Jesse Prinz

Mental Pointing

Phenomenal Knowledge Without Concepts

It is one thing to have phenomenal states and another thing to think about phenomenal states. Thinking about phenomenal states gives us knowledge that we have them and knowledge of what they are like. But how do we think about phenomenal states? These days, the most popular answer is that we use phenomenal concepts. Phenomenal concepts are presumed to be concepts that represent phenomenal states in a special, intrinsically phenomenal, way. The special nature of phenomenal concepts is said to be important for defending materialism against epistemic arguments for dualism. In this paper I present an account of phenomenal knowledge that does not depend on phenomenal concepts. In fact, I argue that we have no phenomenal concepts. Instead my account appeals to mental pointing, a process that I explain in terms of phenomenal demonstratives. Phenomenal demonstratives are sometimes referred to as concepts in the literature, but I suggest that this is a mistake. I also present a theory of phenomenal demonstratives that equates them with attentional control structures in working memory. In a concluding section I describe how this theory can be used to defuse the knowledge argument for dualism. That is only a subsidiary goal, and my response to the knowledge argument echoes others in the literature. I think the project of developing a substantive, empirically informed theory of phenomenal knowledge has interest independent of debates about mental ontology. That is my central focus. Thinking about phenomenal knowledge can shed light on the relationship between consciousness, attention and memory. This paper has a philosophical agenda and an empirical agenda. Those who reject my philosophical claims about the nonexistence of phenomenal concepts, the conditions

Correspondence:
Department of Philosophy, CB # 3125, Caldwell Hall, University of North Carolina, Chapel Hill, NC 27599-3125, USA. *jesse@subcortex.com*

Journal of Consciousness Studies, **14**, No. ??, 2007, pp. ??–??

required for phenomenal knowledge, and the truth of physicalism, could accept some of my empirical claims about the neurofunctional correlates of consciousness and the resources available for accessing phenomenal states. Of course, I think the empirical claims support my philosophical conclusions.

1. The AIR Theory of Consciousness

1.1. Outline of the theory

Any account of how we think about our phenomenal states must depend on the nature of those states. Therefore, this discussion cannot get off the ground without a theory of consciousness. I cannot defend such a theory here, but I will summarize a theory that I have defended elsewhere (Prinz, 2000; 2001; 2005; forthcoming). The account of phenomenal knowledge will draw on ideas from this theory, but is compatible with other approaches to consciousness as well.

The theory I endorse has two components. First, there is an account of the contents of consciousness. I believe that all consciousness is perceptual, which is to say that consciousness attends only representations in sensory systems (Prinz, 2007). I also believe that consciousness arises only at an intermediate stage of perceptual processing (Jackendoff, 1987). Sensory systems are organized hierarchically, and there is a progression from representations of very local features of a stimulus all the way up to highly abstracted categorical representations that capture features of stimuli that are comparatively invariant across perspectives. In between are representations that are coherent rather than local and vantage point specific rather than invariant. Following Jackendoff, I locate consciousness here. Activity in the neural correlates of intermediate-level areas correlates with conscious experience, and cells in these areas have contents that agree with experience. Low-level representations are more fragmented than experience, and high-level representations are too abstract. Consider the cells that are used to recognize something as a face. The cells are typically invariant across a range of viewing angles, which is to say the same cell fires if the face is seen straight on or nearly in profile. In sharp contrast, a phenomenal experience of a face is orientation-specific. I do not want to deny that high-level representations can have conscious effects. For example, they often result in conscious verbal imagery (‘there’s a chair’), conscious action tendencies (the preparation to sit) and shifts in attention. The last of these will be important in the discussion below.

The intermediate-level hypothesis is a theory of the contents of consciousness. But we also need a theory of how these contents become conscious. Mere activation of intermediate-level perceptual states is not sufficient for consciousness. In subliminal perception, there is activation throughout the perceptual hierarchy without any corresponding experience (e.g. Moutoussis and Zeki, 2002). Therefore consciousness requires not mere intermediate-level activation, but activation of a particular kind. But what kind? The second part of my account extends Jackendoff's theory by addressing this question. I am persuaded that consciousness arises when and only when we are paying attention. Consider research on inattentional blindness. In these studies, subjects are given a task that demands considerable attention, such as comparing the length of two similar lines or counting how many times moving objects collide, and while they are engaged, an unexpected stimulus is presented in plain view (Mack and Rock, 1998; Most *et al*., 2005). Many subjects fail to notice the unexpected stimulus under these conditions, even though the stimulus would be readily perceived under conditions that were less attentionally demanding. Or consider the attentional blink (Vogel and Luck, 2002). In these studies subjects are presented with a sequence of stimuli and are asked to look for two target stimuli in the sequence; if the second target appears shortly after the first, many subjects fail to perceive it, because the first target has temporarily consumed their attention. Or consider visual neglect. In this disorder, brain injuries in structures that control attention result in blindness for part of the visual field. All these phenomena suggest that attention is necessary for consciousness. Putting this conclusion together with the conjecture about the content of consciousness, we are left with the following theory: phenomenal states are attended intermediate-level representations, or AIRs.

To avoid circularity, the AIR theory must incorporate an account of attention that is not defined in terms of consciousness. Ultimately, that theory might be specified in precise neurocomputational terms. Short of that goal, we can give an analysis of attention in psychological terms and then relate the psychology to gross neuroanatomy. On the view I favour, attention is a process that allows information to flow from perceptual systems to systems involved in working memory. Working memory is actually an umbrella term for a number of different capacities: the capacity to retain information, to transform perceptual states imaginatively, to select behavioural responses and to carry out various executive responses, such as object tracking, comparison and verbal reporting. In a popular phrase, working memory allows for

maintenance and manipulation. When we attend, aspects of what we perceive become available thereby to be maintained for brief periods and manipulated. If this is what attention does and consciousness requires attention, then consciousness arises when perceptual states send signals to working memory. Anatomically, attention is associated with structures in parietal cortex, and working memory is associated with lateral frontal cortex. Intermediate-level visual perception is located in temporal cortex. So visual consciousness, on this view, involves a circuit that includes temporal, parietal and frontal cortices. Other senses may involve other structures, but the basic principles will be the same: consciousness arises when perception plus attention allows working memory access.

1.2. Encoding or availability?

This formulation of the theory raises a question. The phrase 'working memory access' is ambiguous. It can refer to the process by which information in perceptual systems becomes accessible or it can refer to the processes by which working memory systems actually access perceptual states. Metaphorically, we can talk of broadcasting or receiving. Presumably these processes are distinct. Attention leads to specific changes in how perceptual states are processed and those changes allow perceptual states to send information forward to working memory systems, which then receive, or 'encode', the states. So the question is, does consciousness arise when perceptual systems broadcast or only when working memory systems receive the broadcast? I want to suggest three lines of evidence for thinking broadcasting is sufficient.

First, consider change blindness. When presented with a change-blindness display, subjects actively scan every corner of the image in an effort to detect the change. Presumably, this process involves the allocation of attention and whatever we attend to becomes available to working memory. A person staring at a change-blindness display could report on any feature that she happens to be staring at. But there is reason to think that the features that are available to working memory are not actually getting recorded in working memory, because, if they were, the changes would be easier to detect. This interpretation gets some support from neuroimaging. In a change blindness task, Beck *et al*. (2001) found greater activation in working memory areas when the change was detected as compared to when a change went undetected. This is not to say that there is no working memory activation when no change is detected. Working memory seems to

automatically store the gist of an image, even when details are lost. Dramatic changes affecting the significance of the display (such as the gender of a person we are looking at) are typically noticed (Simons and Levin, 1997). Moreover, we can, by act of control, focus on any given detail of a display and retain it from one moment to the next, but this takes effort, and we don't know antecedently what feature to focus on. When trying to find a change, we might arbitrarily focus on one feature after another, until we hit on the one that changed. Once we find it, the change is easy to detect. But, while searching for the change, the majority of details in a display are not retained, even though they can be plainly seen. This suggests a difference between two kinds of processes: attention without retention and attention with retention. That distinction can be explained by presuming that attention makes each part of the image available to working memory, as we scan it, but working memory receives only select portions of that signal. We scan strategically for bits that we think might be likely to change, and then when, by arbitrary chance, our eyes cross such a bit, we maintain a representation of it in memory.

The second reason for thinking that there can be consciousness without encoding in working memory comes from studies in which stimuli are presented very briefly. If a stimulus is very briefly displayed (e.g. 16 milliseconds) and followed by a mask, it will not be seen at all. Subjects will have no idea that something was presented. If there are some trials in a study in which a stimulus is presented for 16 milliseconds, and others in which a blank screen is presented, subjects will be at chance in guessing whether a stimulus was presented, even though 16 milliseconds is long enough for priming to take place (e.g. Dehaene *et al*., 1998). But now suppose the stimulus is presented for a longer duration, say 50 milliseconds. Then, depending on the stimulus, subjects may know that *something* was flashed but they won't know what it was. In other words, as we approach the threshold for conscious recognition in priming studies, subjects report a phenomenal experience, but they cannot describe it. Importantly, it's not merely that subjects fail to identify the stimulus; they don't even retain information about it's shape (if they did, recognition would be possible). This is comparable to the experience of participating in the widely discussed experiments of Sperling (1960). Sperling found that, when a grid of letters is presented for 50 milliseconds without a mask, subjects can report the numbers in one row, but not the others. Indeed subjects can be cued to recall *any* row. The other numbers are experienced but not identified. Sperling's study was important because it proved the existence of iconic memory: a very brief period in which a perceptual

stimulus is retained after offset. Iconic memory is different from working memory: it is really a perceptual after-effect, caused presumably by the slow delay rate of visual states. It does not require encoding or permit retrieval. So we should not infer from Sperling's conclusion about iconic memory that consciousness involves encoding in working memory. Indeed, I interpret these results as evidence for the opposite conclusion. I would say that all the numbers are *accessible* to working memory, but only some are *accessed* by working memory. The Sperling case and near-threshold priming provide an interesting contrast to inattentional blindness and very fast priming. In the latter two cases, subjects have no awareness of the stimulus; they have no idea that something was displayed. In the Sperling case and the near-threshold case, subjects experience something, but they do not retain its specific properties. This suggests a three-way distinction between being encoded in working memory (the cued numbers in a Sperling display), merely being broadcast to working memory, and not being broadcast at all.

The third line of evidence against the necessity of working memory encoding may be the most important of all. There is good reason to think that working memory *cannot* encode perceptual states at the same level of resolution at which they are consciously experienced. We experience the world in incredible detail. Subtly different colours, for example, can be distinguished. Yet such fine details are lost when we try to retain memories of what we have just perceived over temporal intervals. Imagine being presented with a blue colour patch and then having to select that very same blue five seconds later from an array of three similar shades. You won't be able to do it. Accuracy is nearly perfect for simultaneous side-by-side colour matching, but highly inaccurate with a delay (e.g. Pérez-Carpinell *et al*., 1998). This suggests that working memory stores colours (and other visual features) in a code that abstracts away from their precise details. In other words, it looks as if working memory uses something more like high-level visual representations. Indeed, there is evidence from cognitive neuroscience that working memory actually uses high-level perceptual representations rather than it's own proprietary codes. On this view, brain structures associated with working memory in lateral frontal cortex work in concert with structures in inferotemporal cortex during visual working memory tasks; frontal structures maintain representations in temporal structures (Postle, 2006). If this is right, then working memory encoding is a matter of maintaining activity in the higher levels of the sensory pathways. I already argued that high-level perceptual representations fall outside experience; they are more abstract

that anything we experience phenomenologically. This implies that consciousness cannot depend on encodings in working memory, because such encodings are far less detailed than the contents of experiences. It also follows that perceptual memories, both short-term and long-term, are stored in a code that lies outside consciousness. Of course, a perceptual memory can become conscious through the construction of a mental image. On my view, that happens when high-level representations are used to generate intermediate-level representations through efferent connections in the perceptual hierarchy. This story explains why mental images are often vague or imprecise. In imagery, we generate intermediate-level perceptual states from high-level records that have abstracted away crucial details. When we try to fill in missing details, we are forced to guess what they were. As a result, images can be inaccurate, unstable or sketchy.

Together the foregoing considerations support the contention that consciousness does not require encoding in working memory. I conclude that consciousness involves broadcasting to working memory rather than reception by working memory. The last consideration about the grain of representation raises a bit of a puzzle. If working memory cannot encode the details of experience, then what sense does it make to say that experience involves broadcasting to working memory. How and why would the visual system broadcast information that cannot be received? Isn't this like sending TV broadcasts into outer space? This puzzle threatens to undermine the theory of consciousness that I have been defending — a theory that says consciousness arises when intermediate-level visual states are broadcast to working memory.

I think the answer to this puzzle requires that we move away from the broadcasting metaphor and think about the relationship between consciousness and working memory in a slightly different way. The intermediate level of perceptual processing presents the world as a collection of features and objects, presented in space from a particular point of view. Working memory selects from this array, but to understand the selection process we need to distinguish two stages, each of which might be understood by a different metaphor. First of all, attention functions like a spotlight, illuminating some proper subset of the perceived features, objects and locations. These are the things that are consciously experienced. Then, working memory systems select from this subset. We can compare that selection process to an artist's schematic sketch of the objects on display. The sketch is not a faithful copy, but merely a rough approximation of items in the spotlight. The sketch is also selective, leaving out many things, and transient: the

artist will discard most sketches immediately, placing only a few into long-term storage. Like any metaphor, this one is far from perfect, but it improves over broadcasting in two crucial respects. First, it emphasizes a two-stage selection process: the spotlight and the sketch. Second, it emphasizes the loss of fidelity at the second stage.

2. Phenomenal Knowledge

2.1. Having, categorizing, noticing and pointing

The preceding discussion of consciousness will be helpful in addressing the main topic under investigation here: phenomenal knowledge. I use this term to designate knowledge of our conscious states as such. I equate consciousness with phenomenality (see Block, 2002). To be conscious is to feel like something, or to have 'phenomenal qualities'. Phenomenal knowledge is knowing what a state feels like. It is usually assumed that we have such knowledge. It seems obvious to many that if we know anything at all, we know what our conscious states are like. I want to problematize this assumption. I will not deny that we have phenomenal knowledge, but I will try to suggest that phenomenal knowledge is, in some respects, extremely limited.

Before offering an account of phenomenal knowledge, it is useful to distinguish several different relationships we can have with a phenomenal state. First of all, we can experience the state. Grammatically, this phrase is somewhat misleading. When we experience external things, such as earthquakes or sunsets, there is a clear distinction between the object of experience and the experience itself. But this distinction collapses when we experience our phenomenal states. When we experience a phenomenal state, there are not two things in play — the state and our experience of it — but rather one thing: the state being experienced. Experiencing a phenomenal state is simply being in that state. On the theory of consciousness that I endorse, an experienced state is just a perceptual state that is being processed in a way that makes it available to working memory. Experiencing a phenomenal state can be thought of as a relation, because there is an organism who has the experience, but it is not a representational or intentional relation: the organism is not representing the state.

Note that this last claim marks a sharp contrast between the view I defend and higher-order thought theories (HOT). I prefer the AIR theory to the HOT approach because I think it is supported by the evidence that I have summarized, and because I think higher-order thought theories erroneously assume that we have representations as finely detailed as the features available to us in consciousness. If we

had such representations of everything we can experience, as the HOT approach requires, then I don't see why we wouldn't be able to recall precise colours after brief delays. I also don't see why the nervous system would include a redundant system of representation. Obviously, it would take much more space to develop these objections in a compelling way. I mention these points of concern to indicate my own reasons for preferring AIRs to HOTs, and to underscore the difference between these approaches. This difference — the lack of metarepresentation in the AIR theory — will be important in what follows.

My claim about metarepresentation is that we cannot represent our phenomenal states using representations that have the same precise detail. In making this claim, I do not want to deny that we can represent our phenomenal states. We can, and often do, categorize our experiences. Categorization is the second relation to phenomenal states that I want to consider. If I am looking at the sky on a clear day, and having an experience of blue, I can place that experience under the category blue.

Categorization typically involves the deployment of concepts. As I will use the term, concepts are mental representations that have two important properties. First, concepts can be activated by the organism that possesses them, rather than being activated only by an external stimulus (Prinz, 2002). In Kantian jargon, concepts are capable of being used spontaneously, not just receptively. Second, concepts must be capable of being used to re-identify the things that they represent (Millikan, 2000). We must be able to re-deploy a concept on different occasions to keep track of things. It follows that concepts must be representations that we can store in memory. A representation that was not stored could not be re-tokened or actively used, outside stimulus control, by an organism. Notice that this requirement does not entail that concepts be amodal symbols. Perceptual states can be stored in memory and re-deployed by an organism. Stored perceptual states can be concepts (Prinz, 2002). I think that ordinary colour concepts are stored records of visual states. But, for reasons I have given, the only perceptual states that we store are high-level perceptual representations. The concept of blue is a high-level perceptual representation that could be activated by a range of spectral properties — the range we call 'blues'. We can store records and recognize relatively specific shades of blue as well (e.g. the blue used in the paintings of Yves Klein), but there are limits on this. High-level colour representations are invariant across small changes in hue and value. Retention and recognition of very precise shades may be impossible (unless, of course,

we measure the spectral properties of a sample that we are observing and store a verbal record of its precise scientific description).

In cases where we cannot categorize our phenomenal states, I think we can still notice them. Noticing is less demanding then categorizing, because we can notice things that we do not have the capacity to recall or re-identify. I think noticing is best analysed in terms of working-memory encoding. Something gets noticed if and when it gets encoded in working memory.

I want to distinguish noticing from another relationship that we can have with our phenomenal states. It seems that while we are having a phenomenal state, we can mentally point to it. When having an experience we can say, 'that's spicy' or 'that's blue' or 'that's what a D-minor sounds like'. In each of these cases, we are applying a concept (SPICY, BLUE, D-MINOR). These concepts may be stored high-level perceptual representations. The concept, spicy, for example, may be a high-level gustatory or, more likely, nocireceptive representation that has been stored on previous encounters with spicy food. But what is expressed by the word 'that' when we say 'that's spicy' or 'that's blue'? It seems that we are somehow able to point inwardly to an experience. The thought expressed by the sentence, 'that's blue' seems to have two components, one corresponding to the word 'blue' and the other corresponding to the word 'that'.

Unfortunately, not much is known empirically about what goes on when people point to their experiences in thought. There seem to be two possibilities. One possibility is that inner pointing does not involve anything above and beyond having an experience. On this view, when I say 'that's blue', the non-linguistic thought that this expresses contains no representations other than an experience of blue (an AIR) and a high-level perceptual representation (the concept BLUE). If so, the word 'that' doesn't really express any mental representation. It is just a way of verbally labelling the experience itself. Another possibility is that the word 'that' expresses an actual pointer in the mind. On this view, there are states that refer to phenomenal states by means of a relationship that is analogous to pointing. Some philosophers think about these internal states as demonstratives, but I will suggest in a moment that if mental pointers exist they differ from ordinary linguistic demonstratives in an important respect.

I am not aware of any decisive empirical evidence that can distinguish between the hypothesis that we point using inner pointers and the hypothesis that pointing to a mental state is a merely verbal act — an act of saying 'that' as we have an experience. There are, however, three related lines of phenomenological evidence that lead me to favour

the view that there are internal pointers. First, phenomenologically, there seems to be a difference between mentally pointing and having an experience accompanied by the word 'that'. Try to say 'that' while having an experience, and then try to mentally point, and see if you find a difference. Some people will undoubtedly be baffled by the instruction to mentally point. The exercise will not work for them. I find the idea of mentally pointing somewhat intuitive and it seems to be different from merely uttering the word 'that' while having an experience. Second, it seems to me that I can point in the absence of silent speech. For example, it seems to me I can have thoughts that I would express by saying 'now, that's delicious!' without actually saying that's delicious to myself. Third, it seems to me that when I do silently utter the word 'that' during an experience it's very clear to me which item of my experience I am ostending. While looking at a complex visual scene, I can shift the apparent reference of the linguistic demonstrative without changing my direction of gaze. Phenomenological arguments are tricky because not everyone reports the same phenomenology, but these considerations are collectively convincing to me.

If mental pointers are real, we need an account of what they are. The metaphor of mental pointing is a reasonable starting place, but we need to move from metaphor to mechanism. Let's consider the case of looking at a complex scene. It seems to be that there are two ways in which we can mentally point to an item in a scene. First. we can identify a particular region of space in which the item is located. Second, we can apply a perceptual concept. Perceptual concepts, I suggested, are stored records of high-level representations. If you are looking at a table setting, you might focus on the left and see what's there or you might look for the salad fork and focus on it, by using a high-level salad fork template. Notice that I used the word 'focus' in both cases. The phenomena I have just described is known in psychology as top-down selective attention. It is well established that top-down attention can be driven by object-representations or by spatial locations (e.g. Hayden and Gallant, 2005). I think mental pointing is achieved by top-down attention.

I have already argued that consciousness requires attention. To avoid confusion, I want to make it clear that not all attention is top down. When task demands do not require a strategic visual exploration of the world around us, attention may be applied quite diffusely, or we may scan, in a more or less bottom-up way from object to object in our surround. Attention can be captured by things we see, and we can attend in a way that does not single out any specific object or

region of space (consider the command 'pay attention!'). Top-down selective attention is a distinctive phenomenon. Attention is, on my view, just a process by which perceptual information becomes available to working memory, but that process can be controlled by a variety of different mechanisms. Top-down selective attention is attention that occurs under the control of object representations or spatial maps. I think this process captures the phenomenology of mental pointing very well. For example, when I silently utter the word 'that' while looking at a scene, it seems to apply to whatever object I have brought into focus by top-down control. When this happens, I think the object to which we are attending comes into sharper view, or higher resolution. There is evidence that when we attend, receptive fields in the visual system expand so that more cells than usual respond to the attended object (Olshausen *et al*., 1993). The attended object may also receive more attention then the surrounding space. Attention is not an all or nothing affair. I think it can come in degrees, and hence consciousness too, which depends on attention in grades. When we use top-down attention to single out an object, it becomes more conscious than the surround.

This account suggests a three-way distinction between kinds of conscious episodes. In Section 1, I distinguished cases in which a perceived stimulus gets encoded in working memory from cases in which it is merely available to working memory. Using the terminology introduced in this section, I label that distinction by the contrast between noticing something and merely experiencing it. I have also introduced a distinction between objects that are within the spotlight of attention and those to which we attend by an act of top-down control. We can label these by referring to experiencing and noticing on the one hand, and mentally pointing on the other. All these three conditions are states of consciousness and, in all three, consciousness arises via the same mechanism: AIRs. The differences have to do with what gets encoded outside consciousness, in working memory, and what unconscious mechanism-control attention. These unconscious differences can affect the degree and allocation of focal attention, as well as what information gets maintained or manipulated.

Let us assume that this account of mental pointing is correct. It is instructive to compare mental pointing to linguistic demonstratives, such as 'this' and 'that'. In his highly influential analysis, Kaplan (1989) argues that a demonstrative is a term that refers rigidly to something that is made salient in a context. The demonstrative itself has no descriptive content, so it cannot determine which object has been designated. But demonstratives are used in conjunction with

other representational resources, which Kaplan calls 'demonstrations' to reference. If you say 'that' while pointing, for example, the direction of your finger serves as the demonstration. Or one might refer to 'that man' in a room full of women, and the word 'man' will serve to make the man salient in that context, and the word 'that' will refer to him. Kaplan uses the schema 'dthat[]' to represent the structure of a demonstrative use of the word that, where the brackets get filled in by a demonstration. Mental pointers can be regarded as phenomenal demonstratives. They refer to the conscious perceptual states that are made salient by a mental demonstration. If the analysis I've offered is right, a mental demonstration is a high-level perceptual representation of a representation of a region in space. A phenomenal demonstrative has the structure 'pthat[]' where the brackets are filled in by a mental demonstration. If I mentally point to a salad fork, my phenomenal demonstrative refers to my conscious experience of the fork by using a schematic, high-level representation of a fork to draw attention to that part of the visual scene.

On this story, there are parallels between the way mental pointers work and the way linguistic demonstratives work. But there are also differences. In language, demonstratives are words. I don't think it's helpful to think of phenomenal demonstratives as mental words — symbols in a language of thought — much less as images (e.g. an image of a pointing finger). It's better to think of them as control structures. By this, I simply mean that they are things that have causal control over things. Phenomenal demonstratives use representations of objects in space to direct focal attention on a perceived scene. They are individuated by their causal powers.

In summary, there are at least three ways we can be related to our phenomenal states. We can experience them (or have them), we can categorize them, or we can point to them. In the next subsection, I will argue that these relations allow for phenomenal knowledge only in a very limited sense.

2.2. Knowing what it's like

Having distinguished experiencing, categorizing and pointing to phenomenal states, we can now ask which if any of these constitutes phenomenal knowledge. Recall that phenomenal knowledge is knowledge of what our phenomenal states are like. I will argue that the notion of phenomenal knowledge is actually more problematic than is sometimes appreciated.

Let's begin with experience. Does experience of a phenomenal state constitute knowledge of what that state is like? I don't think so. This may sound surprising, or even heretical. After all, it's trivially true that there is something it is like to have an experience, and it is often said that experiences give us direct knowledge of what phenomenal qualities are like. Indeed, traditional epistemologies, such as sense data theory, have been taken to imply that phenomenal qualities are the only thing we really know directly. I think that it's a mistake to describe experience as a kind of phenomenal knowledge. Experiences are like something, but they do not qualify as *knowledge* of what it's like. Knowing is a transitive relation. It is a relation we bear to something. When we talk about knowing something, we imply that there is both an object of our knowledge and an epistemic relation to that object. Now consider a phenomenal state. What is the object and what is the epistemic relation? What is it that we know when we have a phenomenal state? It is arguable that phenomenal experiences give us knowledge of what they represent. If I am experiencing a particular colour, my experience may be said to give me knowledge of that colour. But this is knowledge that I can have in the absence of consciousness. An unconscious colour representation gives me (unconscious) knowledge of that colour. Patently, this is not knowledge of what it's like to have an experience. So the question is, does a phenomenal state yield knowledge of *itself* in addition to knowledge of what it represents? Here, I want to suggest a negative answer. I don't think there is any *direct* entailment from having a phenomenal state to knowing what that state is like. In general, knowing some object *o* seems to require representing *o*. On the theory of consciousness I favour, unlike higher-order representation theories, one can have a conscious experience without representing it. If knowledge requires representation of the object of knowledge, then having an experience does not entail knowing what it's like.

Another possibility is that knowing what an experience is like is a matter of categorizing the experience. Suppose you've eaten a guava fruit, and you can recognize one by taste. It is tempting to suppose that knowledge of what guava fruit is like consists in your ability to recognize one when you taste it. On this view, knowing what it's like is a matter of acquiring a concept that allows you to re-identify the things you've perceived before. On closer examination, however, such conceptual capacities are neither necessary nor sufficient for phenomenal knowledge (knowing what it's like). Suppose you have never eaten guava before, but now, at this moment, have a first taste of guava in the form of a sauce on the fish you ordered. You can't identify the

flavour you are tasting, and you might not recognize it if you tasted it again, but it might be true of you that you know what guava is like, nevertheless. As you taste it and focus on the flavour, you can say things such as 'this sauce is delicious' and when you do so, you are referring to the flavour of guava. You base your appraisal of the sauce on your current knowledge of what it's like. Because you would not recognize the flavour again, it would be wrong to say you have a concept of the flavour, but it would certainly be appropriate to say that you know, then and there, what it's like. This shows that concepts used to categorize phenomenal states are not necessary for phenomenal knowledge.

Nor are they sufficient. It is noteworthy that when we apply concepts to our experiences, the concepts typically refer to features of the world. The flavour of guava is, arguably, a feature of the fruit itself, not a feature of our minds. Such concepts don't satisfy the criterion for conveying phenomenal knowledge, because they are not representations of phenomenal states. Of course, there can be concepts that refer to phenomenal states; consider the concept expressed by the phrase 'the phenomenal character of guava'. Is this concept sufficient for conferring phenomenal knowledge? Perhaps not. If the AIR theory is right, phenomenal states get their character from their perceptual qualities. Two states are phenomenally alike if and only if they represent the same perceptual features in the same sense modality. If that is right, then a concept referring to the phenomenal character of guava can be possessed and tokened in the absence of phenomenal experience. Such a concept is just a mental representation that can be used to distinguish the gustatory and olfactory state that we have when tasting guava from other gustatory and olfactory states. There is no obvious reason why such a concept cannot be applied unconsciously. That suggests that application of such concepts is not sufficient for phenomenal knowledge.

For these reasons, I do not think that the concepts we use in perceptual classification are the best candidates for explaining phenomenal knowledge. That leaves us with one more candidate: mental pointing. Predictably, I think that mental pointing is the key. Mental pointing can confer knowledge of what something is like. As I have suggested, merely having an experience does not qualify as knowledge of the experience, because knowledge is a transitive relation. When we mentally point, there is a mental control structure that uses representations of objects or spatial locations that focuses in on some aspect of experience. I said this control structure works in ways that parallel linguistic demonstratives. I would argue that, like linguistic demonstratives,

phenomenal demonstratives qualify as representations. They represent the phenomenal qualities to which they point. If I am right, then phenomenal demonstratives can confer phenomenal knowledge even though mere experience cannot. When we merely experience our phenomenal states, we do not represent them. When we mentally point, experiences are represented; pointing is transitive. This distinction has some intuitive plausibility to me. My intuition (for what it's worth) is that we would not credit a person who merely experienced something with knowledge of what that experience is like, but, when that person focuses on the experience in a top-down way, we do credit her with knowledge of what it's like. If you share this intuition, it is another reason for thinking that phenomenal demonstratives can constitute phenomenal knowledge even if experience itself does not.

The present proposal overcomes the objections that I raised to the suggestion that we attain phenomenal knowledge when we categorize our phenomenal states. I said above that the concepts used in phenomenal categorization can also be applied unconsciously, and this raises doubts about whether they are sufficient for phenomenal knowledge. I also said that they are not necessary, because we can attain phenomenal knowledge without storing concepts of the phenomenal qualities we come to know. Phenomenal demonstratives, in contrast, cannot occur without consciousness. Mental pointing serves to direct attention, and attention gives rise to consciousness. Mental pointing can also be achieved without using concepts of the qualities that we represent. For one thing, we can point using spatial representations. We can also point using object representations that are much more abstract than the quality to which we point. We can focus in on *that taste* without forming a concept that would allow us to reidentify the taste we are now experiencing.

Suppose that this is a correct analysis of phenomenal knowledge. One striking implication is that phenomenal knowledge is much more limited than we might have imagined. By that I mean the vehicles by which we come to know our phenomenal states are much coarser in detail than that which they represent. Phenomenal demonstratives do not refer by fully describing what they represent. They refer by directing attention. They have two components: a control structure and a demonstration. The control structure presumably has no descriptive content. It is not a description of any particular phenomenal state. The same control structure is used to point to different phenomenal states. Taken on it's own, these control structures do not represent anything. Demonstrations have richer representational content, but they do not represent what our phenomenal states represent. Some

demonstrations simply represent regions in space. Others represent objects (or features), but they are high-level representations that are comparatively abstract. When combined, the control structure and representation serve as a representation of whatever precise phenomenal quality they end up focusing on, but they do so in virtue of the causal impact they have on the allocation of attention.

It follows from this that two token-identical phenomenal demonstratives could have different content. Consider a case when we focus attention on the upper-rightmost region of visible space. The content of such a focusing will depend on what happens to be located in that region at a particular time. Thus, the vehicles of phenomenal knowledge cannot be used to 'read-off' what experience is like. Like a pointing finger, mental pointers represent; but if we were to examine a pointer itself we wouldn't learn much about what it represents. Paradoxically, we come to know what it's like via representations that aren't like anything — representations that do not re-present what it's like.

One might think that this analysis of phenomenal demonstratives is problematic for the following reason. Imagine looking at two colour patches and thinking, demonstratively, that one is different from that one. If phenomenal demonstratives do not describe what they designate, then it is not immediately clear how such thoughts would be possible. Wouldn't the mental pointings be the same and hence indistinguishable? I think this puzzle has a simple solution, but the solution reveals an interesting fact of phenomenal knowledge. I suggested that phenomenal demonstratives are, on any occasion, partially constituted by high-level perceptual representations or representations of space. These components can distinguish two pointings. If the colour patches are in different locations, then the mental pointers can contain different spatial representations, and if the colours differ in some dimension that we can conceptualize (e.g. if they belong to colour categories that we can store in memory and distinguish), the pointers can use different high-level perceptual representations. But suppose that the two colours are located in the same space (i.e. they are interspersed at a frequency that has finer spatial resolution than attention), *and* suppose they belong to categories that we cannot conceptually distinguish. If both of these conditions are met, then I would say we cannot have a demonstrative thought of the form that one differs from that one. Of course we can point to both and form a thought about that mix of colours, but we cannot conceive of the colours in the mix separately. This is an empirical consequence of the theory that I have been defending and it could be tested empirically. I'd like to think it's a

virtue of a theory of phenomenal knowledge that it makes testable predictions about what kinds of thoughts are possible or impossible. My point here is that the range of atomic thoughts we can have using phenomenal demonstratives is narrower than the range of states we can experience.

2.3. There are no phenomenal concepts

Before drawing philosophical conclusions from this account, I want to point out that it differs from an approach that is gaining popularity in the literature. A number of authors believe that we attain phenomenal knowledge via phenomenal concepts (e.g. Loar, 1990; Perry, 2001; Papineau, 2002). Loar characterizes a phenomenal concept as a special kind of recognitional concept that is dependent on a phenomenal state that it classifies. Perry treats phenomenal concepts as concepts that pick out phenomenal states indexically, rather than by merely describing them. Papineau suggests that phenomenal concepts are quotational: they contain the phenomenal states to which they refer. As I understand these proposals, they can be captured by a common definition: phenomenal concept is a concept that both represents phenomenal qualities and cannot be possessed without having those phenomenal qualities. I think no such concepts exist.

Notice first that the concepts we use to categorize our experiences are not phenomenal concepts. This follows directly from my argument that these concepts can be possessed without consciousness. It also follows from the fact that these concepts are too coarse-grained. They do not represent the precise qualities that are available to us in experience. These concepts represent comparatively abstract distal features of the environment.

The natural move at this point would be to say that phenomenal demonstratives are phenomenal concepts. A number of defenders of phenomenal concepts have offered demonstrative accounts. They equate phenomenal concepts with phenomenal demonstratives. I think this proposal is indefensible. It is widely agreed among concepts researchers that concepts are mental entities that can be generated by an organism to re-identify things. To qualify as a concept, a mental representation must be capable of being stored. I think that phenomenal demonstratives cannot be stored in the way that matters. Phenomenal demonstratives work by focusing on an aspect of current experience. I have argued that records of experiences cannot be stored, because experience resides at a level of representation that cannot be encoded in memory. Now one might think that one could

store a phenomenal demonstrative. For example, one might think that one can store a record of the phenomenal demonstrative we use to focus on a particular region of space. Perhaps each day we spend an hour focusing on the upper left, and we store a record of the command to stare in that direction. Does this stored demonstrative qualify as a phenomenal concept? I claim that it doesn't. The stored command does not represent a phenomenal quality. It only represents a phenomenal quality on those specific occasions when it is put into use and, on each occasion, it may represent a different quality. So phenomenal demonstratives cannot be stored in a content-preserving way. They cannot qualify as concepts of phenomenal qualities.

Of course, there is one way to store a precise record of what we experience: we can store precise descriptions of phenomenal states in a topic-neutral language. For example, I can store the concept 'the blue that I would experience when looking at a such-and-such colour chip under such-and-such viewing conditions' or 'the taste I would experience when I'm in such-and-such brain state'. Such descriptive representations may be able to represent specific phenomenal qualities if fully spelled out, but they are not phenomenal concepts because they can be possessed without ever having had the particular experience that they represent.

Before closing this section, let me mention one quick objection. I have been claiming that there are no phenomenal concepts, and my basic argument for that claim is that concepts are stored representations, and we don't store records of phenomenal states. Against this, one might argue that there are obvious examples of stored phenomenal qualities. We can recall what various colours, tastes and sounds are like, and we can form conscious mental images of these. Doesn't this show that there are phenomenal concepts?

Earlier I suggested that conscious imagery is mediated by stored high-level perceptual representations. If so, the existence of imagistic recall does not show that we have phenomenal concepts. High-level perceptual representations are not phenomenal concepts. They do not represent phenomenal states (they represent objects), and they can exist without phenomenal qualities. These representations can be used to generate intermediate-level perceptual states through efferent pathways. The states that are generated in this way are mental images, and when we attend to those images they are consciously experienced. But notice that the experience of images is not mediated by phenomenal concepts. Notice too, as remarked above, that high-level perceptual representations abstract away from many features of perception and, as a result, the intermediate-level representations that they generate in

imagery are correspondingly inexact. It is difficult to imagine a specific shade of blue, for example. What we get in imagery is an unstable, ephemeral and pale counterpart of the kind of blues we experience in perception. If you imagine blue and then try to match your image precisely to a colour chip, you will find it difficult or impossible. The range of blues we can point to in experience outstrips those we can imagine. In sum, the stored records used to generate images do not qualify as phenomenal concepts, and in any case they also cannot be used to fully explain phenomenal knowledge.

I conclude that we have no phenomenal concepts. I see no way to store a representation that represents a specific phenomenal quality in a way that depends on the experience of that quality. This conclusion depends on the specific definition of concepts that I am using, but that definition is not unusual. Concepts researchers widely and routinely assume that concepts are stored records that can be used to re-identify their referents. I would encourage consciousness researchers to drop the term 'phenomenal concept' on the grounds that is gives a misleading impression of the way we represent our phenomenal states.

3. The Knowledge Argument

3.1. What Mary learns

Recent interest in phenomenal concepts has been driven by the hope that phenomenal concepts can help to undermine the 'knowledge argument' for dualism (Loar, 1990; Hill, 1997; Perry, 2001; Papineau, 2002). The most influential version of the knowledge argument has been put forward by Frank Jackson (1982). He uses epistemic premises to support the conclusion that some of the facts about phenomenal experience are not physical. Like all other physicalists, I think the argument is fallacious but, because I reject phenomenal concepts, I don't think that they can be used to expose the fallacy. That said, my explanation of the fallacy, which calls on phenomenal demonstratives, closely parallels what others have said under the rubric of phenomenal concepts. So my treatment here will be brief.

Jackson's most widely discussed version of the knowledge argument hinges on a thought experiment about a brilliant neuroscientist named Mary, who has been trapped inside a black and white room. Mary learns everything about what goes on in the brain when people see colours and, as a result, she can be said to know every physical fact about colour experience. But when she is finally exposed to colours for the first time, she learns something new. She learns, for instance, what the colour that people call 'blue' is like. This, Jackson maintains,

is a fact she didn't know before. Thus, some facts about colour experience are not included in the sum of all physical facts. Some facts are non-physical. This, we are invited to infer, means that physicalism is false.

The standard version of the phenomenal concepts reply goes like this. Mary does not learn a new fact; she learns a new way of representing facts she already knew. That new way of representing old facts arises because she acquires a new concept. Initially, Mary just has neural concepts, learned from her textbooks on neuroscience, to describe phenomenal states. But when she sees colours for the first time, she acquires phenomenal concepts, according to the standard story. These concepts represent the same brain states as her neural concepts, but they do so in a phenomenal way.

At first, this appeal to phenomenal concepts may not appear especially helpful. Usually, when two concepts co-refer they refer via different properties. For example, 'triangle' and 'trilateral' refer to the same polygons by reference to their interior angles and sides, respectively. If phenomenal concepts were co-extensive with neural concepts, the very fact that these concepts differ would suggest that they would still imply that they refer via different properties, and that would suggest that phenomenal concepts refer via properties that are not physical. This is one reason why phenomenal demonstratives are so appealing. Demonstratives do not refer by fully describing their contents; they refer by being in the right relation to their contents. So postulating phenomenal demonstrative does not lead to proliferation of properties. Moreover, phenomenal demonstratives work by pointing to occurrent phenomenal states, which means that one cannot have a phenomenal demonstrative without having the state to which it refers. This explains why Mary needs to experience colours to know what they are like.

I find this account congenial. My main quibble is that I don't think that phenomenal demonstratives are concepts. Mary can learn what it's like without acquiring any concepts; she can think to herself that red is like that (while mentally pointing), without being able to recognize, recall or imagine red on a future occasion. This distinguishes my proposal from Loar's seminal recognitional concepts account. Perry (2001) also claims that Mary acquires a recognitional concept, but my account is, perhaps, closer to his, because he emphasizes the role of attention in mental pointing. In denying that Mary can necessarily recognize red, I am also implicitly departing from Lewis's (1990) and Nemirow's (1990) suggestion that what Mary acquires is a new ability (compare Tye's, forthcoming, response). I do think that there is a kernel

of truth to the ability proposal, however. As remarked above, I think phenomenal demonstratives are control structures. The deployment of a phenomenal demonstrative involves a kind of procedural knowledge. But rather than saying that Mary learns how to recall and recognize colours, I'd say that she acquires the ability to focally attend to them. My view is not a radical departure from the others in the literature; it builds on the lessons of each. But the subtle differences invite a shift away from the focus on concepts and onto the topic of attention as investigation (both philosophical and empirical) of phenomenal knowledge moves forward.

It's worth noting here that Mary would not necessarily learn what colours are like if she were merely able to perceive them. Colours can be perceived unconsciously. For example, if you are briefly presented with a red flash and a green flash, you will experience neither red nor green, but yellow. Suppose Mary's first encounters with colour took this form. She would know what yellow is like, but not red and green. The story that I have been telling explains this. Attention takes time. When red and green are flashed briefly, one cannot attend to them. By the time attention allows information to flow forward to working memory, the aftereffects of the red and green stimuli have blended together, and we experience the result as yellow. Nevertheless, the red and green are perceived. We know this because images embedded in briefly flashed colours cause stimulus specific brain responses, even though the figures are not perceived (Moutoussis and Zeki, 2002). This underscores the suggestion that phenomenal knowledge is not constituted by perceptual states, but rather requires attention to such states.

The phenomenal demonstratives story can be used to refute the core assumption underlying the knowledge argument. The assumption, often discussed under the rubric of '*a priori* physicalism' says: if phenomenal states were physical, knowledge of what they are like could be deduced from knowledge encoded in physical vocabulary (Chalmers and Jackson, 2001). I argued above that phenomenal knowledge is very limited. We know what our experiences are like via demonstratives that refer by their causal relations to what they represent, not by describing them (cf. Loar, 1990). One can make a deductive inference about the applicability of a concept only if that concept decomposes into descriptive features, and one could infer what phenomenal qualities are like only if our knowledge of what they are like was encoded in concepts that decompose into features that are functional or physical. Phenomenal knowledge is not encoded in that way and therefore cannot be deduced from physical descriptions of the brain. Above I

also distinguished knowing what an experience is like and having that experience. Having an experience is not knowledge of the experience. It should be obvious that knowledge of the brain is not sufficient for having an experience, any more than knowledge of how wings work is sufficient for having wings.

That said, I do think that knowledge of the brain might be useful for inferring *something* about phenomenal states. Such knowledge could be used to deduce facts about which qualities are similar to each other and which features of the world cause our qualitative experiences. These deductions might allow us to determine what concepts people have for classifying phenomenal states. It might not suffice for *acquiring* such concepts, but it would at least allow us to generate plausible lists of what those concepts are. Mary might deduce the categorical boundaries of colours, just as we can deduce perceptual similarity spaces from creatures with sensory systems that differ from our own. The concepts used to categorize phenomenal states encode information about similarities between experiences, as well as information about distal features, and such information can be discovered by observing brain mechanisms and behaviour. But, I have argued that these concepts are not themselves phenomenal. They do not constitute a form of phenomenal knowledge, and they are not necessary or sufficient for knowing what phenomenal states are like. The fact that Mary can deduce these things without knowing what it's like to have phenomenal states suggest that phenomenal knowledge is not merely knowledge of what our phenomenal states represent. Thus, I think the knowledge argument is a good argument against some forms of representationalism about consciousnesses; in particular, it counts against the view that the phenomenal character of an experience is nothing beyond what that experience represents.

In sum, I have endorsed the view that phenomenal demonstratives can explain what's wrong with the knowledge argument. Mary does not learn any new facts when she learns what it's like. Instead, she enters into an epistemic situation in which she can mentally point to a perceptual state caused by seeing something that is coloured. Mentally pointing gives her knowledge of what colours are like. She didn't have this knowledge before and she couldn't have it, if she couldn't focus attention on her perceptual states.

3.2. Three brief objections

Some objections have been raised against the suggestion that what Mary learns can be explained by appeal to phenomenal demonstratives.

In this final section, I will briefly discuss three objections. More detailed discussion can be found in the literature.

One objection is discussed by Nina-Rümelin (2002), though she recognizes that it can be answered. Demonstratives require reference-fixing demonstrations. If I merely say 'that' without a demonstration, my demonstrative will not refer. I must say something like 'that table' or 'that' while pointing, to make some object salient. There is a worry that, in order for phenomenal demonstratives to refer, they would need reference fixes that are phenomenal in nature. How, one might wonder, can you point to a phenomenal state except by demonstrating its phenomenal character? If pointing to phenomenal states requires the use of phenomenal reference fixers, the phenomenal demonstrative account would be no better than the simplest version of the phenomenal concepts account. It would introduce phenomenal properties as modes of presentation for phenomenal states.

My account of phenomenal demonstratives addresses this concern. I have proposed that the demonstrations used by phenomenal demonstratives are not themselves phenomenal. They are high-level perceptual representations or representations of locations in perceptual space. These representations are unconscious on my view, and they merely serve to direct attention. So my account does not introduce a regress of reference-fixing phenomenal properties.

A second objection owes to Chalmers (2002). He distinguishes phenomenal demonstratives from 'pure phenomenal concepts', concepts that refer directly to specific phenomenal qualities. Demonstratives, according to Kaplan, have both content (what they refer to) and character (the rule by which they refer). As a result, two token-identical phenomenal demonstratives could refer to different things in different worlds — a point I made above. Chalmers postulates that pure phenomenal concepts do not admit of a character/content distinction. Moreover, he says, we can have informative thoughts of the kind we might express by 'that is such-and-such' where 'that' expresses a phenomenal demonstrative and 'such-and-such' expresses a pure phenomenal concept. This suggests that these two kinds of representations are different and, Chalmers claims, Mary's new knowledge should be characterized in terms of the acquisition of a pure phenomenal concept, not a phenomenal demonstrative.

I flatly reject Chalmers' claim that there are pure phenomenal concepts. I have argued that such concepts do not exist. We certainly have phenomenal states (I am a phenomenal realist), but it's quite another thing to assume we have concepts that refer to these states, much less concepts that refer in the direct way that Chalmers has in mind (in his

terms, concepts with the same primary and secondary intensions). I see no reason to believe that such concepts exist. In arguing for such concepts, Chalmers alleges that there is an informative identity expressed by 'that is such-and-such'. I don't share his intuition. I frankly have no idea what thought this would be. I can certainly form the thought that I would express by 'that is the colour people call "blue" ', but I cannot imagine an informative thought expressed by 'that is blue' where 'blue' names a concept that is neither functional nor physical nor deferential. Suppose I know nothing about brain or colour vocabulary. Now I have a colour experience. I can imagine thinking about the experience by attending to it. But, when I hold this colour in mind, the only *identity* claim I could come up with would be the trivial one expressed by 'that is that'. I would have only one way of thinking about this colour: namely pointing to it in my mind. To show that there are phenomenal concepts in addition to phenomenal demonstratives, Chalmers would have to show that one could think about phenomenal qualities as such without being able to point to them, and conversely. I just don't see any reason to accept this conjecture.

There are other objections in the literature that are intended to undermine any attempt to explain Mary's knowledge by appeal to new ways of representing old facts. For example, Chalmers (2006) has recently devised a new general-purpose argument against phenomenal concepts, which could be adapted to argue against the non-conceptual phenomenal demonstrative proposal that I have defended. Roughly, he says that if a theory of phenomenal concepts that can be stated in physical terms (hence consistent with physicalism) is such that we could conceive of the theory being realized by an unfeeling zombie, then the theory cannot explain what is special about Mary's epistemic situation. Whatever she learns, it's different from what her zombie counterpart would learn. This argument can be used to challenge the view that Mary's knowledge is given to her by phenomenal demonstratives. I have defined these as mental pointers that give us access to perceptual states by directing top-down attention. Mary's zombie counterpart could have top-down attention. Zombie Mary would, upon first seeing colours, be able to focus in on a visual state caused by a coloured object. This would be new knowledge, knowledge that she might describe as what it's like to see colours, but it would be different from the knowledge that Mary learns.

In a compelling response to this objection, Carruthers and Veillet (this volume) argue that, were zombies possible, zombie Mary would indeed be in the same epistemic situation as Mary. Zombie would

learn just as much as Mary does, and for exactly the same reason. They say that the object of knowledge would be different in the two cases. Mary would have knowledge of a phenomenal state and zombie Mary would not, but that does not mean they are in different epistemic situations towards those objects of knowledge, and therefore the materialist account of what Mary learns can explain her epistemic predicament.

I think that Carruthers and Veillet give the right response to Chalmers. Mary and zombie Mary would be in the same epistemic situation, in the relevant sense. They would both come to know something new, because they would be able to form demonstrative thoughts by mentally pointing to their perceptual representations of colours. Chalmers thinks it is obvious that their epistemic situations are different. I think this intuition derives from the fact that Mary clearly learns something different from zombie Mary. Mary learns that red is like *that*, where that refers to a phenomenal state. But this difference is irrelevant here. An account of phenomenal demonstratives is not supposed to explain phenomenology. It is supposed to explain why phenomenal knowledge cannot be inferred from physical descriptions. It handles this explanatory burden well. Both physicalists and dualists typically agree that phenomenal knowledge uses representations that one could not use if one did not have the states in question. Phenomenal demonstratives prove that this is consistent with physicalism. Demonstratives are representations that work by pointing, and mental analogues of demonstratives cannot apply without being in the mental states to which they point. This is a simple and principled point about representation, and it's enough to prove that some kinds of knowledge, namely demonstrative knowledge, cannot be deduced from physical descriptions. Moreover, there is good reason to think that our knowledge of phenomenal states is demonstrative. So, we have reason to think, antecedently, that phenomenal knowledge is not deducible.

Given the availability of such promising physicalist replies to the knowledge argument, I think we might do well to stop worrying about zombies for a while and dedicate more energy to trying to identify the actual mechanisms underlying phenomenal states and phenomenal knowledge.

4. Conclusions

Let me conclude by listing the main claims that I have made in this discussion. I began by saying that conscious states are AIRs, attended intermediate-level representations. I also claimed that attention works

by making information available to working memory, but I noted that the information does not need to be encoded in working memory to have a conscious experience. I then distinguished three relations with experiences: we can have them, classify them or point to parts of them. The last of these, mental pointing, is achieved by means of phenomenal demonstratives, which I analysed in terms of top-down attention. I argued that phenomenal demonstratives are the source of phenomenal knowledge, but they represent our phenomenal states without fully describing them. I do not think that phenomenal demonstratives are concepts, and I claimed we have no phenomenal concepts, i.e. concepts that both represent phenomenal states and require that we have the phenomenal states that they represent. Finally, I endorsed the view that phenomenal demonstratives can block the dualist conclusion of the knowledge argument.[1]

References

Beck, D.M., Rees, G., Frith, C.D. and Lavie, N. (2001), 'Neural correlates of change detection and change blindness', *Nature Neuroscience*, **4**, pp. 645–50.

Block, N. (2002), 'The harder problem of consciousness', *Journal of Philosophy*, **XCIX**, pp. 391–425.

Chalmers, D. (2002), 'Content and epistemology of phenomenal belief', in *Consciousness: New Philosophical Essays*, ed. Q. Smith and A. Jokic (Oxford: Oxford University Press).

Chalmers, D. (2006), 'Phenomenal concepts and the explanatory gap', in *Phenomenal Concepts and Phenomenal Knowledge*: *New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter (Oxford: Oxford University Press).

Chalmers, D. and Jackson, F. (2001), 'Conceptual analysis and reductive explanation', *Philosophical Review*, **110**, pp. 315–60.

Dehaene, S., Naccache, L., Le Clec, H.G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.F. and LeBihan, D. (1998), 'Imaging unconscious semantic priming', *Nature*, **395**, pp. 597–600.

Hayden, B.Y. and Gallant, J.L. (2005), 'Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4', *Neuron*, **47**, pp. 637–43.

Hill, C.S. (1997), 'Imaginability, conceivability, possibility, and the mind-body problem', *Philosophical Studies*, **87**, pp. 61–85.

Jackendoff, R. (1987), *Consciousness and the Computational Mind* (Cambridge, MA: MIT Press).

Jackson, F. (1982), 'Epiphenomenal qualia', *Philosophical Quarterly*, **32**, pp. 127–36.

Kaplan, D. (1989), 'Demonstratives', in *Themes From Kaplan*, ed. J. Almog, J. Perry and H. Wettstein (New York: Oxford University Press), pp. 481–564.

Lewis, D. (1990), 'What experience teaches', in *Mind and Cognition*, ed. W. Lycan (Oxford: Basil Blackwell).

Loar, B. (1990), 'Phenomenal states', *Philosophical Perspectives*, **4**, pp. 81–108.

[1] I am immensely indebted to Rocco Gennaro and two anonymous referees, who provided extensive comments. This paper would have been far worse without their help.

Mack, A. and Rock, I. (1998), *Inattentional Blindness* (Cambridge, MA: MIT Press).
Millikan, R. (2000), *On Clear and Confused Ideas* (Cambridge: Cambridge University Press).
Most, S.B., Scholl, B.J., Clifford, E. and Simons, D.J. (2005), 'What you see is what you set: Sustained inattentional blindness and the capture of awareness', *Psychological Review*, **112**, pp. 217–42.
Moutoussis, K. and Zeki, S. (2002), 'The relationship between cortical activation and perception investigated with invisible stimuli', *Proceedings of the National Academy of Sciences*, **99**, pp. 9527–32.
Nemirow, L. (1990), 'Physicalism and the cognitive role of acquaintance', in *Mind and Cognition*, ed. W. Lycan (Oxford: Blackwell).
Nina-Rümelin, M. (2002), 'Qualia: The knowledge argument', *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/qualia-knowledge/
Olshausen, B.A., Anderson, C.H. and Van Essen, D.C. (1993), 'A neurobiological model of visual attention and invariant pattern recognition based on DynamicR of information', *Journal of Neuroscience*, **13**, pp. 4700–19.
Papineau, D. (2002), *Thinking about Consciousness* (New York: Oxford University Press).
Pérez-Carpinell, J., Baldovi, R., de Fez, M.D. and Castro, J. (1998), 'Color memory matching: Time effect and other factors', *Color Research and Application*, **23**, pp. 234–47.
Perry, J. (2001), *Knowledge, Possibility, and Consciousness* (Cambridge, MA: MIT Press).
Postle, B.R. (2006), 'Working memory as an emergent property of the mind and brain', *Neuroscience*, **139**, pp. 23–8.
Prinz, J.J. (2000), 'A neurofunctional theory of visual consciousness', *Consciousness and Cognition*, **9**, pp. 243–59.
Prinz, J.J. (2001), 'Functionalism, dualism and the neural correlates of consciousness', in *Philosophy and the Neurosciences: A Reader*, ed. W. Bechtel, P. Mandik, J. Mundale and R. Stufflebeam (Oxford: Blackwell).
Prinz, J.J. (2002), *Furnishing the Mind: Concepts and Their Perceptual Basis* (Cambridge, MA: MIT Press).
Prinz, J.J. (2005), 'A neurofunctional theory of consciousness', in *Cognition and the Brain: Philosophy and Neuroscience Movement*, ed. A. Brook and K. Akins (Cambridge: Cambridge University Press), pp. 381–96.
Prinz, J.J. (2007), 'All consciousness is perceptual', in *Fundamental Debates in Philosophy of Mind*, ed. J. Cohen and B. McLaughlin (Oxford: Blackwell).
Prinz, J.J. (forthcoming), *The Conscious Brain* (New York: Oxford University Press).
Simons, D.J. and Levin, D.T. (1997), 'Change blindness', *Trends in Cognitive Science*, **1**, pp. 261–7.
Sperling, G. (1960), 'The information available in brief visual presentations', *Psychological Monographs*, **74**, pp. 1–29.
Tye, M. (forthcoming), 'Knowing what it is like: The ability hypothesis and the knowledge argument', *Protosociology: Collection of Essays for David Lewis*.
Vogel, E.K. and Luck, S.J. (2002), 'Delayed working memory consolidation during the attentional blink', *Psychonomic Bulletin & Review*, **9**, pp. 739–43.