Extended Consciousness: an Interim Report

Michael Wheeler

Abstract

Advocates of extended cognition hold that the physical machinery of mind sometimes extends beyond the skull and skin. In the first part of this paper, I explain why, and more specifically the precise sense in which, consciousness presents such theorists with an extra hurdle to be cleared. The key challenge is posed by phenomenal consciousness, the what-it’s-like-ness of experience. I consider two arguments for the claim that the physical machinery of phenomenal consciousness sometimes extends beyond the skull and skin. The first – the argument from sensory substitution – suggests that acceptance of extended phenomenal consciousness should follow from a careful analysis of the phenomenon in which technological augmentation enables one sensory modality, for instance touch, to support the kind of environmental access and interaction ordinarily supported by a different sensory modality, for instance vision. The second argument – the argument from the relational character of experience – suggests that acceptance of extended phenomenal consciousness should follow from a particular conception of conscious experience that is mandated by sensorimotor contingency theory. I conclude that neither argument is decisive.

Keywords

consciousness, extended cognition, extended mind, relational character of experience, sensorimotor contingency, sensory substitution.

1. Sunset 53rd Street

Send for director Clark, because Otto, the first star of the extended cognition era, is ready for his close-up. Otto has one last message for (what he thinks of as) his loyal army of adoring fans. And here it comes: ‘it was never about consciousness’. Inga is not impressed. She knew that all along. Inga is, of course, Otto’s devoted companion. For years, she has been keeping all those nasty internalist criticisms of their celebrated (or notorious, depending on your point of view) thought experiment away from Otto. He doesn’t need to know how sceptical some people (especially Fred, Ken and Rob) have
been. Indeed, as far as Otto is concerned, everybody believes in extended cognition. It’s a little sad that Otto doesn’t know the truth, but, in Inga’s view, things are better that way.

Musing on Otto’s words, Inga is perplexed. Why did Otto bother to say that about consciousness? Is there anybody out there really so injudicious as to think that the physical machinery of conscious experience (the processing machinery that enables us to be conscious of things in the world) extends beyond the brain into the non-neural body and the external environment? After all, she exclaims to herself, with a good deal of inner impatience, everybody has seen the neural imaging pictures. If there’s one thing we do know, it’s that the machinery of consciousness is housed entirely in the brain. It seems that Otto has forgotten his own plot. Maybe it’s finally time for him to put down the notebook. Prompted by this thought, Inga reminisces about the good old days. (Of course, it was Clark and Chalmers then, but David went on to other things.) And she wonders: what exactly is the legacy of her legendary partnership with Otto?

Inga’s question is a good one, so let’s follow her down memory lane. Here’s how it was, back in 1998, when Clark and Chalmers introduced Inga and Otto to the world. Cue cinematic dissolve.¹

¹ Inga and Otto famously appeared in the canonical presentation and defence of extended cognition by Clark and Chalmers (1998). Clark’s own more recent treatment of the view may be found in (Clark 2008). For a field-defining collection that places the original Clark and Chalmers paper alongside a range of developments, criticisms and defences of the notion of extended cognition, see (Menary 2010). Fred, Ken and Rob are, of course, the most discussed internalist critics of extended cognition, Fred Adams, Ken Aizawa (e.g. Adams and Aizawa 2008) and Robert Rupert (e.g. 2009). In this paper, when I use the terms ‘internalist’ and ‘externalist’ (and ‘internalism’ and ‘externalism’), I am referring to opposing views about the spatial location of the physical machinery of (alternatively, of the material vehicles that realize) cognition or consciousness. I am not referring to views about what determines or individuates mental content (so-called content internalism and content externalism). So, in the present context, the internalist holds that the physical machinery of cognition/consciousness (disambiguation in context) is wholly internal (paradigmatically, brain-bound), while the externalist holds that some of the physical machinery of cognition/consciousness sometimes lies beyond the skin. Hitting the same note, Clark (2009, 968) brings the hypothesis of extended consciousness into view like this: ‘the question is not whether the states of affairs that the conscious thought concerns are extra-neural: they nearly always are. Nor is it a question about what individuates the thought as, for example, a thought about a tooth... It is a question about what might be dubbed the machinery of mind’.
2. From Extended Cognition to Extended Consciousness

First, the backstory: according to the hypothesis of *extended cognition* (henceforth ExC), the physical machinery of mind sometimes extends beyond the skull and skin. More precisely, ExC is the view that there are actual (in this world) cases of intelligent thought and action, in which the material vehicles that realize the thinking and thoughts concerned are spatially distributed over brain, body and world, in such a way that certain external (beyond-the-skull-and-skin) factors are rightly accorded cognitive status. The term ‘cognitive status’ is here really just a place-holder for ‘whatever status it is that we standardly grant the brain, in cognitive science and naturalistic philosophy of mind, when talking about the causal wellsprings of intelligent thought and action’.

On another day, more would need to be said about exactly what that means.

Now for the plot: Inga is a psychologically normal individual who has committed to her purely organic (neural) memory the address of the New York Museum of Modern Art (MOMA). If she forms the desire to go to MOMA, she accesses her organic memory to retrieve the information that the building is on 53rd Street. Otto, on the other hand, suffers from a mild form of Alzheimer’s which means that he cannot internally retain certain kinds of factual information. He compensates for this shortcoming by recording salient facts in a notebook that he carries with him constantly. If Otto forms the desire to go to MOMA, he automatically and unhesitatingly pulls out the notebook and, without a hint of any critical scrutiny of the information stored within, retrieves the relevant fact, viz. that the museum is on 53rd Street. At this point in the drama, the so-called parity principle (Clark and Chalmers 1998) makes its entrance. The parity principle asks us to start by considering a distributed system like Otto and his notebook, that is, a system (a) that generates some psychologically interesting outcome and (b) whose operation involves not only neural clankings and whirrings, but also an important functional contribution from certain externally located (with respect to the skin) physical elements. It then encourages us to imagine a hypothetical scenario in which the functional contribution identified in (b), to an equivalent outcome to the one identified in (a), is made not by any external elements, but by certain internally located factors.

Having taken this imaginative step, if we then judge that the internal realizing elements in the latter hypothetical case count as bona fide parts of a genuinely cognitive system, we ought to conclude that the very same status – that is, cognitive status (see above) –

---

2 I am going to use the terms ‘mind’ and ‘cognition’ interchangeably. Although this upsets some philosophers, it strikes me as standard practice in cognitive science. From this perspective, ‘extended cognition’ and ‘the extended mind’ are alternative names for the same view.
should be granted to the external realizing elements in the environment-involving case with which we began. To do otherwise would be to succumb to neural chauvinism, which is to beg the question against the externalist. The idea is that, if we follow the steps just specified, in order to implement the parity principle, we should conclude that Otto’s memory is extended into the environment. Moreover, argue Clark and Chalmers, just as, prior to recalling the information in question, Inga has the non-occurrent dispositional belief that MOMA is on 53rd Street, so too does Otto, although while Inga’s belief is realized in her head, Otto’s is realized in the extended, notebook-including system.3

For present purposes, I am not in the business of determining whether Clark and Chalmers succeed in establishing that Otto has an extended memory and an extended dispositional belief. Let’s just agree that he does. I am interested in the following line of thought. Dispositional beliefs are paradigm cases of psychological states that are not conscious. A few minutes ago, I had the belief that my daughter is 9 years old, even though that content wasn’t part of my conscious mental life, but was merely poised to become so when certain triggering conditions were present, which they were when I wrote this sentence. So, a few minutes ago, I had the belief in unconscious dispositional form. A related story could be told about the unconscious status of the information that is stored away in memory prior to conscious recall. So, if the parity-centred argument rehearsed above establishes anything at all, it establishes only that certain kinds of

3 Interestingly, given this understanding of the parity principle (which I am confident is the only sensible reading of it – see Wheeler 2011a), the conclusions regarding Otto’s extended memory and his extended dispositional belief may be drawn without any reference to Inga. This is because, when we deploy the parity principle, the wholly internal system that plays the regulating role in driving the case for extended cognition is an imagined system that need not coincide with any existing example. As it happens, Clark and Chalmers themselves make the claim that there is a functional equivalence between the behaviour-governing causal role played by Otto’s notebook, and the behaviour-governing causal role played by the part of Inga’s brain that stores the same item of information. With the benefit of hindsight, this claim was, in my view, unfortunate, because it encouraged a host of internalist critics to hunt for functional differences between Otto and Inga (that is, between cases of putatively extended cognition and cases of extant inner cognition), in an attempt to demonstrate a lack of parity between the two and thereby block the externalist argument. But once the parity principle is understood independently of the Inga and Otto example, it seems clear enough that, in relation to the plotline of what that principle may or may not establish, this was something of a garden-path diversion. In truth, Otto always knew that Inga had no more than a bit part in the drama. He just didn’t like to mention it.
unconscious psychological phenomena may be extended. This makes sense of Clark’s (2009, 967) claim that ‘[the extended mind] was concerned only with the vehicles of non-conscious mental states such as states of dispositional believing’. (That’s why Inga was bemused by Otto’s pronouncement. Didn’t he know that?)

Let’s now turn to the situation that obtains after Otto has looked in his notebook and is wending his way to 53rd street. It is possible, I think, that, in this situation, Otto may have an occurrent, conscious belief that remains a case of extended cognition. Perhaps the information that MOMA is on 53rd street never manages to lodge itself in Otto’s brain in any remotely stable way, while the notebook stays open and the externally stored information therein remains accessible, not only as a kind of anchor for his behaviour, but also as the material vehicle of the content of an intermittent conscious belief that MOMA is on 53rd Street. In other words, when instantiated – perhaps when a passer-by asks Otto where he is going and why he is going there – what results is a conscious belief that is widely realized in Otto’s brain, body and notebook. There are undoubtedly some tricky issues here, but note that the internalist cannot simply insist that any time Otto has the relevant conscious belief, the content must be represented in his brain, for fear of (once again) begging the question. That said, however, it is (at least) equally possible that what happens when Otto looks in his notebook and forms the conscious belief that MOMA is on 53rd Street is that the salient information is transferred to, and remains stably stored and accessible in, his brain, until he arrives at his destination. Of course, because of his Alzheimer’s condition, that information will be lost from his brain later, but nevertheless, during the relevant episodes of conscious belief, what we confront may be a conventional case of wholly inner consciousness.

The lesson from all this uncertainty is that, whatever else the case of Inga and Otto may show, nothing about it decides between the two scenarios just sketched, one of which is externalist in form, the other of which is internalist. So, as far is this particular, but undeniably canonical, argument for ExC goes, the physical machinery of conscious experience may remain wholly internal. And although, as mentioned already, the agreed rules of engagement mean that the internalist is not allowed simply to insist that she is right, she might well suggest that existing practice in cognitive science, when coupled with a perfectly reasonable conservatism about scientific theory-change (i.e.,

---

4 For what it’s worth, this suggestion doesn’t strike me as even remotely odd. Perhaps I’m weird, but sometimes, when I’m using a mobile navigation app, I could tell you the name of the place for which I’m heading – e.g. MOMA – but I could not tell you its address, unless or until I look at the app. Arguably, when I look at the app, I have the conscious belief that, say, MOMA is on 53rd Street, while at the other times I don’t.
replace an empirically fruitful existing theory only if a challenger has demonstrable explanatory advantages), gives her view of consciousness a kind of default status.

At this juncture, the fan of extended consciousness might be moved to complain that we are placing too much emphasis on the Inga and Otto case. Held spellbound by their star billing, we are blind to other arguments for ExC that might plausibly generalize to consciousness. There is undoubtedly some force to this complaint. That said, here is an instructive general consideration. It has become a common move in the literature to link ExC in some way to what might be judged the house philosophy of mind in cognitive science, namely functionalism. Hence the fact that ExC is sometimes cashed out as extended functionalism (see Clark 2008a, b; followed by Wheeler 2010a, b, 2011a, 2014).

The functionalist about cognition holds that what makes a state or process cognitive is not its material composition, but the functional role it plays in generating psychological phenomena, by intervening causally between systemic inputs, systemic outputs and other functionally identified, intrasystemic states and processes. Now, the fact is that the possibility of extended cognition is a straightforward consequence of functionalism. This might seem surprising because, historically, the assumption (in philosophy and cognitive science) has tended to be that the functionally identified states and processes that matter for mind, at least in the case of human beings and other minded life on earth, will be realized by the nervous system. In truth, however, there isn’t anything in the letter of functionalism that mandates this internalism (Wheeler 2010a, b). After all, what functionalism requires is that we specify the causal relations that exist between some target element and a certain set of systemic inputs, systemic outputs and other functionally identified, intrasystemic elements. Nothing about this demands internalism, since the boundaries of the functionally identified system of interest – i.e., the cognitive system – may in principle fall beyond boundaries of the organic sensory-motor interface.

What this tells us is that functionalism makes extended cognition a conceptual possibility. Indeed, functionalism plausibly provides a theoretical backdrop for the operation of the parity principle. For although it is possible to conduct a debate that revolves around the functional contributions of certain elements without assuming functionalism as such (Chalmers 2008), the parity principle nevertheless does assume that, in principle at least, the very same type-identified cognitive state or process is realizable in either a purely organic medium (the imagined wholly internal system) or one that involves an integrated combination of organic and non-organic structures (the Otto-and-notebook system). In other words, the parity principle assumes that cognitive phenomena are multiply realizable. And even if a fan of ExC chooses to turn her back on the parity principle, she will presumably still want to hold onto the surely attractive thought that the very same type-identified cognitive process may, on some occasions,
take place wholly inside the head, while on others it may take place in an extended brain-body-environment system. Take mathematical cognition: there are some mathematical calculations for which it seems plausible to say that although, on Monday, I may carry them out using pen and paper, on Thursday I may call only on my organic resources. Now, if we are to describe these alternative problem-solving routines as two realizations of the very same type-identified process of mathematical cognition (which, assuming certain conditions are met, is what the advocate of ExC will be inclined to do), then it must be possible for the very same psychological reasoning process to enjoy (at least) two different material instantiations. In other words, once again, the target cognitive phenomenon must be multiply realizable. But if the advocate of ExC requires multiple realizability, it will be tempting for her to sign up to functionalism. For while functionalism may not be necessary for multiple realizability, it is standardly thought to be sufficient, since a function is something that, as we have seen, enjoys a particular kind of independence from its implementing material substrate.

So far, so good, but when we turn to consciousness, this orgy of reciprocal back-slapping between ExC and functionalism grinds to a halt. This is because, famously, the phenomenon of consciousness provides functionalism with some daunting philosophical challenges, in the form of thought experiments such as the single system comprising the entire Chinese nation, organized so as to satisfy the functional definition of a mind (Block 1980) and the functionally-identical-to-one-of-us zombie (Chalmers 1996). In cases such as these, the intended message is (roughly) that since we enjoy conscious experience, yet certain systems functionally identical to us plausibly don’t, no purely functional characterization can explain consciousness. In truth, rather more precision is needed here. If one focusses on so-called access consciousness – where a state is access conscious if the information it carries is flexibly available for use in reasoning and the rational guidance of speech and action – then arguably the functionalist about the machinery of consciousness is at least in with a fighting chance. Access consciousness, so defined, seems to be amenable, in principle anyway, to a functionalist (e.g. information processing) treatment. But, if that’s so, then presumably extended functionalism about the machinery of consciousness is also still in the game, since that’s just multiple realizability playing itself out. In fact, the same guarded open-mindedness about the prospects for extended consciousness is encouraged, I think, if one interprets the conscious belief enjoyed by Otto as he journeys to 53rd Street as a case of access consciousness. This is unsurprising, since, as we have seen, functionalism may provide support for the parity argument in which Otto plays his part. So the sound of back-slapping may still be heard. However, when we narrow our sights to phenomenal consciousness – that fabled what-it’s-like-ness of experience, such as what it’s like for me to see blue, taste wasabi peas, or, as I’m doing right now, listen to The
Strypes – we confront the full might of the Chinese nation and their zombie friends. It is here that functionalism is often thought to struggle, and to struggle badly.\(^5\)

If the extended functionalist is an optimist by nature, she might be tempted to spin this point positively. After all, if there is an explanatory gap between functionalist specification and phenomenal consciousness, that gap obtains whether the realizing machinery is wholly neural or an extended matrix of elements in the brain, the non-neural body, and the beyond-the-skin environment. It’s the functional basis of the explanation that causes the difficulty, not where the realizing machinery happens to be spatially located. So although functionalism may indeed struggle in the face of phenomenal consciousness, extending one’s functionalism doesn’t make things any worse than they already were.

Unfortunately, at first sight, this sort of positive spin looks to be nigh on baseless. After all, being no worse off than someone who has no idea at all is hardly a victory. However, perhaps this is a case where first impressions are misleading. To see why, we can draw on an analysis due to Hurley (2010), in which she exposes a tension between two widely held intuitions regarding the machinery of phenomenal consciousness.\(^6\) The first of these intuitions is the internalist intuition, which is that if phenomenal qualities can be realized physically or functionally at all, it could only be in terms of internal factors. Hence all the fuss in cognitive neuroscience and naturalistic philosophy of mind about the neural correlates of consciousness. The second intuition is what Hurley calls the autonomy metaintuition. This is Hurley’s recoding of the alleged explanatory gap highlighted above. Thus, in general terms, ‘[the] autonomy metaintuition for phenomenal qualities is an expression of the intuition that there is an intractable explanatory gap between physical or functional properties and phenomenal qualities’ (Hurley 2010, 104). Applied specifically to the machinery of phenomenal consciousness, the autonomy metaintuition is an expression of bewilderment about precisely how any suite of purely physical or functional states and processes could realize phenomenal consciousness. The tension between these two intuitions is clear enough. To echo some of Hurley’s own rhetorical words, if we have no understanding of how phenomenal qualities could be realized in neural elements, why is the conditional intuition so strong?

\(^5\) The distinction between access and phenomenal consciousness is due to Block (1995).

\(^6\) As I shall explicate it here, the tension in question concerns intuitions about the machinery of phenomenal consciousness. Hurley herself initially explicates the tension as concerning intuitions about what determines phenomenal quality, but she is clear that the tension surfaces at the level of machinery too (Hurley 2010, 148 n.10; for further discussion, see Wheeler 2015). The distinction between the ‘what-determines’ and the ‘machinery’ levels of analysis will be relevant later in this paper.
that if phenomenal qualities can be realized physically or functionally at all, it could only be in terms of internal (neural) factors?

Notice that, in the wake of this tension, it’s the autonomy metaintuition that cuts the ground from under the feet of the internalist intuition, rather than the other way round. This becomes clear if we try to express the tension differently, by formulating it as follows: if it is true that, if phenomenal qualities can be realized physically or functionally at all, it could only be in terms of internal (neural) factors, why is the intuition so strong that we have no understanding of how phenomenal qualities could be realized in neural elements? This thought is suspect, because the claim that phenomenal qualities must be realized in neural elements sheds no light on how that might be true. Put another way, the bald commitment to internalism does not undermine the explanatory gap. On the other hand, the explanatory gap does undermine the bald commitment to internalism. That, I think, is why Hurley expresses the tension the way round that she does. Of course, this analysis doesn’t favour externalism about the machinery of phenomenal consciousness over internalism – because the autonomy metaintuition applies to purely physical or functional properties, wherever they are located – but it does level the playing field. For it suggests that, when it comes to the machinery of phenomenal consciousness, the prior plausibility of internalism and externalism (prior, that is, to further philosophical argument and empirical investigation) turns out to be equal. And what this means is that the previously-lodged claim by the internalist, that her view has default status in the debate (see above), is rendered insecure. In truth, neither internalism nor externalism about the machinery of phenomenal consciousness enjoys that privileged status. Game on.

For the remainder of this paper, I will be concerned only with phenomenal consciousness. According to the hypothesis of extended phenomenal consciousness (henceforth ExPC), there are actual (in this world) cases of phenomenal consciousness in which the material vehicles that realize the phenomenally conscious states and processes concerned are spatially distributed over brain, body and world, in such a way that certain external (beyond-the-skull-and-skin) factors are rightly accorded whatever status it is that we standardly grant the brain, in cognitive science and naturalistic philosophy of mind, when talking about the causal wellsprings of phenomenal consciousness. If ExPC is true, although the machinery that realizes conscious phenomenal experience includes, and maybe necessarily includes, neural elements, it is not restricted to such elements. There will be cases in which that machinery additionally includes not only non-neural bodily elements, but also elements located beyond the skull and skin. To keep our thinking straight, it is worth emphasizing the fact that ExPC, as we are conceiving it, does not entail that the worldly object or state of affairs that one is phenomenally conscious of is part of the relevant machinery. The objects of
phenomenal consciousness may remain external to the machinery of such consciousness, even if the machinery is now partly external with respect to the brain and body.

So, should we accept ExPC? In the space available here, I cannot hope to review all the treatments of this question that appear in the literature. So I shall concentrate on assessing just two arguments that, if sound, would produce affirmative answers, and I shall comment on a connection between them. These arguments are what I shall call the argument from sensory substitution and the argument from the relational character of experience.

3. The Argument from Sensory Substitution

Sensory substitution, as it will concern us here, occurs when technological augmentation enables one sensory modality to support the kind of environmental access and interaction ordinarily supported by a different sensory modality. The seminal work in this area is Paul Bach y Rita’s (1972; Bach y Rita and Kercel 2002) research on tactile-vision sensory substitution (henceforth TVSS). In this work, blind subjects were equipped with a head- or shoulder-mounted camera that conveyed information, from video images, via the activation of an array of vibrators located on the subject’s back, abdomen or thigh. After a short period of adaptation, those TVSS subjects who actively controlled the information received, either by manipulating their bodies or by manipulating the camera, were able to make reliable judgments about things such as the number, relative size and position of distal objects in three-dimensional space, and to perform actions such as reaching out and picking up objects. TVSS subjects have also been successful at making perceptual judgments involving effects such as looming and object occlusion, and (this time with image-sourced information transmitted via vibrators on the tongue) have reported experiencing illusory movement effects such as the waterfall illusion (Bach y Rita and Kercel 2002). TVSS is not the only form of sensory substitution. For example, in auditory-vision substitution (e.g. Auvray et al. 2005), video images from a camera are converted into

7 Most notably, I shall not discuss the prominent pro-ExPC arguments offered by Hurley and Noë (2003), the criticisms of those arguments lodged by Clark, or Clark’s own positive argument against extended consciousness (Clark 2009). I comment on the last of these in (Wheeler 2015, 131 n.2).

8 This section develops a variation on an argument that I first offered in (Wheeler 2015). In the present treatment, various aspects of that argument have been reorganized and enhanced. In that prior paper, I additionally consider (and reject) a different modulation of the argument from sensory substitution due to Kiverstein and Farina (2013).
sounds (e.g. the vertical positions of pixels may be correlated with different audio frequencies) and conveyed to the subject via headphones. Again after short periods of adaptation, subjects equipped with such devices have been able to localize, and to recognize the shapes of, distal objects in three-dimensional space.

Noë (2009) argues that a careful analysis of sensory substitution indicates that we should accept ExPC. As I understand it, the reasoning goes like this: (i) sensory substitution involves a transformation in perceptual experience; (ii) that transformation is achieved without there being any change in the underpinning neural activity; (iii) if we assume internalism about the machinery of phenomenal consciousness, the transformation in question is, from the perspective of naturalism, mysterious; (iv) if we adopt ExPC (i.e., externalism about the machinery of phenomenal consciousness), the transformation is naturalistically explicable; (v) given some target domain of phenomena, we should defeasibly accept the existence of those states and processes that are required by the most powerful explanations that we have available in that domain;9 (vi) we should accept ExPC. As we shall see, steps (i) and (ii) are explicit in Noë’s treatment of the issue. Steps (iii)-(v) admittedly involve some creative filling-in on my part, but are implicit in, or at least consistent with, Noë’s approach and would be sufficient to complete the argument. The conclusion, (vi), is expressed by Noë in his claim that ‘the world itself can be described as belonging to the very machinery of our own consciousness’ (Noë 2009, 65).

Let’s start with (i). Some blind users of sensory substitution systems report experiences that might be categorized as visual qualia, such as experiences of phosphenes (the seeing of light without light actually entering the eye) (Ortiz et al. 2011). It is because of such reports that Noë (2009, 57) plausibly understands TVSS, thought of as a paradigmatic instance of sensory substitution, to be a ‘full-fledged, bona fide example of... [a] transformation in perceptual consciousness’. What Noë means by a ‘transformation in perceptual consciousness’ is that there has been a transition in experience from one sensory modality to another. The most straightforward (but also most controversial) claim in the vicinity here would be that post-adaptation TVSS subjects genuinely see (enjoy authentic visual phenomenal consciousness), in spite of the fact that the relevant channel of proximal stimulation is tactile (vibrations on the skin caused by the TVSS technology). In truth, however, Noë doesn’t need such a radical claim, in order to construct his argument for ExPC. All he needs is that TVSS engenders a transformation in perceptual consciousness such that, even though the proximal stimuli remain tactile in character, the post-adaptation conscious experience in question is not correctly categorized as one of touch, even if it isn’t vision (Noë 2009, 65).

---

9 My formulation of this principle is indebted to (Rupert 2013).
Given the perceptual sensitivity of TVSS subjects to distal objects, and given that touch standardly involves perceptual sensitivity through surface-to-surface contact (pressure reception), that much certainly seems to be defensible.

Step (ii) in Noë’s argument is the claim that the kind of transformation in perceptual consciousness just described is achieved without there being any change in the neural activity underpinning it. As Noë (2009, 54) puts it, the ‘character of conscious experience can vary even though the neural activity underpinning it does not change’. Taking it that the kinds of variation in conscious experience that matter here are those that involve transitions in perceptual consciousness from one sensory modality to another, what does Noë mean by a ‘change in underpinning neural activity’? An answer to this question is suggested by the following observation. According to Noë (2009, p.56), TVSS cases show that ‘[neural] rewiring isn’t necessary for changes [transformations] in conscious experience’. In other words, the marker for a change in underpinning neural activity is neural rewiring. The inspiration here comes from experiments on neural plasticity due to Sur and colleagues (Sur et al.1999, described by Noë 2009, 53-4). In these experiments, the neurophysiology of new-born ferrets was rewired so that their eyes ended up being connected to the parts of their brains ordinarily used for hearing. The result was that those regions of the ferret brain that standardly process auditory information were recruited for vision. This constitutes a transformation in sensory experience of the right kind (from one modality to another), and it is obvious that an important factor in effecting that transformation was neural rewiring. In fact, if there were any neural rewiring in the TVSS case (which Noë claims there isn’t), it would presumably not be quite the same as that in evidence in the ferret model. The TVSS phenomenon would require not that the path from a proximal input be rewired so as to recruit a different part of the brain to the proximal input’s standard sensory modality, but rather that the path from a proximal input be rewired so as to activate a different part of the brain to usual, so as to produce experiences in that part of the brain’s standard sensory modality. However, one gets the idea: the relevant sense of a ‘change in underpinning neural activity’ is a physical modification to the wiring of the brain.10

10 To head off a possible exegetical challenge, let me register the fact that Noë’s statement that the ‘character of conscious experience can vary even though the neural activity underpinning it does not change’ actually appears not during his discussion of TVSS, but during his discussion of Sur’s ferrets. Indeed, Noë (2009, 54) describes it as ‘the basic lesson of Sur’s studies’. To my mind this is an odd remark, since, as we have seen, in the ferret case, there is a change in neural activity, in the sense of neural rewiring. This might make one suspicious that whatever Noë means by a ‘change in neural activity’, at least in connection with Sur’s ferrets, it can’t be neural rewiring. One
Why precisely is neural rewiring supposed to be ruled out as an explanation of the transformation in perceptual consciousness present in TVSS? For one thing, TVSS subjects adapt swiftly to their augmenting devices, with the transition in experiential sensory modality happening far too quickly for the explanation to turn on the kind of substantial neural rewiring that we see in the ferret case. For another, TVSS subjects are typically adult human beings whose brains simply do not have the plasticity exhibited by the brains of new-born ferrets. With thoughts such as these in the background, Noë describes the post-adaptation TVSS scenario as follows:

Stimulation of the skin gives rise to neural activity in touch areas of the brain (the so-called somatosensory cortex). But for a person who has adapted to the sensory substitution system, activation in somatosensory touch areas gives rise not to the experience of being touched (or at least not only to the feeling of being touched) but to a visual experience of the scene in front of him. (Noë 2009, 58)

If this is right, then point (ii) of Noë’s argument goes through: what TVSS cases indicate is that, in the relevant sense, a change in underpinning neural activity isn’t necessary for a transformation in perceptual consciousness.

Points (iii) and (iv) do their ExPC-supporting work as a team. Given that what Noë means by a change in the underpinning neural activity is neural rewiring, and given that no neural rewiring has taken place in the TVSS case, there has been a transformation in perceptual consciousness without any change in the underpinning neural activity. If we then assume internalism (neuro-centrism) about the physical possibility is that, for Noë, there is no change in neural activity in the ferret case because, post rewiring, cells in (what would normally be) auditory cortex remain active, even though they now underlie visual experiences. On this alternative interpretation, ‘a change in neural activity’ would require that a different area of the brain become active. For present purposes, it really wouldn’t matter if we did interpret the target phrase in this way, since, as we are about to see (in the main text), Noë claims that TVSS cases involve no change in neural activity in this sense too, and so there would be no implications for my subsequent argument. However, my favoured interpretation, according to which a change in neural activity demands neural rewiring, fits significantly better with the observation that Sur’s ferrets are examples of neural plasticity (Noë 2009, 53, 58), which naturally indicates a change in neural activity. Moreover, there’s no doubt that, for Noë, TVSS cases involve no change in neural activity in the sense of there having been no neural rewiring.
machinery of phenomenal consciousness, and also that any perspective worthy of the name ‘naturalism’ rules out substance dualism, then it seems that we have no naturalistic explanation of the target transformation. So, from the perspective of naturalism, what confronts us is a mystery. That’s point (iii). If, by contrast, we accept externalism about the physical machinery of phenomenal consciousness, then it’s the addition of some new externally-located realizing factors, in the guise of the TVSS-delivering technology, that promises to supply the missing naturalistic explanation. That’s point (iv). But now, assuming the surely reasonable principle that, given some target domain of phenomena, we should defeasibly accept the existence of those states and processes that are required by the most powerful explanations that we have available in that domain (point (v)), we are permitted to draw the conclusion that we should accept ExPC (point (vi)).

With the whole argument in place (including the parts that tend to be hidden in Noë’s own treatment), one can appreciate, I think, why Noë concentrates on a case of sensory substitution to provide his model of a technology-driven transformation in perceptual consciousness (yielding the requirement that the transformation occur between sensory modalities), rather than on an example such as improving poor eyesight by putting on the right spectacles (which would yield a technology-driven transformation in perceptual consciousness within a single sensory modality). The proposal that the latter may occur without any neural rewiring engenders no sense of mystery analogous to the sense of mystery generated at point (iii) in the argument as stated. And what that means is that no momentum is established to include the spectacles as a constitutive part of the realizing machinery of phenomenal consciousness, as opposed to a causal input to such machinery, with the machinery itself remaining internal.

How should we respond to the argument from sensory substitution? I think we should grant points (i), (iv) and (v). TVSS cases do involve a transformation in sensory experience from one modality to another, ExPC would provide a genuine naturalistic explanation of the phenomenon, and we should let our metaphysics be guided by our best explanations. However, I shall argue that Noë’s own way of unpacking point (ii) – the claim that the relevant transformation is achieved without there being any change in the underpinning neural activity – seriously under-estimates the naturalistic theoretical resources available to the internalist. With those resources brought into view, point (ii) falls, and with it point (iii). There is a change in the underpinning neural activity, so there remains a perfectly respectable internalist account of the transformation in question and internalism does not result in any mystery. The consequence is that the argument from sensory substitution to ExPC fails, because internalism and externalism
are back on the level playing field instituted by Hurley’s reflections on the autonomy metaintuition.\textsuperscript{11}

As we have seen, according to Noë, the claim that the relevant transformation in phenomenal consciousness is achieved without there being any change in the underpinning neural activity should be interpreted as the claim that there has been no ferret-style neural rewiring. But this sort of rewiring surely does not exhaust the idea of a change in the underpinning neural activity. For example, let’s assume that what results from adaptation to TVSS technology is a fundamental change in the mathematical structure of the neural activation patterns in somatasensory cortex. We know, from well-established connectionist research in artificial intelligence and cognitive science, that phase transitions in the activation dynamics of neural-style networks can generate very different systemic behaviour, without any structural reconfiguration of the units and connections (without any neural rewiring). It is far from crazy to think that such phase transitions might produce transformations in the modality of the perceptual experience, from, say, touch to vision or quasi-vision. Indeed, as long as we continue to hang on to functionalism as our explanatory creed, such a possibility looks eminently sensible. And even if, with the zombies at the door, one doesn’t want to be quite so gung-ho about functionalism, Hurley’s autonomy metaintuition (this time deployed in the service of internalism) might serve to blunt any worries. For if we have no understanding of how phenomenal qualities could be realized in physical or functional elements, why is the intuition so strong that if phenomenal qualities can be realized physically at all, that realization must be specified at the level of neural wiring, rather than at the level of patterns of activation? I conclude that the argument from sensory substitution, as stated, gives us no decisive reason to accept ExPC.

4. The Argument from the Relational Character of Experience

So far I have considered the argument from sensory substitution in isolation from Noë’s enactivist view that perceptual experience is constituted by implicit knowledge of what he calls sensorimotor contingencies – the law-like effects that either my movement or the

\textsuperscript{11} Another way of responding to Noë’s argument would be to submit empirical evidence that, in TVSS subjects, some region of the brain other than somatasensory cortex – a region such as visual cortex that might plausibly realize the transformed phenomenal quality – is somehow activated crossmodally along with somatasensory cortex. I discuss this strategy in (Wheeler 2015). For today I want to grant Noë as much as possible, including the claim that it is somatasensory cortex alone that is activated by the TVSS technology.
movements of objects in my sensory field have on the sensory input that I receive (see e.g. Noë 2004, 2009; O’Regan and Noë 2001). Although this respects the fact that, in his 2009 discussion of sensory substitution, Noë draws the externalist conclusion before introducing sensorimotor contingency theory (Noë 2009, 58-9), perhaps this is a mistake. Sensorimotor contingency theory is certainly marshalled, by Noë, as part of the wider context in which the argument from sensory substitution is developed. So perhaps the heavy lifting in the case for ExPC is in reality performed by sensorimotor contingency theory, and ultimately by its commitment to what I shall call the relational character of experience. Let’s see how this might work.

A nice illustration of the sensorimotor contingency approach to perceptual experience is provided by Noë’s response to Dennett’s wall of Marylins example (Dennett 1991; Noë 2004, 54-9). Imagine walking into a room where the wallpaper features nothing other than a repeated identical image of Marylin Monroe’s face, in the style of Andy Warhol’s famous silkscreen paintings. If anyone asks you about your visual experience, you will presumably report that you see a wall of identical Marylins. But although that is how you will report your experience, there is a snag. To see any particular Marylin would require foveating on it, since peripheral vision simply doesn’t deliver information of sufficient detail. However, the human visual system functions in such a way that you would be able to foveate on only a tiny number of the available images at any one time. So what explains your conscious report of seeing a wall of Marylins? Noë’s account goes like this: while one is attentively gazing at any particular point on the wall, one doesn’t in fact have a high-definition visual impression of a wall covered in identical Marylins. What one has is (a) a high-definition visual impression of a tiny part of the wall containing a Marylin or two, plus (b) a sense that there are other Marylins present, in part because one knows implicitly that one could access them if one actively looked around, that is, because one has mastery of certain sensorimotor contingencies.

Importantly, the sensorimotor contingency approach plausibly has the resources to explain the fact that TVSS subjects have experiences that are vision-like in at least some respects, since the sensorimotor contingencies in play in TVSS involve relations characteristic of visual phenomena such as looming and occlusion effects. If that is right, then we have a good reason to add sensorimotor-contingency theory to the mix. And if that theory licences externalism about the machinery of phenomenal consciousness, then we have a new argument for ExPC. But how do we connect sensorimotor contingency theory with externalism in this way? The first step is the observation that sensorimotor contingency theory requires a relational conception of phenomenally conscious experience. This seems clear: for the sensorimotor contingency theorist, conscious perceptual experience essentially involves mastery of law-like relations between movement and sensory input. The second and final step is to claim that the
The relationality of perceptual consciousness licences ExPC (Ward 2012). This is the argument from the relational character of experience.

Before we turn to the details of this argument, we need to deflect an in-principle worry. If we follow the lead of some thinkers in this area, the proposed appeal to relationality will establish the wrong conclusion, because it will secure not the extended (in space) character of the machinery of phenomenal consciousness, but rather the blanket non-spatiality of the phenomenon. In other words, phenomenal consciousness isn’t extended, or indeed wholly internal, since, in truth, it isn’t anywhere. Thus Froese et al. claim that ‘[i]f cognition [consciousness] is a relational phenomenon, it logically cannot be located inside the brain (or anywhere)’ (Froese et al. 2013, 1420; for similar claims, see di Paolo 2009, Thompson and Stapleton 2009). Here one shouldn’t be distracted by the fact that Froese et al. make their claim explicitly about cognition rather than consciousness. The allegedly compelling inference is from relationality to non-spatiality, so any relational phenomenon will be non-spatial, and, on the sensorimotor account, perceptual experience is relational. If this is right, the argument from the relational character of experience to ExPC is a non-starter.

Fortunately, it’s the inference from relationality to non-spatiality that, it seems to me, fails to leave the blocks (Wheeler 2011b). Consider: towing is a phenomenon that constitutively involves a relation or set of relations (e.g. between two cars and a rope), but there seems no doubt that towing takes place in space and that the realizing vehicle (or vehicles) are located in space. In other words, even though towing is constitutively a relational phenomenon, the machinery of towing is spatial. Examples of this sort could surely be multiplied indefinitely. Of course, as Thompson (2011, 218) notes, ‘it does not make sense to think of cognition [or consciousness] as spatially located in the way that the “vehicles” enabling cognitive processes [or consciousness] are spatially located’, but that’s just fine. It’s the spatiality of the vehicles (the machinery) that’s needed for ExPC. One might clarify the point here using conceptual material borrowed, once more, from Hurley (2010), who draws a distinction between claims about what determines phenomenal quality and claims about how those qualities are materially realized. Framed in terms of this distinction, which is, of course, closely related to other distinctions in the vicinity, such as the personal-subpersonal distinction and the constitutive-enabling distinction (e.g. McDowell 1994), what we have just learned is that even if phenomenal consciousness is relationally constituted at the what level, it can be spatially located at the how level. And with relationality and spatiality back on the same side, one is free to argue that relationality at the what level entails externalism at

---

12 Thanks to Peter Sullivan for the towing example.
the how level, that is, that the relational character of experience entails that we accept ExPC.

But why should the relational character of experience at the what level have externalist consequences for our understanding the machinery of phenomenal consciousness? The driving thought here is that what-level relationality is inconsistent with how-level (machinery-level) internalism. Here is Ward (2012, 746): ‘[i]f we adopt the [sensorimotor] enactivist’s conception of experience as a relationship between perceiver and environment then the subpersonal [how level, mechanistic] underpinnings of experience must include more than the internal properties of the subject, since those properties leave it underdetermined whether the requisite relationship obtains’.

Although I do not have the space here to do justice to all the subtleties of Ward’s discussion, as far as I can tell, and in general terms, his argument faces the following dilemma: we can interpret the idea of ‘the requisite relationship obtaining’ so that it delivers the intended inconsistency, but then the resulting relational view of experience has an unpalatable consequence; or we can interpret that idea in a way that avoids the unpalatable consequence, but then the intended inconsistency does not materialize.

To explain: Let’s begin by interpreting the idea of ‘the requisite relationship obtaining’ as demanding that, for the sensorimotor contingency theorist, a conscious experience happens when and only when the organism actually interacts with its environment. On this view, assuming that the wall of Marylins is stationary, having a conscious experience of that object requires actively looking around, so that the relevant sensorimotor contingencies are actually traversed. Given this unpacking of the key phrase, ExPC would indeed follow from the relational character of experience, since any appeal to purely internal machinery could, at best, tell us only that the organism is in a state such that, if certain appropriate interactions were to take place, it would have a conscious experience. But this has the unpalatable consequence that there is no perceptual consciousness without behaviour, which flies in the face not only of intuition, but also of Noë’s claim that what matters for perceptual experience (e.g. of a wall of Marylins) is the sense the perceiver has that certain sensory inputs are accessible, that is, available if one looked around. We can avoid the unpalatable consequence, and thereby stay closer to what Noë says about the wall of Marylins case, by interpreting the idea of ‘the requisite relationship obtaining’ as demanding only that the organism be poised to masterfully traverse the relevant sensorimotor contingencies. On this view, having a conscious experience of the wall of Marylins requires only knowing implicitly that one could access more Marylins if one actively looked around. Now, however,
there seems to be no barrier to understanding such knowledgeable readiness to act as fully compatible with internalism about the realizing machinery.\textsuperscript{13}

It is important to stress that the foregoing dilemma exists only for the advocate of the argument from the relational character of experience. Someone enamoured with the relational character of experience who wishes to be an internalist about the machinery of phenomenal consciousness can cheerily follow the second option. Moreover, I suspect that there is a perfectly good further relationalist option out there in philosophical space, which is to develop the notion of experiential relationality at the what level of analysis in such a way that it has no implications whatsoever for the outcome of the internalism-externalism debate over the underlying physical machinery. What does seem clear, however, is that the argument from the relational character of experience gives us no decisive reason to accept ExPC.

5. Endings and Beginnings

In the opening scene of \textit{Sunset Boulevard}, the body of murdered scriptwriter Joe Gillis floats lifeless in the swimming pool of Norma Desmond’s mansion. Is a similar fate awaiting the fan of extended consciousness? I have certainly not argued for the general conclusion that the machinery of consciousness is always wholly internal. Rather, I have sought to bring some clarification to the structure of the debate and to expose the shortcomings of two arguments for extended phenomenal consciousness – the argument from sensory substitution and the argument from the relational character of experience. If I am right, these arguments fall short. So my interim report on extended consciousness is that, if the externalist goal in this arena is to establish that the physical machinery underpinning the what-it’s-like-ness of experience is extended beyond the skin, then different arguments will be needed. The internalist is certainly not home and dry, but the externalist has significant work to do. Still, if there had been a sequel to \textit{Sunset Boulevard}, maybe we would have discovered that Joe survived.

Acknowledgments

Some passages in section 3 of this paper have been adapted with revision from (Wheeler 2015). For useful critical discussion of the ideas presented here, many thanks

\textsuperscript{13}Although the dilemma I have developed here is (as far as I know) new, the issue of what exactly is involved in the exercise of sensorimotor understanding already has a central presence in the debate over sensorimotor enactivism (see e.g. Hutto and Myin 2012, chapter 2).
to audiences in Aberdeen, Memphis, Stirling and Warwick, and especially to Philip Ebert and Peter Sullivan.

References


