The Blackwell Companion to Consciousness
Edited by Max Velmans, Susan Schneider

33

The Neurophilosophy of Consciousness

PETE MANDIK

The topic of phenomenal consciousness concerns what it means for mental states to be conscious states (as opposed to unconscious mental states) and what it means for such states to have phenomenal character, that is, to have properties in virtue of which there is "something it's like" for a subject to be in such a state. Traditional philosophical issues that phenomenal consciousness raises involve the relation of phenomenal consciousness to the rest of the world, especially as that world is conceived of by the natural sciences. Thus much philosophical discussion concerns whether the world as conceived of by physical theory can adequately accommodate phenomenal consciousness or if instead we are left with a dualism that cleaves reality into, for example, a nonphysical phenomenal consciousness and a physical everything else. Even among philosophers who agree that phenomenal consciousness is consistent with physicalism, there is much disagreement, for there are several proposals for how best to spell out the consistency of a physicalistic worldview that makes room for phenomenal consciousness. One way of portraying this cluster of issues is in terms of which natural science is best suited to study phenomenal consciousness and how to conceive of the relation between that science and the sciences involving the most basic aspects of reality (the physical sciences). One major view is that psychology is the proper science for understanding phenomenal consciousness and furthermore, that psychological investigation of phenomenal consciousness should be regarded as autonomous from sciences such as the neurosciences. In opposition is the view that the proper science is neuroscience and whatever contributions come from psychology are only valid insofar as psychological theories are reducible to neuroscientific theories. Increasingly, proponents of the latter view identify themselves as practitioners of neurophilosophy.

Neurophilosophy is a sub-genre of naturalized philosophy – philosophy that embraces Quine's (1969) vision of philosophy as continuous with the natural sciences – wherein the natural science in primary focus is neuroscience. It is perhaps worth addressing here in further detail what is distinctive of neurophilosophy as opposed to other kinds of naturalism. The role that neuroscience plays is, of course, key, but not just any mention of the brain in a philosophical theory will suffice to make it neurophilosophical. Neurophilosophical appeals to neuroscience involve explicit and detailed use of contemporary neuroscientific literature. Furthermore, neurophilosophy is not to be distinguished from other forms of naturalism by the philosophical *conclusions* that might be reached but by the role that contemporary neuroscience plays in the *premises* of the arguments for those conclusions. These

points about different styles of naturalistic philosophizing may be illustrated in terms of some recent examples. For example, Jaegwon Kim is a kind of naturalist and even advocates a reduction of mental state types to physical state types. However, he is not thereby a neurophilosopher. His identification of the relevant physical state types makes no explicit reference to contemporary neuroscientific findings. The state types in question involve no familiarity with the typologies specific to either neurophysiology or neuroanatomy. In contrast, the research of neurophilosophers like Kathleen Akins makes explicit reference to contemporary neuroscientific findings in the arguments for various naturalistic conclusions. For example, she argues (1996) against traditional views of the role that sensory states play in grounding the contents of intentional states. Crucial to her arguments are detailed examinations of the neurophysiology of thermoreception (see Bickle & Mandik 1999 for a longer discussion of examples of neurophilosophical work such as Akins's).

Some authors draw a distinction between neurophilosophy and philosophy of neuroscience wherein the former involves the application of neuroscientific results to topics of philosophical concern, usually in the philosophy of mind, and the latter is a sub-discipline of the philosophy of science. Though often neurophilosophers are also philosophers of neuroscience, the current chapter focuses on the activities distinctive of the former group.

The term "neurophilosophy" entered philosophical parlance with the publication of Patricia Churchland's *Neurophilosophy* (1986), the aims of which were to introduce neuroscience to philosophers and philosophy to neuroscientists, with an emphasis on the former. Patricia Churchland and husband Paul Churchland are paradigmatic examples of neurophilosophers. Their professional training is primarily philosophical, their appointments are in philosophy departments, and they publish in philosophy journals. Because of this, neuroscience and philosophy do not have equal influence over neurophilosophy. Instead the primary forces that drive its development as an academic pursuit emanate from conventions of philosophical institutions. Thus neurophilosophical work on phenomenal consciousness proceeds largely by bringing neuroscientific theory and data to bear on philosophical questions concerning phenomenal consciousness.

Such questions are diverse. However, a useful way to focus the discussion – as well as to understand what has been of primary concern to neurophilosophical theories of phenomenal consciousness – will be to focus on just three questions: the question of state consciousness, the question of transitive consciousness, and the question of phenomenal character. (The terms "transitive consciousness" and "state consciousness" are due to David Rosenthal. For discussion, see Rosenthal 1993; Tye, chapter 2.) The question of state consciousness concerns in what consists the difference between mental states that are conscious and mental states that are unconscious. We have conscious mental states, such as my conscious perception of the words I type. Mental states vary with respect to whether they are conscious. Consider, for example, your memory of your mother's name. You may have had that memory for years but it obviously was not a *conscious* memory for the entire time between its initial acquisition and its current retrieval. In what does the difference between conscious and unconscious mental states consist? The question of transitive consciousness concerns what it is that we are conscious *of*. When one has a conscious state, typically, if not always, one is conscious *of* something, as when I am conscious of a buzzing insect. Things may vary with respect to whether I am conscious of them, as when I am only intermittently conscious of the conversation at a nearby table in a restaurant. What does it mean to be *conscious of* something? The question of phenomenal character concerns the so-called qualia of conscious states. Conscious states have certain properties – their phenomenal

character – properties in virtue of which there is "something it's like" to be in that state. When I have a conscious perception of a cup of coffee there is, presumably, something it's like for me to have that perception and, for all I know, what it's like for you to have a conscious perception of a cup of coffee is quite different. What makes a conscious state have "something it's like" to be in that state? The phrase "phenomenal consciousness" does not denote a kind of consciousness distinct from state consciousness but is instead a term of art used by authors (e.g., Block 1995; Chalmers 1996) who are primarily interested in a certain aspect of conscious states, namely their phenomenal character (for a longer discussion see Mandik 2005).

Given the centrality of these questions, we will have several occasions to return to them throughout the present chapter. In brief summary they are:

The Question of State Consciousness:
In what consists the difference between mental states that are conscious and mental states that are unconscious?

The Question of Transitive Consciousness:
When one has a conscious mental state, what is one thereby conscious *of*?

The Question of Phenomenal Character:
When one has a conscious state, in what consists the properties in virtue of which there is something it's like for one to be in that state?

Neurophilosophical theories of consciousness bring neuroscience to bear on answering these three questions of consciousness. The question arises, of course, of what motivates the neurophilosophy of consciousness. The primary answer is that neurophilosophy has a certain appeal to those with an antecedent belief in physicalism, in that neurophilosophy seems especially well-suited to bridge the gap between the phenomenal and the physical. Attempting to bridge the gap by reducing the phenomenal all the way down to chemistry or microphysics may strike many as too far a distance to traverse. More plausible is to seek a higher-level physical set of phenomena, as offered in biology. Of the biological phenomena, the most plausible candidates are neural. The appeal of neurophilosophical approaches to phenomenal consciousness may become more evident upon examination of some sample theories.

Before examining the neurophilosophical theories, it will be useful to look at a small sample of some of the relevant neuroscience. Vision is one of the most important and best understood senses. Accordingly, most of the fruitful progress in combining philosophy and neuroscience has occurred in the domain of visual consciousness.

Neuroscience and Visual Consciousness

The processing of visual information in the brain can be understood as occurring in a processing hierarchy with the lowest levels in the retina and the highest levels in areas of the cerebral cortex. Processing begins after light is transduced by the rods and cones in the retina and electrochemical signals are passed to the retinal ganglia. From there, information flows through the optic nerve to the lateral geniculate nucleus (LGN) in the subcortex. From

the LGN, information is passed to the first stage of cortical processing in the primary visual area of occipital cortex (area V1). From V1, the information is sent to other areas of occipital cortex and is then sent along a "ventral stream" from the occipital to the infero-temporal cortex as well as along a "dorsal stream" from the occipital to the posterior parietal cortex (Milner & Goodale 1995; Prinz, chapter 19; Crick & Koch, chapter 44; Goodale, chapter 48). Beyond that, information is sent to areas of the frontal cortex (Olson et al. 1999) as well as the hippocampus (Milner & Goodale 1995). As will be discussed further, information does not simply flow from lower levels to higher levels but there are many instances in which it flows from higher levels down to lower levels (Pascual-Leone & Walsh 2001). Furthermore, information is processed in various ways in different regions of the different levels and can be briefly characterized in the following ways. Information at the lowest levels is represented by neural activations that serve as detectors of features in specific locations defined relative to the retina (AKA retinocentric locations). Thus, at the lowest levels, neural activations in LGN and V1 constitute egocentric representations of visual features as in, for instance, the detection of an oriented line by a cell with a relatively small retinocentric receptive field. At progressively higher-level areas (such as visual areas V2 through V5), locally defined visual features are "grouped" or integrated as when local information about shading is grouped to give rise to representations of depth. Progressively higher levels of information processing increasingly abstract away from the egocentric information of the lower-level representations and give rise to progressively allocentric ("other-centered") representations as in view-point invariant representations in inferior temporal cortex that underwrite the recognition of objects from multiple angles and other viewing conditions. Thus, information represented at progressively higher levels of processing becomes progressively less egocentric and progressively more allocentric, the most allocentric representations being in the frontal areas and hippocampus (Mandik 2005).

The question arises of how best to apply the concepts of consciousness of interest to philosophers – state consciousness, transitive consciousness, and phenomenal character – in the context of a neuroscientific understanding of visual perception. We may make the most progress in this regard by focusing on breakdowns and anomalies of normal vision. We will briefly examine two such cases. The first is blindsight, a condition that results from a certain kind of brain damage (Weiscrantz, chapter 13). The second is motion-induced blindness, a condition that occurs in normal subjects under certain unusual conditions.

Blindsight is a condition in which lesions to V1 cause subjects to report a loss of consciousness in spite of the retention of visual ability. For so-called blind regions of their visual fields, blindsight subjects are nonetheless better than chance in their responses (such as directed eye movements or forced-choice identifications) to stimulus properties such as luminance onset (Pöppel, Held, & Frost 1973), wavelength (Stoerig & Cowey 1992), and motion (Weiskrantz 1995). Lack of consciousness is indicated in such studies by, for example, having the subject indicate by pressing one of two keys "whether he had any experience whatever, no matter how slight or effervescent" (Weiskrantz 1996).

Blindsight subjects' responses to stimuli in the blind portions of their visual fields give evidence that the stimuli are represented in portions of the brain. However, it is clear that these representational states are not conscious states. Thus, the kind of consciousness that seems most relevant in describing what blindsight patients lack is state consciousness. Furthermore, blindsight patients arguably also lack transitive consciousness with respect to the stimuli in the blind regions of their visual field. One consideration in favor of this view arises when we take the subject's own reports at face value. They claim not to be

conscious of the stimuli in question. It would be difficult to affirm that blindsight subjects do have transitive consciousness of the relevant stimuli without affirming that all instances of representation are instances of transitive consciousness, and thus instances of unconscious consciousness.

Regarding the question of qualia, of whether there is anything it's like for blindsight subjects to have stimuli presented to the blind regions of their visual fields, I take it that it is quite natural to reason as follows. Since they are not conscious of the stimuli, and since the states that represent the stimuli are not conscious states, there must not be anything it's like to have stimuli presented to those regions. Of course, the reader may doubt this claim if the reader is not a blindsight subject. It will be useful in this regard to consider a case that readers will be more likely to have first-person access to. For precisely this reason it is instructive to look at the phenomenon of motion-induced blindness (Bonneh et al. 2001).

Motion-induced blindness may be elicited in normal subjects under conditions in which they look at a computer screen that has a triangular pattern of three bright yellow dots on a black background with a pattern of blue dots moving "behind" the yellow dots. As subjects fixate on the center of the screen, it appears to them that one or more of the yellow dots disappear (although in reality the yellow dots remain on the screen). The effect is quite salient and readers are encouraged to search the internet for "motion-induced blindness" and experience the effect for themselves. There are several lines of evidence that even during the "disappearance" the yellow dots continue to be represented in visual areas of the brain. The effect can be influenced by transcranial magnetic stimulation to the parietal cortex (a relatively late stage of visual processing in the brain). Additionally, the effect can be shown to involve nonlocal grouping of the stimulus elements. So, for example, if the yellow dots are replaced with a pair of partially overlapping circles, one yellow and one pink, sometimes an entire circle will disappear leaving the other behind even though some parts of the two different circles are very close in the visual field. As mentioned previously, the brain mechanisms thought to mediate such object groupings are relatively late in the visual processing hierarchy.

We may turn now to the applications of the concepts of transitive consciousness, state consciousness, and qualia to motion-induced blindness. First, motion-induced blindness looks to be a phenomenon involving transitive consciousness since in the one moment the subject is conscious of the yellow dot, in the next they are not conscious of the yellow dot, and along the way they are conscious of a yellow dot seeming to disappear. Second, we can see that motion-induced blindness allows for applications of the concept of state consciousness, since studies of motion-induced blindness provide evidence of conscious states that represent the presence of yellow dots as well as unconscious states that represent the presence of yellow dots.

Let us turn now to ask how the concept of phenomenal character applies in the context of motion-induced blindness. The best grip we can get on this question is simply by asking what it's like to see yellow dots disappear. When there is an unconscious state that represents the yellow dots or no transitive consciousness of yellow dot, there is, with respect to the yellow dot, nothing it's like to see it. Or, more accurately, what this instance of motion-induced blindness is like, is like *not* seeing a yellow dot. When the state representing the yellow dot is conscious, what it's like to be in that state is like seeing a yellow dot. One might suppose then, as will be discussed later, that what it's like to be in the conscious state is determined, at least in part, by the representational content of that state. In this case, it is the content of the representation of a yellow dot.

Neurophilosophical Theories of Consciousness

I will now turn to examine a sample of neurophilosophical theories of consciousness. In keeping with the definitions of neurophilosophy as well as the three questions, the discussion of this section will be centered on philosophical accounts of state consciousness, transitive consciousness, and phenomenal character that make heavy use of contemporary neuroscientific research in the premises of their arguments.

In keeping with the paradigmatic status of the work of the Churchlands in neurophilosophy, my primary focus will be on Paul Churchland's neurophilosophical work on consciousness. However, other philosophers have produced neurophilosophical accounts and I will discuss their work as well.

Paul Churchland articulates what he calls the "dynamical profile approach" to understanding consciousness (2002). According to the approach, a conscious state is any cognitive representation that is involved in:

1 a moveable attention that can focus on different aspects of perceptual inputs;
2 the application of various conceptual interpretations of those inputs;
3 holding the results of attended and conceptual interpreted inputs in a short-term memory that
4 allows for the representation of temporal sequences.

Note that these four conditions primarily answer the question of what makes a state a conscious one. Regarding the question of what we are conscious of, Churchland writes that "a conscious representation could have any content or subject matter at all" (p. 72) and he is especially critical of theories of consciousness that impose restrictions on the contents of conscious representations along the lines of requiring them to be self-representational or meta-representational (pp. 72–4).

Much of Churchland's discussion of the dynamical profile account of consciousness concerns how all of the four conditions may be implemented in recurrent neural networks. A recurrent neural network may be best understood in terms of contrast with feedforward neural networks, but we should first give a general characterization of neural networks. Neural networks are collections of interconnected neurons. These networks have one or more input neurons and one or more output neurons. They may additionally have neurons that are neither input nor output neurons and are called "interneurons" or "hidden-layer" neurons. Neurons have, at any given time, one of several states of activation. In the case of input neurons, the state of activation is a function of a stimulus. In the case of interneurons and output neurons, their state of activation is a function of the states of activation of other neurons that connect to them. The amount of influence the activation of one neuron can exert on another neuron is determined by the "weight" of the connection between them. Learning in neural networks is typically thought to involve changes to the weights of the connections between neurons (though it may also involve the addition of new connections and the "pruning" of old ones). In feedforward networks, the flow of information is strictly from input to output (via interneurons if any are present). In recurrent networks there are feedback (or "recurrent") connections as well as feedforward connections. (For further discussion of artificial neural networks, see Garson 2002.)

Let us turn now to Churchland's account of how the four elements of the dynamical

profile of conscious states might be realized in recursive neural networks. It helps to begin with Churchland's notion of the conceptual interpretation of sensory inputs and we do well to begin with what Churchland thinks a concept is. Consider a connectionist network with one or more hidden layers that is trained to categorize input types. Suppose that its inputs are a retinal array to which we present grayscale images of human faces. Suppose that its outputs are two units, one indicating that the face is male and the other indicating that the face is female. After training, the configuration of weights will be such that diverse patterns of activation in the input layer provoke the correct response of "male" to the diversity of male faces and "female" for female faces. For each unit in the hidden layer, we can represent its state of activation along one of several dimensions that define activation space. A pattern of hidden layer activation will be represented as a single point in this space. This space will have two regions: one for males and one for females. Regions in the center of each of the two spaces will constitute "attractors" that define what, for the network, constitutes prototypical female faces and prototypical male faces, respectively.

The addition of recurrent connections allows for information from higher layers to influence the responses of lower layers. As Churchland puts the point:

> This information can and does serve to "prime" or "prejudice" that neuronal population's collective activity in the direction of one or other of its learned perceptual categories. The network's cognitive "attention" is now preferentially focused on one of its learned categories at the expense of the others. (Churchland 2002, p. 75)

Churchland is not explicit about what this might mean in terms of the example of a face categorization network, but I suppose what this might mean is that if the previous face was a prototypical female, then the network might be more likely to classify an ambiguous stimulus as female. We can construe this as exogenous cueing of attention. Churchland goes on to further describe shifts of attention in recurrent networks that we might regard as endogenous. "Such a network has an ongoing *control* of its topical selections from, and its conceptual interpretations of, its unfolding perceptual inputs" (p. 76).

Recurrent connections allow for both a kind of short-term memory and the representation of events spread out over time. In a feedforward network, a single stimulus event gives rise to a single hidden layer response, then a single output response. With recurrence however, even after the stimulus event has faded, activity in lower layers can be sustained by information coming back down from higher layers, and that activity can itself reactivate higher layers. Also, what response a given stimulus yields depends in part on what previous stimuli were. Thus, recurrent connections implement a memory. Decreasing connection weights shorten the time it takes for this memory to decay. The ability to hold on to information over time allows for the representation of events spread out over time, according to Churchland, and the representation in question will not be a single point in activation space but a trajectory through it.

Churchland (2002) does not go into much neuroanatomical or neurophysiological detail, but adverts, though tentatively, to the account in Churchland (1995) wherein he endorses Llinas's view whereby consciousness involves recurrent connections between the thalamus (a bilateral structure at the rostral tip of the brainstem) and cortex. Part of the appeal of localizing consciousness in these structures presumably involves the role hypothesized for recurrence as well as the ideas that consciousness involves systems responsible for wakefulness and arousal (thalamus), diverse "higher" functions (the various portions of

the cortex), and a system that can act as a relay between the various "higher" functions (the thalamus again).

I will have more to say about this later, but for now we may briefly summarize Churchland's dynamic profile acount with respect to the three questions of consciousness as follows. With respect to the question of state consciousness, according to Churchland, conscious states are neural representations that have a particular dynamic profile. With respect to the question of transitive consciousness, Churchland's account imposes no limitations on what one can be conscious of; one could be conscious of just about anything according to Churchland. With repect to the question of phenomenal character, "what it's like" to have a conscious state is going to be determined by the representational content of that state. More will be said about these points after we have had the opportunity to examine some other neurophilosophical theories of consciousness.

The neurophilosophical account of consciousness by Prinz (2000, 2004) is relatively similar and fills in a lot of neuroanatomy and neurophysiology that Churchland leaves out. (For further detail see Prinz, chapter 19.) Prinz characterizes the processing hierarchy we discussed earlier and then notes that the contents of consciousness seem to match it with representations at the intermediate level of processing (areas V2–V5). This means that the contents of conscious states do not abstract entirely from points of view as does the highest level of the processing hierarchy, but neither are they the same as the representations at the lowest level. However, Prinz argues that intermediate representations are alone insufficient for consciousness. They must additionally be targeted by attention. Prinz thinks attention is required because of considerations having to do with the pathology of attention known as "neglect." Prinz cites Bisiach's (1992) study of neglect patients who were able to demonstrate certain kinds of unconscious recognition. Prinz infers from such results that not only did high-level areas in the visual hierarchy become activated (they are necessary for the kinds of recognition in question) but also that intermediate levels had to have been activated. Prinz seems to be assuming that information can only get to higher levels of cortical processing by way of the intermediate level, but one wonders if perhaps the intermediate level was bypassed via a subcortical route.

Given the large role that Prinz assigns to attention in his theory of consciousness, the question naturally arises as to what Prinz thinks attention is and what it does. Prinz endorses the account of attention by Olshausen, Anderson, and van Essen (1994), wherein attention involves the modulation of the flow of information between different parts of the brain. Furthermore, Prinz endorses the speculation that the attention crucial in making intermediate-level representations conscious, involves a mechanism whereby information flows from intermediate areas, through high-level visual areas (infero-temporal cortex) to working memory areas in the lateral prefrontal cortex. Pieces of information in working memory, "allow the brain to recreate an intermediate-level representation by sending information back from working memory areas into the intermediate areas" (2004, p. 210). Prinz (2000) summarizes, emphasizing attention's role, as follows:

> When we see a visual stimulus, it is propagated unconsciously through the levels of our visual system. When signals arrive at the high level, interpretation is attempted. If the high level arrives at an interpretation, it sends an efferent signal back into the intermediate level with the aid of attention. Aspects of the intermediate-level representation that are most relevant to interpretation are neurally marked in some way, while others are either unmarked or suppressed. When no interpretation is achieved (as with fragmented images or cases of agnosia), attentional mechanisms might be deployed somewhat differently. They might "search" or "scan"

> the intermediate level, attempting to find groupings that will lead to an interpretation. Both the interpretation-driven enhancement process and the interpretation-seeking search process might bring the attended portions of the intermediate level into awareness. This proposal can be summarized by saying that visual awareness derives from Attended Intermediate-level Representations (AIRs). (p. 249)

Prinz's account of attention's role in consciousness seems a lot like Churchland's conceptual interpretation, short-term memory, and of course, attention requirements on consciousness. Tye raises objections to the sort of view advocated by Churchland and Prinz. Tye is critical of accounts of consciousness that build in constitutive roles for attention. Tye's claim is based on introspective grounds (1995, p. 6). The thought here is that one might have a pain for a length of time but not be attending to it the entire time. Tye insists that there is still something it's like to have an unattended pain. Tye infers from these sorts of considerations that the neural correlate of visual consciousness is lower in the processing hierarchy than an attention-based theory would locate it. Tye thus locates the neural correlates of conscious states in "the grouped array" located in the occipital lobe and, regarding the phenomenon of blindsight, rejects "the hypothesis that blindsight is due to an impairment in the linkage between the spatial-attention system and the grouped array" (Tye 1995, pp. 215–16) Tye accounts for the retained visual abilities of blindsight subjects (p. 217) in terms of a "tecto-pulvinar pathway" from retina to superior coliculus that continues through the pulvinar to various parts of the cortex, including both the parietal lobe and area V4. Thus, Tye seems to think consciousness is in V1. Prinz (2000) argues against this, citing evidence against locating consciousness in V1 (see Crick & Koch 1995 and Koch & Braun 1996 for reviews). Prinz writes:

> As Crick and Koch emphasize, V1 also seems to lack information that is available to consciousness. First, our experience of colors can remain constant across dramatic changes in wavelengths (Land 1964). Zeki (1983) has shown that such color constancy is not registered in V1. Second, V1 does not seem responsive to illusory contours across gaps in a visual array (von der Heydt, Peterhans, & Baumgartner 1984). If V1 were the locale of consciousness, we would not experience the lines in a Kanizsa triangle. (pp. 245–6)

Turning from disagreements to agreements, we may note that Churchland, Prinz, and Tye all agree that conscious states are representational states. They also agree that what will differentiate a conscious representation from an unconscious representation will involve relations that the representation bears to representations higher in the processing hierarchy. For both Churchland and Prinz, this will involve actual interactions, and further, these interactions will constitute relations that involve representations in processes of attention, conceptual interpretation, and short-term memory. Tye disagrees on the necessity of actually interacting with concepts or attention. His account is "dispositional," meaning that the representations need only be poised for uptake by higher levels of the hierarchy.

Turning to the question of transitive consciousness, we see both agreements and disagreements between the three authors. Churchland, Tye, and Prinz all agree that what one is conscious of is the representational content of conscious states. In all cases, what the subject is conscious of is what the representational contents of the conscious states are. However, these theorists differ somewhat in what they think the contents can be. Churchland has the least restrictive view: any content can be the content of a conscious state. Prinz's is more restrictive: the contents are not going to include high-level invariant contents. Tye's is the

most restrictive: the contents will only be first-order and non-conceptual. Tye thinks that they are non-conceptual since he thinks that creatures without concepts – perhaps non-human animals and human infants – can have states for which there is something it's like to have them even though they possess no concepts. Tye says little about what concepts are, and for this, among other reasons, it is difficult to evaluate his view. The reason Tye thinks the contents of consciousness are first-order is because he believes in the pre-theoretic obviousness of the transparency thesis whereby when one has a conscious experience, all that one is conscious of is what the experience is an experience of. Thus, if one has a conscious experience of a blue square, one is only aware of what the mental state represents – the blue square. One is not, Tye insists, able to be conscious of the state itself. So, for example, if the state itself is a pattern of activity in one's nervous system, one will not be able to be conscious of this pattern of activity, but only be able to be conscious of external world properties that the pattern represents. Mandik (2005, 2006) argues that Churchland's (1979) thesis of the direct introspection of brain states provides the resources to argue against the kinds of restrictions on content that Tye makes.

I will not spell out the full argument here, just indicate the gist of it. Conceptual content can influence what it's like to have a particular experience. What it is to look at a ladybug and conceive of it as an example of *Hippodamia convergens* is, intuitively, quite different from what it would be like to conceive of it as one's reincarnated great-great-grandmother. Thus, if a person had the conceptual knowledge that consciously perceiving motion involved activity in area MT, and acquired the skill of being able to automatically and without conscious inference apply that conceptual knowledge to experience, then that person would be able to be conscious of the vehicular properties of that experience.

I turn now to what neurophilosophical accounts have to say about phenomenal character. I focus, in particular, on the suggestion that phenomenal character is to be identified with the representational content of conscious states. I will discuss this in terms of Churchland's suggestion of how qualia should be understood in terms of neural state spaces.

Our experience of color provides the most often discussed example of phenomenal character by philosophers, and Churchland is no exception. When Churchland discusses color qualia, he articulates a reductive account of them in terms of Land's theory that human perceptual discrimination of reflectance is due to the sensory reception of three kinds of electromagnetic wavelengths by three different kinds of cones in the retina (Land 1964). In keeping with the kinds of state-space interpretations of neural activity that Churchland is fond of, he explicates color qualia in terms of points in three dimensional spaces, the three dimensions of which correspond to the three kinds of cells responsive to electromagnetic wavelengths. Each color sensation is identical to a neural representation of a color (a neural representation of a spectral reflectance). Each sensation can thus be construed as a point in this 3-D activation space and the perceived similarity between colors and the subjective similarities between corresponding color qualia are definable in terms of proximity between points within the 3-D activation space. "Evidently, we can reconceive [sic] the cube [depicting the three dimensions of coding frequencies for reflectance in color state space] as an internal 'qualia cube'" (1989, p. 105). Churchland thinks this approach generalizes to other sensory qualia, such as gustatory, olfactory, and auditory qualia (ibid., pp. 105–6). Bringing this view in line with the thesis of the direct introspection of brain states, Churchland writes:

> The "ineffable" pink of one's current visual sensation may be richly and precisely expressible as a 95 Hz/80 Hz/80 Hz "chord" in the relevant triune cortical system. The "unconveyable"

> taste sensation produced by the fabled Australian health tonic Vegamite [sic.] might be quite poignantly conveyed as a 85/80/90/15 "chord" in one's four-channeled gustatory system (a dark corner of taste-space that is best avoided). And the "indescribable" olfactory sensation produced by a newly-opened rose might be quite accurately described as a 95/35/10/80/60/55 "chord" in some six-dimensional system within one's olfactory bulb.
>
> This more penetrating conceptual framework might even displace the common-sense framework as the vehicle of intersubjective description and spontaneous introspection. Just as a musician can learn to recognize the constitution of heard musical chords, after internalizing the general theory of their internal structure, so may we learn to recognize, introspectively, the *n*-dimensional constitution of our subjective sensory qualia, after having internalized the general theory of *their* internal structure. (Ibid., p. 106)

Three particular and related features of Churchland's view of qualia are of special note. The first is that qualia are construed in representational terms. The second follows from the first, namely, that qualia so construed are not intrinsic properties of sensations, and thus overturns a relatively traditional view of qualia. The third is that it allows for intersubjective apprehensions of qualia. To see these points more clearly it will be useful to briefly examine the traditional account of qualia noting the role of supposedly intrinsic properties in the account.

It is difficult to say uncontroversial things about qualia; however, there are several points of agreement among many of those philosophers who believe that mental states have such properties. These philosophers describe qualia as (i) intrinsic properties of conscious states that (ii) are directly and fully knowable only by that subject and (iii) account for "what it's like" for a subject to be in that state. More briefly, qualia are (i) intrinsic, (ii) subjective, and (iii) there is "something it's like" to have (states with) them. Less briefly, we can start with (iii) and work our way to (i) as follows. When I have a conscious perception of a cup of coffee there is, presumably, something it's like for me to have that perception and, for all I know, what it's like for you to have a conscious perception of a cup of coffee is quite different. Furthermore, for all that you can tell me about your experience, there is much that cannot be conveyed and thus is subjective, that is, directly and fully knowable only by you alone. The supposition that qualia are intrinsic properties of conscious states serves as a possible, though questionable, explanation of their subjectivity. (See Mandik 2001 for a neurophilosophical account in which subjectivity is consistent with qualia being extrinsic.) The inference from subjectivity to the intrinsic nature of qualia may be articulated as follows. If something is defined by the relations that it enters into, then it is fully describable by the relations it enters into, and if it is not fully describable by the relations it enters into, it must not be defined by the relations it enters into.

To construe qualia in terms of representational content, however, is to construe them as no longer intrinsic, since typical accounts will spell out representational content in terms of:

1 causal relations that sensory states bear to states of the external world;
2 causal relations that they bear to other inner states; or
3 some combination of the two sorts of relations.

In neural terms, a pattern of activation in a neural network is the bearer of representational content in virtue of:

1 the distal or proximal stimuli that elicit the activation;
2 other patterns of activation that influence it via, e.g., recurrent connections; or
3 some combination of the two.

While it is relatively clear how Churchland's view is supposed to rule out the view of qualia as being intrinsic, it is not so clear that it is equally able to rule out their being subjective. The above quoted passage contains Churchland's view that properties of neural states previously inexpressible could, if one acquired the relevant neuroscientific concepts and the skill to apply them introspectively, become expressible. However, this view seems to be in tension with the earlier-mentioned view that concepts influence phenomenal character. The phenomenal character of an experience prior to the acquisition and introspective application of a concept will not, then, be the same as the phenomenal character of an experience after the acquisition and introspective application of that concept. Thus, even within a general neurophilosophical view of consciousness, there may remain certain representational contents of neural states that are directly and fully knowable only by the subject who has them. Neurophilosophy, then, may be fully compatible with the subjectivity of phenomenal consciousness.

See also 2 Philosophical problems of consciousness; 13 The case of blindsight; 19 The intermediate level theory of consciousness; 44 A neurobiological framework for consciousness; 48 Duplex vision: separate cortical pathways for conscious perception and the control of action.

Further Readings

Churchland, P. S. (1986) *Neurophilosophy*. Cambridge, MA: MIT Press.
Churchland, P. M. (1989) *A Neurocomputational Perspective*. Cambridge, MA: MIT Press.
Churchland, P. M. (2002) Catching consciousness in a recurrent net. In A. Brook, and D. Ross (eds.), *Daniel Dennett: Contemporary Philosophy in Focus*, 64–80. Cambridge: Cambridge University Press.
Prinz, J. (2000) A neurofunctional theory of visual consciousness. *Consciousness and Cognition* 9, 243–59.

References

Akins, A. (1996) Of sensory systems and the "aboutness" of mental states. *Journal of Philosophy* 93, 337–72.
Bisiach, E. (1992) Understanding consciousness: clues from unilateral neglect and related disorders. In A. D. Milner and M. D. Rugg (eds.), *The Neuropsychology of Consciousness*, 113–39. London: Academic Press.
Block, N. (1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 2, 227–88.
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*. New York: Oxford University Press.
Churchland, P. M. (1979) *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Churchland, P. M. (1989) *A Neurocomputational Perspective*. Cambridge, MA: MIT Press.
Churchland, P. M. (1995) *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, MA: MIT Press.
Churchland, P. M. (2002) Catching consciousness in a recurrent net. In A. Brook, and D. Ross (eds.), *Daniel Dennett: Contemporary Philosophy in Focus*, 64–80. Cambridge: Cambridge University Press.
Churchland, P. S. (1986) *Neurophilosophy*. Cambridge, MA: MIT Press.
Crick, F. and Koch, C. (1995) Are we aware of activity in primary visual cortex? *Nature* 375, 121–3.
Garson, J. (2002) Connectionism. In Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Winter 2002 edn.), <http://plato.stanford.edu/archives/win2002/entries/connectionism/>.
Koch, C. and Braun, J. (1996) Towards a neuronal correlate of visual awareness. *Current Opinion in Neurobiology* 6, 158–64.
Land, E. H. (1964) The retinex. *Scientific American* 52, 247–64.
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, A. Elitzur, and M. Saniga (eds.), *Endophysics, Time, Quantum and the Subjective*. Singapore: World Scientific Publishing Co.
Mandik, P. (2006) The introspectability of brain states as such. In B. Keeley (ed.), *Paul M. Churchland: Contemporary Philosophy in Focus*. Cambridge: Cambridge University Press.
Milner, A. and Goodale, M. (1995) *The Visual Brain in Action*. New York: Oxford University Press.
Olshausen, B. A., Anderson, C. H., and van Essen, D. C. (1994) A neurobiological model of visual attention and invariant pattern recognition based task. *Journal of Neuroscience* 14, 6171–86.
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*, 359–80. New York: Oxford University Press.
Pascual-Leone, A. and Walsh, V. (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–12.
Pöppel, E., Held, R., and Frost, D. (1973) Residual visual functions after brain wounds involving the central visual pathways in man. *Nature* 243, 295–6.
Prinz, J. (2000) A neurofunctional theory of visual consciousness. *Consciousness and Cognition* 9, 243–59.
Prinz, J. (2004) *Gut Reactions*. New York: Oxford University Press.
Quine, W. (1969) Epistemology naturalized. In *Ontological Relativity and Other Essays*, 69–90. New York: Columbia University Press.
Rosenthal, D. (1993) State consciousness and transitive consciousness. *Consciousness and Cognition*, 2: 4 (December), 355–63.
Stoerig, P. and Cowey, A. (1992) Wavelength discrimination in blindsight. *Brain* 115, 425–44.
Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984) Illusory contours and cortical neuron responses. *Science* 224, 1260–2.
Weiskrantz, L. (1995) Blindsight: not an island unto itself. *Current Directions in Psychological Science* 4, 146–51.
Weiskrantz, L. (1996) Blindsight revisited. *Current Opinions in Neurobiology* 6: 2, 215–20.
Zeki, S. (1983) Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colour. *Neuroscience* 9, 741–56.