are ways things seem when we perceive them and there are ways our mental states seem when we introspect them. The account of perception and introspection I favor is one developed by Churchland (1979) and one I've elaborated elsewhere (Mandik 2006). The view of perception at play here is that "perception consists in the conceptual exploitation of the natural information contained in our sensations or

sensory states" (Churchland , 1979 , p. 7). The view of introspection is analogous: introspection of sensations is the conceptual exploitation of natural information that our sensations contain about themselves. The notion of the conceptual exploitation of information contained in sensations can be spelled out, following Churchland, in terms of two senses in which a sensation may have intentionality, that is, two senses in which a sensation may be a sensation *of* something. These two senses of "sensation of" are an objective sense and a subjective sense. Adapting Churchland's formulations (1979, p. 14) yields:

> A given (kind of) sensation one has is a sensation of X (in the objective sense of "of") if and only if under normal conditions, sensations of that kind occur in one only if something in one's perceptual environment is indeed X.
>
> A given (kind of) sensation one has is a sensation of X (in the subjective sense of "of") if and only if under normal conditions, one's characteristic non-inferential response to any sensation of that kind is some judgment to the effect that something or other is X.

These notions allow for the explication of many important features about perception, in particular: (1) the distinction between what *can* be perceived and what actually is perceived and (2) the distinction between what is perceived and what is inferred. What can be perceived is determined by the objective intentionality of sensations. My vision is quite poor without my contact lenses and my inability to see the small print at the bottom of an eye chart is due to the lack of information present at my irradiated retina. Wearing my contacts increases the information my sensory states carry about distal objects and thus increases the number of things I can perceive. This does not however, alone suffice for what I actually do perceive. I may fail to perceive some distant object not because my eyesight is insufficiently acute but because I simply have not noticed it. What I do perceive depends on what concepts are brought to bear in my non-inferentially elicited judgments about the causes of my sensations. Thus, if my contacts are on, I'm looking in the right direction, and I have the concept of an insect, then I am in a position to actually perceive some small insect flying through my line of sight.

Regarding the distinction between what is perceived and what is inferred, consider the following example adapted from Mandik (2006). Jones and Smith both witness a man in a realistic gorilla suit perform a realistic imitation of a gorilla. Both suit and performance are convincing to the untrained eye and Smith, having an untrained eye, is initially convinced. Jones, in contrast, is a special-effects expert and thus is not fooled. Jones can see that this is a man in a costume. Suppose that at some later point Jones tells Smith that this is merely a man in a suit. Smith trusts Jones and believes him. Nonetheless, Smith cannot shake the impression that this is a real gorilla. What is going on with Jones and Smith? Suppose that they both have equally acute eyesight and in a sense, then, see the same things insofar as they have visual sensations with the same objective intentionality. Jones and Smith differ, however, in that Jones is, in virtue of his expertise in special effects, able to automatically apply the concept of a man in response to his sensations of the man in the suit. Smith, in contrast, is able to apply the concept of

a man only as the consequence of an inference and further is having difficulty overcoming his tendency to automatically apply the concept of a gorilla.

With this sketch of perception in hand, we can go on to sketch an account of introspection. Focusing on the introspection of sensations, introspection is analogous to perception in that each involves the automatic application of concepts in judgments elicited by sensations. Introspection differs from perception in the concepts that are brought to bear and the information thereby exploited. Perception involves concepts of external world objects and properties for the exploitation of information that sensations carry about those external world objects and properties. The introspection of sensation involves concepts about sensations for the exploitation of information that sensations carry about themselves.

Sensations, despite *being* brain states seldom *seem* like brain states, and this is due in large part to the facts that (1) most people lack the requisite neuroscientific concepts and (2) fewer still have sufficient training to *automatically* apply neuroscientific concepts in response to their own sensations. But as argued in Churchland (1979) and Mandik (2006), the barriers to introspecting ones own brain states as such are not insurmountable. But this is beside the crucial point here. More important for present purposes is how the above accounts for the ways things seem in perception and introspection, for herein lies an explanation of knowledge of what its like to have conscious experiences.

The account of perception and introspection supplies a means for distinguishing both what is perceived and what is introspected from what is inferred. Combining this account of perception and introspection with the account of consciousness spelled out earlier provides a means for distinguishing conscious from unconscious perceptions. With these various distinctions in place, we are in a position to give an account of appearance that can achieve what Clark could not: an explanation of why the ability to know the difference between various states entails that there is something it is like to be in those states (in a way that isn't simply an instance of the "trivial inference" mentioned in §1).

The strategy in what follows will be two-fold. First, I will discus the notion of appearance that needs explaining: a phenomenal notion of appearance or ways things seem. Second, I will look at the minimal account of appearance that comes along with the mere existence of conceptual states, an epistemic notion of appearance. Then I will discuss how, building upon an epistemic notion of appearance, we are able to recover the requisite notion of phenomenal appearance. To appreciate the notion of appearance in need of explaining, consider the following scenario.

The Blue Dog Scenario:

> Smith and Jones see a dog that is in fact white but due to a trick of the electric lighting, seems blue. Smith is unaware of the facts about the lighting and so believes that the dog is blue. Jones knows about the lighting situation and so believes the dog is white. Jones would agree, though, that in spite of his believing it to be white, the dog seems blue.

What is going on in the minds of Smith and Jones that constitute the ways things appear to them? We will return to this question after the consideration of a different scenario.

The Monty Hall Scenario:

> Smith and Jones are playing *Let's Make a Deal* with Monty Hall. There are three doors for Smith and three for Jones. Behind one of Smith's doors is a car. Likewise for Jones. They each pick their door number one. Before door number one is opened, Monty Hall opens door number three and reveals that there is a goat behind it. Monty asks if they'd like to keep door number one or switch to door number two. Smith figures there is a fifty/fifty chance that the car is behind door number one, so he believes door number two to not be a superior choice. Jones knows the explanation of the relevant probabilities and so believes correctly that there is an advantage in switching. Jones admits, though, that while he trusts the explanation, he doesn't totally understand it, and sympathizes with Smith's urge to not switch.

What is going on in the minds of Smith and Jones in the Monty Hall Scenario that constitutes the ways things seem to them relevant to their decisions in the game? Here it looks like a purely *epistemic* notion of appearance can do the work. Things seem to Smith and Jones to be such-and-such in so far as they apply various concepts in their judgments that things are such-and-such. We may need, of course, to appeal to various judgments and further, various dispositions of varying strength toward distinct judgments, but none of this obviously takes us out of the realm of epistemic appearance. So, for example, here is a straightforward and uncontroversial explanation of what is going on. Smith has a disposition to judge door number two to not be a superior choice and is aware of no overriding considerations against resisting his disposition. Jones similarly has a disposition to judge door number two to not be a superior choice, but is aware of overriding considerations in favor of resisting this disposition, so he resists. He believes door two to be superior but agrees that it seems not to be superior. In what does this latter seeming consist? It consists in his overridden disposition to make a certain judgment.

What, then, should we say of the Blue Dog Scenario? It is worth noting, first, just how much mileage we can get out of an explanation constructed to be analogous to the explanation of the Monty Hall Scenario. Smith has a disposition to judge the dog to be blue and is aware of no overriding considerations against resisting this disposition. Jones similarly has a disposition to judge the dog to be blue, but is aware of overriding considerations in favor of resisting this disposition, so he resists. He believes the dog to be white but agrees that it seems to be blue. In what does this latter seeming consist? It consists in his overridden disposition to make a certain judgment.

One might object at this point that the epistemic appearances appealed to in the explanation of the Blue Dog Scenario are *mere* epistemic appearances, that is, phenomenal appearances have not yet been taken into account. While there is some truth to this, the problem is not so much with the explanation of the scenario, but with the scenario itself. Note that nothing going on in the scenario has essentially to do with consciousness. However, we can modify the scenario slightly to add that the relevant perceptions are conscious perceptions, that is, that Smith and Jones are have conscious visual experiences of the blueness of the dog. Adding consciousness to the scenario requires that we add appeal to a theory of consciousness in an explanation of the relevant notions of appearance.

So-called phenomenal appearances are reducible to a sub-class of epistemic appearances. There's nothing going on in the mind in these scenarios that can't be

explained in terms of information bearing states (the sensations) and our conceptual reactions to them (the judgments). So, what are qualia? They are introspectible properties of conscious states, where what is introspectible is in part determined by what is there to be introspected and in part determined by what concepts a subject is able to automatically bring to bear in the judgments elicited by the sensory states. Conscious states are hybrid states of mutually causally interacting judgments and sensations.

It might be objected at this point that the application of concepts is unnecessary, that whatever phenomenal consciousness consists in, it is the sort of thing that may exist independently of our conceptual states. However, this objection garners no support from either first-person or third-person views on consciousness, and we can develop this point along two lines of thought. On the first line of thought, we look to sensory processing hierarchies and note, as was noted before, that pre-conceptual states are insufficient for consciousness. The states at the lowest levels of sensory processing hierarchies are states that act as detectors of stimuli in egocentric space, and as such are essentially no more complex than measuring devices such as thermometers, devices that are themselves devoid of phenomenal consciousness.

The second line of thought against this objection points out that alleged counter examples to the view that conscious states are conceptual-states would be inaccessible from the first-person point of view. Insofar as what is seen can be known, then visual experience has conceptual content. If visual experiences have distinctive contents, then there is no need to postulate a distinctive attitude to account for what is distinctive about experience. If visual experiences have conceptual contents then what is seen can be believed. If what is seen can be believed and there is such a thing as a counter-example to the claim that conscious states are conceptual states, then you would not be able to tell the difference between you and a being that had all of the same beliefs but differed in what visual experiences it had (or even in whether it had visual experiences). You would not know whether you were a counterexample. But this is absurd.

Let us return to the kind of task that Clark discussed in his argument that a certain case of access consciousness entails phenomenal consciousness. Suppose that Jones knows, by introspection of his perceptual experience of the dog, that he is arriving at the judgment that the dog is blue by seeing a blue dog (as opposed to, say, being told that there is a blue dog in the vicinity). In such a case, Jones is having a conscious visual experience of a dog as being blue insofar as (1) Jones has a sensory state carrying information that there is a blue dog (2), Jones judges that there is a blue dog, and (3) there is reciprocal causal interaction between Jones' sensation and Jones' judgment. The appearance to Jones that there is a blue dog is not a *mere* epistemic appearance because the judgment in question is automatically elicited by the sensation. When Jones introspects, again the resultant judgment is not a *mere* epistemic appearance because his judgment that he is having a sensation of blue is automatically elicited by the sensation. I close with one final question and one final answer. Question: How do we know that Jones is not a zombie? Answer: Because we know that we are not zombies and we know that we are just like Jones. Thus has an epistemological theory of consciousness lead us to a zombie-free zone where one and the same theory accounts for what consciousness is and how consciousness is known.

Acknowledgments

This work was supported in part by grants from The National Endowment of the Humanities and The James S. McDonnell Foundation's Project for Philosophy and the Neurosciences. I am grateful to audiences of versions of this material at the McDonnell Project "Neurophilosophy: The State of the Art" conference at the California Institute of Technology; the City University of New York Graduate Center Cognitive Science Symposium and Discussion Group; and the 2006 "Toward a Science of Consciousness" meeting in Tucson, Arizona. I am especially grateful for discussions of this and related material with the following individuals: Jared Blank, David Chalmers, Ron Chrisley, Andy Clark, Tanasije Gjorgoski, Uriah Kriegel, Clayton Littlejohn, Doug Meehan, Phillip Pettit, Jesse Prinz, David Rosenthal, Eric Schwitzgebel, Anders Weinstein, Josh Weisberg, Chase Wrenn, and Tad Zawidzki.

References

Andersen, R. (1995). Coordinate transformations and motor planning in posterior parietal cortex, in M. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge, MA, 519-532.

Block, N. (1978). Troubles with functionalism. In *Perception and Cognition: Issues in the Foundations of Psychology*, ed. C. Savage, Minneapolis: University of Minnesota Press. 261-326

Block, N. (1995). On a confusion about a function of consciousness, *Behavioral and Brain Sciences* **18** (2), 227-288.

Bonneh, Y., Cooperman A., and Sagi, D. (2001). Motion induced blindness in normal observers, *Nature*, **411**(6839), 798-801.

Chalmers, D. (1996). *The Conscious Mind*, Oxford University Press, New York.

Churchland, P. (1979). *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge.

Clark, A. (2000a). A case where access implies qualia? *Analysis* 60: 1: 30-37.

Clark, A. (2000b). Phenomenal immediacy and the doors of sensation. *Journal of Consciousness Studies* 7: 4: 21-24.

Dennett, D. (1991). *Consciousness Explained*. New York: Little Brown & Co.

Dennett, D. (1995). The path not taken. *Behavioral and Brain Sciences* 18: 2: 252-53.

Hubel, D. and Wiesel, T. (2001) Brain mechanisms of vision in Bechtel, W., Mandik, P., Mundale, J., and Stufflebeam, R. (eds.), *Philosophy and the Neurosciences: A Reader,* Basil Blackwell, Oxford, pp. 179-198.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32: 127-36.

Lamme, V. A. F., et al. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. Current Opinion in Neurobiology, 8, 529 – 535.

Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA, MIT Press.

Mandik, P. (1999). Qualia, space, and control. *Philosophical Psychology* 12: 1: 47-60.

Mandik, P. (2001). Mental representation and the subjectivity of consciousness, *Philosophical Psychology* **14** (2), 179-202.

Mandik, P. (2005). Phenomenal consciousness and the allocentric- egocentric interface. In: R. Buccheri et al. (eds.); Endophysics, Time, Quantum and the Subjective World Scientific Publishing Co.

Mandik, P. (2006). The introspectability of brain states as such. In Brian Keeley, (ed.) Paul M. Churchland: Contemporary Philosophy in Focus. Cambridge: Cambridge University Press.

Milner, A. and Goodale, M. (1995) *The Visual Brain in Action.* Oxford University Press, New York.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review 83*, 435-50.

Olson, C., Gettner, S., and Tremblay, L. (1999). Representation of allocentric space in the monkey frontal lobe, in N. Burgess, K. Jeffery, and J. O'Keefe (eds.) *The Hippocampal and Parietal Foundations of Spatial Cognition*, Oxford University Press, New York, pp. 359-380.

Pascual-Leone, A. and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness, *Science,* **292**, 510–512.

Robinson. W. (1982). Causation, sensations and knowledge. *Mind* 91: 364: 524-40.

Stone, J. (2001). What is it like to have an unconscious mental state? *Philosophical Studies* 104: 179-202.

Taube, J., Muller, R., and Ranck, J. (1990). Head direction cells recorded from the postsubiculum in freely moving rats. Description and qualitative analysis. *Journal of Neurosciences,* **10**, 420-435.

Weiskrantz, L. (1996). Blindsight revisited, *Curr. Opin. Neurobiol*, **6**(2), 215-220.