The brain and its states

Richard Brown
LaGuardia Community College, NY, USA

1. Introduction

In recent times we have seen an explosion in the amount of attention paid to the conscious brain from scientists and philosophers alike. One message that has emerged loud and clear from scientific work is that the brain is a dynamical system whose operations unfold in time. Any theory of consciousness that is going to be physically realistic must take account of the intrinsic nature of neurons and brain activity. This important idea is often taken to be in conflict with the more traditional way of thinking about the mind in terms of static states like beliefs, pains, or representations of blue. One of the aims of this chapter is to show that this is not the case. We must use the traditional theories as a way to pick out the brain activity that we are interested in. In this way the two ways of thinking depend on each other. To corrupt a common Kantian line: uninterpreted brain data is meaningless and interpretation without brain data is blind. Once we adopt this model of scientific explanation and reduction we can build a case that phenomenal consciousness may turn out to be nothing but patterns of synchronized neural activity in specific frequencies against a dynamically changing chemical background.

To give some of the overall structure of my argument, in the next section I distinguish between Creature, Transitive, State, Access, and Phenomenal consciousness. Creature consciousness could plausibly turn out to be nothing but the global chemical state of the brain (in particular the exact ratio of aminergic to cholinergic neuromodulators) and transitive consciousness to be nothing but synchronized neural activity in various frequencies. Once we have identified these two fundamental kinds of consciousness in the brain the other three can be reduced to transitive consciousness. To see how phenomenal consciousness could be reduced to a kind of transitive consciousness I introduce the Higher-Order Representation of a Representation (HOROR) theory of phenomenal consciousness, which is a variant of the traditional higher-order thought theory. The main difference consists in the claim by the HOROR theory that phenomenal consciousness just is a kind of representation, albeit a higher-order representation (of a suitable kind).

As such there is no relation between the higher-order representation and any other representation needed. This distinguishes the HOROR theory from at least some versions of the traditional higher-order thought theory.

The HOROR theory strikes many people as counter-intuitive in that many people wonder how it could possibly be the case that the conscious experience of a pain – the painfulness of the pain – could be a higher-order representation. I agree that it is counter-intuitive, but it may be right nonetheless. After introducing the HOROR theory I offer some empirical evidence that suggests that we do have conscious experience in the absence of the appropriate first-order states and that disrupting activity in pre-frontal areas (where higher-order representations presumably live) produces a blindsight-like state in normal subjects. Taken together these empirical results show that the HOROR theory is empirically viable despite our intuitions.

If the HOROR theory is right then phenomenal consciousness is nothing but a particular kind of brain activity. This is because phenomenal consciousness is a kind of representation and representations of the right kind turn out to be patterns of synchronized neural activity (possibly in the frontal areas of the brain). What this means is that I will be arguing for what philosophers call a type-type identity theory. The type-type identity theory claims that mental state types are identical to physical state types. After some fancy footwork, which we can avoid here, it can be shown that this amounts to the claim that brains are necessary, and patterns of synchronized neural activity against global chemical background states are sufficient, for consciousness. If the type-type identity theory is right then if there is no brain you do not have mental states or consciousness at all.

Some philosophers might have thought that an identity theory in terms of dynamical states of the brain is inconsistent with any kind higher-order theory, which on their face seem to allow that there may be many ways to have a higher-order representation (in particular ways that are not biological). As in the above case I think it is an empirical question whether we can actually make synthetic or artificial consciousness. It may turn out that something distinctly biological is required for consciousness. If this turns out to be right then we can see our various folk-psychological platitudes about the various kinds of consciousness as a way of picking out or identifying what in the brain we are interested in. On this way of thinking it is the states in the brain that are the representations in question. This kind of view allows powerful responses to various anti-physicalist arguments like Descartes' argument from the conceivability of the distinctness of his mind from his body as well as the more recent arguments based on zombies and the possibility of machine consciousness.

The questions addressed in this chapter are still wide open and there may be many possible routes to physicalism being true. At heart, I am an optimist about the prospects for a complete account of consciousness in physical terms. At the

very least, I hope that the arguments in this chapter can help to support that optimism. I would not go so far as to say that we know that consciousness is a dynamical brain process but I do think that – for all we know – it could be.

2. Some concepts of consciousness

The territory here is at this point well worn so I will provide only a brief exposition of the various concepts of consciousness that I will discuss (for more detailed discussion see Block 1995 & Rosenthal 2005).

Creature consciousness consists in a creature being awake and responding to stimuli. When a creature is unconscious it is not responding to stimuli.

Transitive consciousness consists in our being conscious of things and so consists in sensation, perception, and thinking. A state is transitively conscious when it makes us aware of something in the world. To be aware of something is to be informationaly responsive to it.

State consciousness is a property of mental states. Unconscious mental states lack this property and conscious mental states have it. Intuitively a conscious mental state is one that we are, in some way, aware of ourselves as being in.

Access consciousness, as understood by Ned Block (1996), amounts to the idea of a global workspace (Baars 1988; Dehaene & Naccache 2001). That is to say that a state is access conscious when it is broadcasted in such a way so as to be available for the use in reasoning or action.

Phenomenal consciousness is the property of there being something that it is like for one to have a conscious mental state. When a state is phenomenally conscious there is a distinctive way that my experience seems to be. So, when I am phenomenally conscious of, say, a blue patch, there is a particular way that things appear to me. We might say things appear blue but we must also note that the blueness is for me in a particular way. I experience the blue as mine.

This brief survey does not aim to be exhaustive but rather is meant to delineate the topics of the chapter. One concept that I will not discuss in this chapter is the notion of self-consciousness understood as the awareness of oneself as a self.

We can see that there are relationships between these various concepts. For instance, when a creature is conscious it may be in some mental states that are conscious but also be in some mental states that are unconscious. So too when the creature is unconscious it may turn out to be the case that the creature has conscious mental states (perhaps while dreaming). These conscious states may be ones that there is something that it is like for the creature that has them but we must also leave open the possibility that some mental states will not be like anything for the creature that has it.

I will now turn to discussing each of these notions and its relation to the brain.

3. Creature consciousness

Asking the question 'when is a creature conscious?' is just asking when is the creature alert and awake? There has to date been a lot of work done on the neurophysiology of sleep and wakefulness. Brainstem areas are implicated in regulating the brain between its waking and sleeping phases by controlling the kinds of neuromodulators that are being released, thereby controlling which neurons are active and inactive. In the awake state the brain is aminergic, meaning that neurons that use histamine, noradrenaline and serotonin are active, and cholinergic, meaning that neurons that use acetylcholine and dopamine are active, in the REM state it is only cholinergic, meaning that noradrenaline, serotonin, and histamine neurons are offline, and in NREM the milieu is chemically intermediate between the two (Hobson 2009: 810).

Hobson and his collaborators have developed a state space model that is very useful. They call it the AIM model (Kahn et al. 1997; Hobson 2009) which allows the state of the brain to be mapped by the amount of activation as measured by global EEG (A), the flow of information as measured by the level of input-output gating (I) and neuromodulatory effects as measured by the excitability of spinal neurons (M). This allows them to distinguish a state space in which we can see waking, sleep, and dreaming each occupying a unique place. For instance when a creature is awake you will find high levels of activation and neuromodulation (indicating that the brain is aminergic) and low values of input-output gating (which is just to say that there is mental activity in response to input and actions being generated). When a creature is asleep we will see a large amount of input-output gating (dreaming of running does not usually lead to moving one's legs) and low levels of amenirgic activity.

Hobson argues that the AIM state space approach provides a natural way to understand states besides waking, non-dreaming sleep, and dreaming. He says,

> The 'state space' approach also enables the mapping of exceptional mental states such as lucid dreaming and abnormal conditions such as coma and minimally conscious states. Sleep disorders such as narcolepsy and many psychiatric syndromes (such as depression) also find their place in the AIM state space. (Hobson 2009: 810)

We might hypothesize that mood in general can be analyzed in this way. If so then we can see that states of creature consciousness are nothing but global chemical states of the brain. These chemical states are in a constant state of flux throughout the creature's existence. This is to say that they unfold in time and so can be considered dynamical systems.

Hobson also distinguishes between what he calls primary and secondary consciousness. Primary consciousness on his usage amounts to perception and emotion while secondary consciousness is defined as self-reflective awareness, abstract thinking and metacognition (Hobson 2009: 803). These notions will count as varieties of transitive consciousness according to the present schema since these are all versions of being conscious of something.

4. Transitive consciousness

Transitive consciousness involves being aware of objects (as) in our environment. Thus when we see a red square we are transitively conscious of red and square and perhaps even that the square is red. Transitive consciousness thus involves what we would normally call sensation, perception and thought. This term is unusual but was introduced by David Rosenthal (2005) as a way to capture the fact that this kind of consciousness always involves being conscious of something (the 'of' there is what grammarians would call transitive verb since it takes an object).

The difference between transitive consciousness and creature consciousness highlights a distinction between what we might call states of consciousness and conscious states, which in turn mirrors the distinction between states of the brain and brain states. A state of the brain is a global state which just is the total ratio of aminergic versus cholinergic neuromodulators in the brain at any given moment in time. A conscious state, on the other hand, is a particular state. This is mirrored in the brain by talk about particular brain states. We will be talking about particular representations that can be instantiated against different background states of the brain. Thus we may have the same perception or thought against the very different states of an awake brain versus a dreaming brain.

Wolf Singer is well known for arguing that synchrony in a frequency may be a general strategy that the brain uses to represent various features of objects. In my 2006 paper (Brown 2006) I argued that this general notion could be extended to a hypothesis about the nature of brain states (as opposed to states of the brain). I won't here repeat the experimental data from Singer (1996, 2000), which by this point is well known.

We can extend this idea to offer a viable account of how various mental processes can be understood in terms of synchrony. For instance Gyorgy Buzsaki (1989, 1996) and his lab have argued that we can understand memory in terms of synchrony in various frequencies. Buzsaki assumes that we can understand neural representation in Singerian terms as synchronized neural activity in the gamma range. In the hippocampal formation this gamma activity is modulated by a theta

rhythm that serves to put neurons in 'the appropriate context to receive information'. In essence neurons may 'tune' in to the information that is being broadcast in the theta frequency. At the neural level this happens because the theta rhythm serves to keep the entorhinal neuron's membrane voltage close to but below firing threshold (Buzsaki 1996: p 83). For the full story I would refer the reader to my 2006 paper.

This idea, that some rhythms encode information, or represent things, while others serve as 'carrier' signals on which the information is broadcast and that disparate brain areas can 'tune' in to the information by becoming synchronized in the broadcast frequency, has been recently defended by Edvard Moser and his lab. The gamma frequency comprises a wide swath of frequencies from about 25 hertz all the way to about 150 hertz. The Moser lab has been able to show that neurons can selectively synchronize in either slow or fast frequencies which are themselves phase-locked to different cycles of the theta rhythm (Colgin et al. 2009).

If the foregoing considerations are right, then we can postulate that access consciousness is also nothing but long-range synchronization between different parts of the brain. Neurons in, say CA1 which are firing in synchrony and thereby representing the orientation, say, of something are being broadcast through the hippocampal formation on the theta frequency, meaning that they are phase locked to particular cycles of the theta frequency, then neurons in a later processing stage, CA2, can tune in, or access, that representation by becoming synchronized in the theta frequency, which then it turn disposes those neurons to become synchronized in the gamma range, thereby acquiring the representation.

While in no way conclusive the foregoing empirical theories seem well suited for giving a satisfying physicalistic account of these two fundamental kinds of consciousness in terms of two distinct kinds of dynamic activity of the brain. In the rest of this paper I will argue that the other two notions of consciousness, state consciousness and phenomenal consciousness, can be understood in terms of transitive consciousness. On my view transitive consciousness just is a group of neurons that are firing in sync and this in turn means that the other two kinds of consciousness to be discussed will turn out to be nothing more than this kind of brain activity.

5. State consciousness

At one point in time it was common to assume that all mental states are conscious. Descartes famously argued that the mind was immediately transparent to the person whose mind it was. If I thought, felt, saw, or otherwise experienced

something then I knew that I did. And if I sincerely took myself to be having a thought or experience then I was indeed having that thought or experience. But this is no longer commonplace. We have learned that mental states can occur consciously and that they can occur unconsciously as well. In fact one of the striking discoveries of contemporary cognitive science is just how much of what we do can be done by unconscious processes.

At this point we come to the distinction between first-order and higher-order theories. A first-order theorist, like Fred Dretske (1993), will identify state consciousness with transitive consciousness. A conscious state, they will hold, is one by which I am conscious of something in my environment. One major problem with any first-order view is that there seems to be cases of mental states that are unconscious yet in virtue of which we are aware of something. Classic examples come from priming and masking studies as well as change blindness. In those kinds of cases we have evidence for a mental representation that enables us to perform some task, say completing a word or identifying something quicker, but of which we are completely unaware.

The same kind of problem arises for someone who wants to see state consciousness as merely access consciousness. When we are primed to pick out a red car by being subliminally presented with red, say, the state that represents red is access conscious. It is widely available for control of action (that is why we are primed). Yet it will seem to me as though I saw nothing.

Higher-order theories claim that a mental state's consciousness consists in having a suitable higher-order awareness of being in that state (Armstrong 1968, Rosenthal 2005; Lycan 1996). This amounts to the claim that state consciousness is to be explained in terms of some suitable kind of transitive consciousness. The difference between first-order theorists and higher-order theorists lies in what kind of transitive consciousness is needed. For the higher-order theorist the transitive consciousness must be consciousness of oneself as being in some first-order state. These kinds of theories are divided between higher-order perception and higher-order thought being the right kind of awareness. Hereon we will ignore that distinction.

We cannot settle that debate here but many people find a higher-order theory of state consciousness to be plausible. There is a case to be made that some kind of higher-order theory is part of our common sense thinking about the mind. Intuitively we do not call any mental state of which we are completely unaware a conscious state. If you are completely unaware of believing something what sense is there in calling that state conscious? Granted, the state will be an instance of transitive consciousness, which is to say that it will make me aware of something, but that is not what we mean we talk about state consciousness.

We can have conscious states of consciousness as well as conscious mental states. Thus I can have a mood of which I am not aware myself as being in as well

as a belief of which I am not aware of having. Being aware of these states involves a kind of transitive consciousness and so we can see that state consciousness will turn out to be, on the present view, a kind of synchronized neural activity. In particular it will be the neural activity that is responsible for my being informationally responsive to my own first-order brain states.

6. Phenomenal consciousness

There are many physicalistic theories of phenomenal consciousness but broadly speaking they fall into two categories. There are first-order theorists who see phenomenal consciousness as a particular kind of representation of the world (Tye 2000; Byrne 2001) but there are also first-order theories that see phenomenal consciousness as identical to some brain state even though they go on to deny that the state is representational in any way (Block 1996). The water is muddied here by the view that a state may be representational even if it is not conceptual. I will adopt the inclusive view on which a state can be a representation even if the state involves no concepts. In this sense even Block will think that there are first-order representations of red and that these representations just are what phenomenal consciousness turns out to be.

Thus the question for us is whether we should think of phenomenal consciousness as being identical to first-order representations or higher-order representations. Some philosophers have taken the higher-order theory to be a theory of phenomenal consciousness (Weisberg 2011; Rosenthal 2005). On this view having a conscious pain consists in having a suitable higher-order state that represents the first-order state.

It seems to me that both views are possible and that it is largely a matter of empirical fact which of these turns out to be true. It also seems to me that the balance of evidence is tipped in favor of some kind of higher-order theory. I will develop one such theory that I call the Higher-Order Representation Of a Representation (HOROR) theory of phenomenal consciousness and survey the empirical evidence for it. According to this theory what it is like to consciously experience red is identical to the having of a certain kind of higher-order representation (as usual I assume that the right kind of higher-order representation is one that is seemingly unmediated by inference, etc). This is to say that phenomenal consciousness just is a certain kind of higher-order representation. In particular it is identical to a representation of oneself as having some world-directed (i.e. first-order) representation.

The HOROR theory of phenomenal consciousness has similarities and differences from the traditional forms of both first-order and higher-order theories. It

s similar to a first-order view in that it claims that phenomenal conscious just is a kind of representation. Higher-order theories of phenomenal consciousness claim that the first-order representation of red is not enough for phenomenal consciousness. One needs also to have a higher-order awareness of the first-order state. Thus the traditional higher-order thought and perception views seem to be committed to their being a relationship between the first-order and the higher-order state. This is what has caused some to worry about cases where there is no first-order state (Block 2011). In those cases there would seem to be nothing for the relationship to hold between. This is where the advantages of the HOROR theory come in. According to that view it is not that we must have a relationship between the first-order state and the higher-order awareness. Rather phenomenal consciousness just is the higher-order representation.

The HOROR theory is a theory about phenomenal consciousness. That is, it is a theory about which mental states posses the property of there being something that it is like for the creature to have the states in question. It is not a theory of state consciousness. State consciousness consists in being aware of yourself as being in some state (whether a mental state or a state of mind) or other. It will then be the case that the conscious state is the one of which you are aware of yourself as being in. It is the target of the higher-order representation. But the phenomenally conscious state is not the target of the higher-order representation; it just is the higher-order representation. Thus HOROR theory explicitly denies that any kind of relation is required for phenomenal consciousness. Some versions of the higher-order thought theory do claim this and so this will be a way that the HOROR theory is different from tradition higher-order thought theories.

The higher-order representation is not introspectively conscious – for that it would need to have a third order state targeting it – but it is phenomenally conscious. It is the state in virtue of which there is something that it is like for the subject.

Many people will find the HOROR theory of phenomenal consciousness to be extremely counter-intuitive. Why should we think that phenomenal consciousness just is the having of a representation of a representation? There are roughly two lines of argument that aim to show that this might be the case. The first is the one that Rosenthal has developed (2005) and is based on the role that concepts play in phenomenology and the second is based on recent empirical work from Hakwan Lau's lab (Rahnev et al. 2011).

First we should note the connection between the HOROR theory and phenomenal consciousness. When a state is phenomenally conscious there is something that it is like for me to have that state. It makes sense to think that this entails some kind of awareness must be a part of phenomenal consciousness. If I am experiencing a state as mine, as for me, then there must be something that accounts

for the fact that I do experience it as such. First-order views face the challenge of explaining why some first-order representation would result in there being anything at all that it is like *for* the creature with the representation. If I have a first-order representation of red then I can explain why I am aware of red in the world but it will be a mystery why I take the experience to be mine or for me in the right way. But the HOROR view exploits the resources of higher-order theories to meet this challenge. The specific nature of the higher-order representation involves one representing that one is, oneself, in a particular state. Thus we can explain why a state will be experienced as for the creature. Since one represents oneself as being in a certain state it will seem to you as though you are in that state. Given this we can easily see why having a higher-order representation of a representation would make it the case that there is something that it is like for me. This is a major explanatory advantage of the HOROR theory.

Many will find this picture inviting because there is a straightforward sense in which if one believes that p is the case then it will seem to one as though p. But many will balk at the step from the claim that it seems as though p to the creature to the claim that there is something that it is like for it to seem as though p. This is to say that many will think that there is something illicit going on here. There may be a sense in which we can say that it will appear to me as though I am in a certain state by representing myself as being in that state but why should we think that this is the same as the phenomenal sense of appearance? That is, why should we not think that there is a sense in which just having the thought is not enough for the existence of phenomenal consciousness?

Take the case of listening to an orchestra. If one has no concept of what a bass clarinet is one will not consciously experience the sound of the bass clarinet as such, though one's experience of it may be conscious in some other respect (that is to say one will have the relevant first-order states with their qualitative characters and perhaps even higher-order thoughts about them but not as having bass-clarinet* qualities). Once one acquires the concept 'bass clarinet' one's experience is different in a phenomenological way. What it is like for one to hear the orchestra will differ in precisely the sense that it will now sound like there is a bass clarinet in the orchestra to one. The same case can be made for wine tasting. What cases like this give us is data that learning a new concept results in new conscious phenomenology. If concepts can make this kind of difference in our experience then perhaps having them at all can account for the existence of phenomenal consciousness in the first place.

The argument just sketched here may be suggestive but there is a ready response from those who wish to deny the HOROR theory. One may grant that acquiring new concepts is enough to alter one's phenomenology while still denying that concepts can account for the existence of phenomenology. For instance

perhaps the application of concepts actually alters the first-order representations? If one thought that phenomenal consciousness was a property of first-order states, and one thought those states could be changed by conceptualizing them then one could admit that the acquisition of new concepts altered one's phenomenology while denying that phenomenology consists in the application of concepts. This objection may not be decisive but rather than pursue that I will turn to the second, more recent, empirical argument, which is designed to avoid this confound.

This second line of argument is based on the idea that there are empirical cases that seem to suggest that we can in fact have phenomenally conscious experiences in the absence of first-order representations. If we can show that there is genuine phenomenal consciousness without the appropriate first-order activity then we would have strong evidence for some kind of higher-order theory. In other work Hakwan Lau and I (Lau & Brown forthcoming) have discussed three kinds of empirical cases, each serving a slightly different purpose. Here I will focus just on the experimental results from Lau's lab (Rahnev et al. 2009; Rahnev et al. 2011), and I will do this for two reasons. First, I think the other cases we discuss are fairly well understood but the Rahnev et al. results are new and have yet to be fully appreciated. Second, the purpose of each case is slightly different. For instance, we discuss the case of Rare Charles Bonnet syndrome in which subjects with extensive damage to v1 nonetheless report vivid visual hallucinations. This suggests that a particular view about the nature of first-order representations is false (i.e. the view that first-order representations consist in feedback to V1). However it does nothing to show that the first-order representations may not be in other more intermediate brain areas (Prinz 2005). As I will try to show in what follows, the Rahnev experiments are not susceptible to this response.

In this series of experiments subjects were presented with grating patterns or noise in both attended and unattended locations and asked simply whether they saw a target (as opposed to noise) at the probed location. The luminance contrast was adjusted so that performance on the task was matched as between the attended and unattended locations. Using signal detection theory the authors calculated the d' scores, which in effect are simply a measure of how successful subjects are at performing the task. When the d' scores were matched between the two conditions subjects displayed a more liberal detection bias in the unattended location, even though the amount of information, and so presumably first-order representations, was no different (as measured by d'). In a separate condition subjects were asked to discriminate the orientation of the grating (left versus right tilt) and also to judge how visible the stimulus was. Subjects reported higher-visibility ratings for the gratings that were presented in the unattended locations even though they were just as good at discriminating the tilt, as shown by d'.

The above studies use mathematical methods to estimate the amount of information being processed from behavioral data (specifically hits and false alarms). In order to more directly test these issue the authors performed a follow up study using functional magnetic resonance imaging (fMRI). In this follow up study they tracked brain activity in areas that have been implicated in spatial attention. The results showed that when activity in these areas was low, which is thought to correspond to a low state of attention, subjects were more confident that they had seen something in an unattended location. Yet, just as before, their ability to do the task was no better (as measured by d'). In addition to this the authors point out that the average intensity of activity in early visual areas was not higher in either of the two conditions (attended versus unattended). This is what we would expect given that the d' scores are matched. Subjects are performing at the same level and so we would expect to find comparable activity in areas thought to be responsible for first-order representations.

This series of experiments strongly suggests that there can be differences in conscious phenomenology without differences in first-order representations. This is because subjects are telling us that they see something or that it is more visible and yet they are doing no better (or worse) in these cases. If we take their reports at face value then we should allow that gratings in unattended locations are indeed more visible than ones that subjects are attending to. But yet there is no difference in the amount of information being represented by the system (as given by d' and reinforced via the fMRI data from the follow up study). As noted above this suggests that the response from Prinz doesn't affect the Rahnev studies. There appears to be the same amount of information represented in both cases. If there were additional representations in intermediates brain areas in one case but not the other we should not expect to find d' scores that are matched.

It might be objected that this does not actually show that there is a difference in conscious experience since it may be the case that this merely reflects some kind of cognitive bias. However this seems unlikely. In the Rahnev (2011) experiments multiple controls were performed ranging from paying subjects for better performances to giving trial by trial feedback on performance to adjusting the stimulus length/contrast to tracking eye movements to assure that subjects were performing the task correctly. None of these controls destroyed the effect and this suggests that the effect is actually due to perception. If it were merely a cognitive bias then we would expect that it could be trained away. But it is resistant to these kinds of tactics.

The above considerations suggest that there can be conscious experience that does not depend on first-order representations. The next step in the argument aims to show that we have evidence that conscious experience depends on areas of the brain thought to be engaged in higher-order representations. In particular there

is some evidence that the dorsal lateral pre-frontal cortex is the brain area where the higher-order representations can be found. In particular Lau and Passignham (2006) showed that when it is disrupted with TMS bursts subjects report seeing nothing even though they have very good d' scores. These results together suggest that phenomenal consciousness may be higher-order activity in the dorsal lateral pre- frontal cortex and given that we know that it is associated with metacognition it is reasonable to conclude that activity in the dorsal lateral pre-frontal cortex may be the neural substrate of higher-order representations of representations.

Of course the view that we end up with is still something very much like the higher-order thought theory as Rosenthal defends it. It claims that phenomenal consciousness is the having of a certain kind of representation, a higher-order representation of a representation, and further speculates that these representations are in the pre-frontal cortex. All we have done is to accept phenomenal consciousness as a distinct kind of consciousness and to find it a place in the higher-order theory. The HOROR view aims to make clear what the higher-order theory is committed to and what it isn't, but it is a variant of higher-order theory. The main and most notable difference is that the HOROR theory is explicit about the relational requirement applying only to state consciousness. That is to say that state consciousness is explained via a relation of awareness whereas phenomenal consciousness is explained via the awareness itself.

If the forgoing argument is plausible then we have a decent case for thinking that phenomenal consciousness is itself a kind of synchronized neural activity in pre-frontal areas. This is because we have reason to think that phenomenal consciousness is a certain kind of representation and we, in turn, have reason to think that the relevant kinds of representations are nothing but synchronized neural activity in specific frequencies.

The foregoing is a *prima facie* decent case that HOROR theory is empirically viable. Before concluding this section I will note that Ned Block (2007, 2008) and Rafi Malach (2011) have contested the higher-order account on empirical grounds. Block has argued for what he calls phenomenological overflow, which is the claim that phenomenal consciousness outstrips our cognitive access to it. On his view we should think of phenomenal consciousness as neural activity in the relevant first-order sensory areas of the brain rather than as activity in the prefrontal areas. Malach, on the other hand, has suggested that recent empirical work suggests that we can have phenomenally conscious experience when frontal areas are relatively deactivated. I will say a brief word about each before concluding this section.

Malach has used recent results (Goldberg et al. 2006) that suggest that frontal areas responsible for introspection and self-consciousness seem to be relatively inactive while people are absorbed in an external stimulus. Yet we have very good reason from our own cases to expect that one's phenomenal consciousness is very

vivid in these cases. This looks to present a serious challenge to the HOROR approach. However the mistake here is to assimilate these frontal areas with the dorsal lateral prefrontal cortex (Lau & Rosenthal 2011b). The HOROR theory makes no appeal to self-consciousness or introspection and so we would not expect activity in prefrontal areas involved with self-directed introspection to be activated when one is having vivid conscious sensory imagery. HOROR theory is a theory about phenomenal consciousness and so is about our ordinary pre-reflective conscious experiences and not about those rare cases when we turn our attention to our own experiences.

Block uses results from Sperling (1960) and, more recently, work from Victor Lamme's lab (Sligte et al. 2008; Sligte et al. 2009) to argue that phenomenal consciousness is to be identified with activity in first-order visual areas. In the Sperling paradigm subjects are briefly shown an array of letters arranged in a 4×4 grid. Subjects report that they see all of the letters but when asked to name all of the letters they are able to only name about 4. Yet if cued before hand as to which row to attend to they can get most or all of the letters in any given row. Block reasons that if we take them at face value we seem to have evidence that they have more phenomenally conscious experience than they are able to report (after all, they say they see all of the letters but can only report a subset. And the subset could be any row). However it is not at all clear that the arguments for overflow are persuasive at all (Brown 2011). I will not here repeat the arguments but I will just say that both views are compatible with the reports of subjects. If one sees something through a foggy window one may not be able to see all of the details of the object but will no doubt feel as though one has seen the entire object. So too if I am flashed a grid of letters and I see most of them to some extant or other I will be confident that I have seen all of the letters. Given this there is no case for any kind of overflow and so no threat to HOROR theory. The upshot, then, is that the empirical evidence is not strong enough to take overflow seriously, especially when we factor in the independent evidence we have in favor of the non-overflow HOROR view given above (for additional empirical support for the higher-order approach see Lau & Rosenthal 2011a).

I conclude, then, that for all we know HOROR theory is true and phenomenal consciousness just is the right kind of higher-order representation.

7. Identity, reduction, and explanation

As I see things what we have arrived at is a kind of Type-Type identity theory. According to this kind of view types of mental states are identical to types of brain states and types of states of the brain are identical to types of states of mind. So,

depression, on this view, just is a certain range of a dynamic chemical state of the brain, seeing a red bar just is a certain kind of synchronized neural firing. This is because it is plausible that phenomenal consciousness just is a certain kind of higher-order representation and that representation is likely to be synchronized neural activity in dorsal-lateral pre-frontal cortex. However the view we have arrived at is not exactly the same as any of the familiar kinds of type-type identity theory.

Within the identity camp there are two broad traditions that roughly correspond to how one thinks about scientific identities. One view, championed by U.T. Place (2004) and J.J.C. Smart (1991), two of the originators of this theory in philosophical circles, is that mind-brain identities are postulated because they offer the most parsimonious ultimate theory. Thus on this view the postulated identities are brute facts that cannot be explained by anything else. We identify water with H_2O because it allows us to offer the most simple and parsimonious explanation of a wide range of chemical and common sense data.

On the other hand we have a tradition that traces back to David Lewis (1966). On this view the identities are entailed by the theories that make them true. So in the case of water we arrive at the identity of water and H_2O by first identifying water in common sense terms. Water is the stuff that falls from the skies, fills our lakes, etc. We then find out that the stuff that fills our lakes and falls from the sky is H_2O and so we conclude that water is H_2O.

In the foregoing discussion I have been assuming some version of the Lewis strategy. Notice that in each case we started with a common sense way of identifying the kind of consciousness we were interested in and then we found out that that thing turned out to be a particular kind of dynamic brain activity. I favor this kind of view because it allows us to explain why water is H_2O and why consciousness is a certain kind of dynamic brain activity. It also allows us to answer all of the common objections to the identity theory.

Consider first the kind of objections based on conceivability. Descartes famously argued that he could conceive of himself as existing without his body and concluded that he was not his body. The early philosophers who were interested in the brain and came up with the identity theory modeled identity statements as contingent, which means that they just happened to be true but did not have to be true. Just as the fact that we can conceive of Barak Obama losing the election does not show that he is not currently the President of the United States. So too, they reasoned, just because we can conceive of the mind without the body doesn't show that the mind isn't the body in actuality.

The well-known problem with this move is that it seems plausible that scientific identity statements are necessary. Consider one famous philosophical thought experiment known as Twin Earth. Twin Earth is a place where there is

a clear odorless substance that the inhabitants even call 'water' which turns out not to be H_2O. Its microstructure is something much more complicated (and which philosophers have chosen 'XYZ' to indicate). So, on Twin Earth water is XYZ, not H_2O. This looks like a case where we have something that we might describe as "fool's water" (Kripke 1980). Fool's water is stuff that looks like water but is not. In other words there is a strong tendency to think that there is no water on Twin Earth. Water is H_2O, and there is no H_2O on Twin Earth. Others think that there is water on Twin Earth, it just so happens that water – for them- is XYZ.

Given this simple way of thinking about things (for more see Chalmers 2008) we can see that conceiving of a world with a ghostly mind is no problem. Just as the XYZ world did not impugn the identity of water and H_2O in actuality so too the ghost worlds do not impugn the physical credentials of consciousness around here. The ghost world is just another way that we might get consciousness in the world but its conceivability shouldn't bother us here.

What about zombies? Chalmers (2009) appeals the conceivability of physical duplicates of me that lack phenomenal consciousness. According to the present view the zombie world is akin to a world that is physically identical to our world in that it has H_2O but is stipulated to lack water. This is not even conceivable. Given what we know now we can see that we can in fact deduce water facts from H_2O facts and that shows us that there are no possible worlds like the one described. Just given the H_2O facts alone necessitates water facts. If the mind-brain identity theory is true then the same is the case for mind-brain identities. The zombie world is then inconceivable. What are we to say to the charge that the zombie world seems conceivable? Is this an objection? No. The problem is that it is equally conceivable that consciousness be a physical property. I have previously (Brown 2010) called these creatures 'shombies'. Shombies are creatures that are completely and exhaustively physical but that are conscious in exactly the same way that I am. In fact what we have seen in this chapter is an argument to the effect that we can conceive of shombies.

One other kind of objection comes from thinking about the possibility of minds that are not composed of neurons. Let us discuss the science fiction example of Commander Data from the Star Trek series. Data is portrayed as having a 'positronic' brain, which is supposed to be something like a functional isomorph of the human brain. If it is empirically possible to build something like Commander Data then the type-type identity theory is not true. In that case a kind of functionalism would be true about the mind. This is a possibility but it is an empirical question. If it turns out that there are biological properties of neurons that matter and cannot be reproduced artificially then Commander Data cases will

turn out to be like ghost cases. They will be worlds where consciousness is not a brain state but is rather a positronic state.

As of right now we have no reason to believe in the multiple realizability of consciousness. Instead we have good reason to believe that the mechanisms discussed in this paper hold for all brained species on Earth. We may have intuitions about what could have been the case about consciousness (could it have been positronically based rather than brain based?) but we don't have any empirical reason to think so. So, at least as of now, we cannot take intuitions about machine consciousness as a defeater for the type of reductive view I am arguing for. For all those intuitions machine consciousness just may not be actually possible.

8. Conclusion

The brain is a dynamical system that is constantly evolving in time. There are two faces to this evolution in time. One face is chemical and is the story of how the ratio of aminergic and cholinergic neuromodulators evolves in time. The other face is electrical and is the story of how transient assemblies of neurons are formed via synchronous phase-locked firing and transmitted to disparate areas via long-range synchronization in different frequencies. We discovered these dynamic activities by looking for what in the brain performs various mental tasks. We start with sleep characterized in terms of behavior and then discover the nature of it in the state of the brain. So too we ask how does the brain represent? And we find out that it does so by instantiating a certain pattern of neural activity. It is then natural to deduce that the nature of sensing or sleeping just is the activity in the brain.

This is the way normal scientific identities are established and defended. But it can only be done when one has a theory of the phenomenon that is not couched in neuronal terms. We identify some mental activity as a representation of orientation, say, because we were looking for a representation of orientation. That is, we started with some idea about how we pick those kinds of things out. It is, we think, whatever state we find which reliably tracks this orientation as opposed to some other. It is only because we understand those things in non neural terms that we are able to look at the brain to see what it is that does that. In this way we can see that we cannot have a truly neural theory of consciousness. We must always have some higher-level theory of the thing in question. Neuroscience then is in a position to tell us what in the brain those things are.

Once we recognize that this is the way that scientific identities are discovered we can see that there is no threat to the identity theory from any of the major

objections to it. Neither is there any kind of tension between higher-order theories and biological theories of consciousness.[1]

References

Armstrong, D.M. (1968). *A Materialist Theory of Mind*, London: Routledge.
Baars, B. (1988). *A cognitive Theory of Consciousness*, NY: Cambridge University Press.
Block, N. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Sciences, 18*, 227–47.
Block, N. (1996). Mental paint and mental latex. In E. Villanueva (Ed.), *Perception.* Atascadero, CA: Ridgeview.
Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*, 481–548.
Block, N. (2008). Consciousness and cognitive access. *Proceedings of the Aristotelian Society, Vol. cviii, Part* 3.
Block, N. (2011) The higher-order approach to consciousness is defunct *Analysis*, *17*(3), 419–431.
Brown, R. (2006). What is a Brain State?. *Philosophical Psychology, 19*(6), 729–742.
Brown, R. (2010). Deprioritizing the a priori arguments against physicalism.*Journal of Consciousness Studies*, *17*(3–4), 47–69.
Brown, R. (2011). The Myth of phenomenological overflow *Consciousness and Cognition.* doi:10.1016/j.concog.2011.06.005.
Buzsaki, G. (1989). Two-stage model of memory trace formation: A role for noisy brain states. *Neuroscience, 31*(3), 551–570.
Buzsaki, G. (1996). The hippocamo-neocortical dialogue. *Cerebral Cortex, 6*(2).
Byrne, A. (2001). Intentionalism Defended. In *Philosophical Review,* I: 199–240.
Chalmers, D. (2008). Two-Dimensional Semantics. In E. Lepore & B. Smith (Eds.), *Oxford Handbook of the Philosophy of Language*. Oxford: Oxford University Press.
Chalmers, D.J. (2009). The Two Dimensional Argument Against Materialism. In B.P. McLaughlin & A. Beckermann (Eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press.

1. Some of the ideas in this paper were presented at Columbia University's psychology department as part of their "Cognitive Lunch" Speaker Series under the title "Consciousness and the Tribunal of Experience" March 22 2010. Other ideas in the paper were presented at the Southern Society for Philosophy and Psychology in 2011 as "The Higher-Order Approach to Consciousness: The HOT Ticket or in HOT Water?" and in 2012 as "Phenomenal Consciousness Ain't in the (Back of the) Head". I am grateful to participants for helpful discussion. In particular I have benefitted from discussions with David Rosenthal, Josh Weisberg, Jake Berger, Alex Kiefer, Ned Block, and Pete Mandik. I would also like to especially thank Hakwan Lau for very valuable discussion on many of the ideas in this paper and the data generated by his lab.

Colgin, L.L., Denninger T., Fyhn, M., Hafting,T., Bonnevie, T., Jensen, O., Moser M. & Moser, E.I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature, 462*(7271), 353–357

Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79*, 1–37

Dretske, F. (1993). Conscious experience. *Mind, 102*, 263–283.

Goldberg, I.I., et al. (2006). When the brain loses its self: Prefrontal inactivation during sensorimotor processing. *Neuron, 50*, 329–339.

Hobson, J.A. (2009). REM sleep and dreaming: Towards a theory of protoconsciousness *Nature Reviews Neuroscience, 10*, 803–813.

Kahn, D., Pace-Schott, E., et al. (1997). Conscious in waking and dreaming: The role of neuronal oscillation and neuromodulation in determing similarities and differences. *Neuroscience & Biobehavioral Reviews*, *78*(1), 13–38.

Kripke, S.A. (1980). *Naming and Necessity*. Cambridge: Harvard Univ Pr.

Lau, H. & Brown, R. (Forthcoming). "The Emperors new Phenomenology? The Empirical Case for Conscious Experience without First-Order Representations" in a Festschrift for Ned Block edited by Adam Pautz and Daniel Stoljar (Eds.), MIT Press.

Lau, H. & Passingham, R. (2006). Relative blindsight in normal observers and the neural correlate of vsual consciousness. *Proceedings of the National Academy of Sciences 103*(49), 18763–18768.

Lau, H. & Rosenthal, D. (2011a). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Science 15*(8), 365–373.

Lau, H. & Rosenthal, D. (2011b). The higher-order view does not require consciously self-directed introspection: Response to Malach. *Trends in Cognitive Sciences, 15*(11), 508–509.

Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy, 63*, 17–25.

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

Malach, R. (2011). Conscious perception and the frontal lobes: comment on Lau and Rosenthal. *Trends Cogn. Sci. 15*(11), 507.

Place, U.T. (2004). Is Consciousness a Brain Process?. In G. Graham & E. Valentine (Eds.), *Identifying the Mind:* Selected Papers of U. T. Place. Oxford University Press.

Prinz, J.J. (2005). A Neurofunctional Theory of Consciousness. In A. Brook & K. Akins (Eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movemen*t. Cambridge: Cambridge University Press.

Rahnev, D., Maniscalco, B., Huang, E. & Lau, H.C. (2009). Inattention Boosts Subjective Visibility: Implications for Inattentional and Change Blindness. *Journal of Vision, 9*(8), 157.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F., Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14*, 1513–1515.

Rosenthal, D. (2005). *Consciousness and Mind*. New York: Oxford University Press.

Singer, W. (1996). Neuronal synchronization: A solution to the binding problem? In Llinas R, Churchland P. *The Mind-Brain Continuum.* The MIT Press.

Sligte, I.G., Scholte, H.S. & Lamme, V.A.F. (2008). Are there multiple visual short-term memory stores? *Plos One 3*(2), 1–9.

Sligte, I.G., Scholte, H.S. & Lamme, V.A.F. (2009). V4 Activity predicts the strength of visual short-term memory representations. *Journal of Neuroscience 29*(23), 7432–438.

Singer, W. (2000). Consciousness from a neurobiological perspective. In: Metzinger T, *Neural Correlates of Consciousness*. The MIT Press.

Smart, J.J.C. (1991). Sensations and brain processes. In D.M. Rosenthal, *The Nature of Mind*. Oxford University Press.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, 1–29.
Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge: MA: MIT Press.
Weisberg, J. (2011). Abusing the notion of what-it's-like-ness: A response to Block,' *Analysis*, 71, 438–443.