

Draft of paper published in *Topoi*. Online first DOI: 10.1007/s11245-015-9356-x, 2015.

Michael Wheeler

**The Revolution will not be Optimised:
Radical Enactivism, Extended Functionalism and the Extensive Mind**

Division of Law and Philosophy
School of Arts and Humanities
University of Stirling
Stirling
FK9 4LA
UK

Email: m.w.wheeler@stir.ac.uk

Tel: +44 (0) 1786 467555

The Revolution will not be Optimised: Radical Enactivism, Extended Functionalism and the Extensive Mind

Michael Wheeler

Abstract

Optimising the 4E (embodied-embedded-extended-enactive) revolution in cognitive science arguably requires the rejection of two guiding commitments made by orthodox thinking in the field, namely that the material realisers of cognitive states and processes are located entirely inside the head (internalism), and that intelligent thought and action are to be explained in terms of the building and manipulation of content-bearing representations (representationalism). In other words, the full-strength 4E revolution would be secured only by a position that delivered externalism plus antirepresentationalism. I argue that one view in 4E space – extended functionalism – is appropriately poised to deliver externalism but not antirepresentationalism. By contrast, in the case of a competing 4E view – radical enactivism – even if that view can deliver antirepresentationalism, its pivotal notion of extensiveness falls short of establishing externalism. These conclusions are justified via an examination of, and by responding critically to, certain key arguments offered in support of their view (and against extended functionalism) by the radical enactivists.

Keywords

enactivism, extended cognition, extensive cognition, representation, vehicle externalism

1. The Alternative April Theses

On April 4, 1917, Lenin delivered his so-called ‘April Theses’ at two meetings of the *All-Russia Conference of Soviets of Workers’ and Soldiers’ Deputies*. These meetings took place less than two months after the fall of Imperial Russia, while the country was presided over (temporarily) by the reformist, liberal-socialist Provisional Government. On April 4, 2014, I gave an early version of the present paper at *Varieties of Enactivism – a Conceptual Geography*, a symposium held in London that was designed to map out the theoretical terrain occupied by the approach to cognition known as enactivism. Surprisingly, these two presentations are connected by more than a shared date.¹ Lenin’s April Theses were principles designed to optimise the impending Bolshevik revolution, that is, to guarantee that a genuine revolutionary transformation of society would ensue. He called for a new revolutionary Workers’ International, for a rejection of the Provisional Government, for a republic of Soviets, and for all political officials to be elected, to be displaceable at any time and to receive no more than the average wage of a skilled worker. Of course, revolutions come in many forms. Thus, some thinkers who have recently been marching under the banner of 4E (embodied-embedded-extended-enactive) cognition have suggested that contemporary cognitive science is in its own revolutionary situation, characterized by an incipient paradigm shift in its fundamental theoretical principles. So, in an affectionate nod to dear old Lenin, I intend to offer you what might be dubbed ‘the alternative April Theses’. These principles

¹ For dramatic effect, I am ignoring the mismatch between the modern calendar and the old Russian calendar.

detail what would be needed, if the 4E revolution in cognitive science is to be optimised, that is, produce a truly paradigm-shifting result.

To bring the alternative April Theses into view, consider two guiding commitments made by orthodox thinking in cognitive science: (i) the material realisers (or vehicles) of cognitive states and processes are located entirely inside the skin (paradigmatically, inside the head); (ii) intelligent thought and action are to be explained in terms of the building and manipulation of content-bearing representations. Let's call the view expressed by (i) *internalism* (and its advocates *internalists*) and the view expressed by (ii) *representationalism* (and its advocates *representationalists*). I suggest that the alternative April Theses should be conceived as demands for the rejection of internalism and representationalism. That is, they should be demands for *externalism* and *antirepresentationalism*. In the present context, externalism is the view that, sometimes at least, the material realisers (or vehicles) of cognitive states and processes are distributed over brain, body and world, and so are located partly outside the skin.² Antirepresentationalism is the view that the fundamental forms of intelligent thought and action are to be explained in ways that do not appeal to the building or manipulation of content-bearing representations.³ Any outcome that falls short of delivering these dual demands will be, to borrow a phrase from Rupert's sobering set of conclusions regarding the revolutionary implications of 4E approaches in general, 'more of a nudging than a coup' (Rupert 2009, 242).

So what are the chances of a coup? Rather like the revolutionary left in politics, the 4E cognition community is actually an umbrella movement encompassing a range of partially overlapping, partially diverging and partially conflicting views. Although the various camps may sometimes unite in fragile alliances against (what they jointly perceive to be) the reactionary orthodoxy, they are just as often to be discovered fighting amongst themselves.

² There are, of course, other forms of externalism. Most notably, content externalism is the view that the contents of mental states (e.g. what follows the 'that' clause in a belief attribution such as 'Sharon believes that water is wet') are *partly determined by* external factors, meaning that two thinkers who are internally identical in every way may nevertheless possess intentional states with different contents merely through being located in different environments. Here I am not concerned with this sort of externalism, but rather with what, for reasons that should be obvious from the main text, is sometimes called *vehicle externalism* (e.g. Hurley 2010).

³ For the purposes of this paper, and in line with the way the radical enactivists (more on them soon) tend to set things up, the expressions 'content-bearing' and 'representational' will be treated as equivalent. Roughly speaking, a state is content-bearing (is representational) when it has correctness conditions (conditions under which it can be said to represent accurately). In the cognitive-scientific literature, the term 'representation' is sometimes used in ways that involve extremely weak correctness conditions. For example, sometimes representation seems to require nothing more than causal covariance. Against such an undemanding account of representation, revolutionary antirepresentationalism is a non-starter, since no one seriously doubts that there are causal covariances between inner states and external states of affairs. However, in the philosophical literature, *mere* causal covariance is widely held to be insufficient for genuine semantic representation (see e.g. Cummins 1989). If this is right, then conditions of causal covariance are, in truth, not correctness conditions at all – or at least not in the right sense for representational content. Something richer is required. It's that 'something richer', whatever it turns out to be, that the antirepresentationalist will want to target.

Given this fractured (and sometimes fractious) heterogeneity, there is insufficient room in just one paper to assess the revolutionary credentials of each and every group in the 4E family. So I propose to focus on just two of the available positions, namely *radical enactivism* (Hutto and Myin 2013, Hutto et al. 2014) and *extended functionalism* (Clark 2008a, b, Wheeler 2010a, b, 2014). What makes these two positions especially interesting is that the radical enactivists develop their case in part by diagnosing (what they take to be) the failure of the extended functionalists to secure externalism against the internalist counter-revolution. According to the diagnosis on offer, as I shall interpret that diagnosis in what follows, it is because the extended functionalist fails to turn her back on representationalism that she finds herself unable to secure her position against the key moves made by her internalist critics. The pivotal claim here is that there is an intimate link between internalism and representationalism, such that if one accepts representationalism, one is pulled inexorably toward internalism. Put another way, the only stable externalism will be one that is founded on antirepresentationalism. And it is just such an externalism that the radical enactivist aims to provide, via what they call the *extensiveness* of cognition, an idea to which we shall return. If all this is right, then not merely is it true that the 4E revolution will be optimised only when the dual demands specified in the alternative April Theses are both met, it turns out that meeting one of the component demands (externalism) will require meeting the other (antirepresentationalism). But is this right? Is the radical enactivist critique of extended functionalism compelling? And is the extensive mind in good externalist shape? These are the questions that the present paper will ultimately endeavour to answer.

2. Meet the Neighbours

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 externalist view that there are actual (in this world) cases of intelligent thought and action, in which the material vehicles that realise 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.⁴ It has become a common (although far from universal) move in the 4E literature to link ExC in some way to what might justifiably be called the house philosophy of mind in cognitive science, namely functionalism. This explains why ExC is sometimes cashed out as extended functionalism (Clark 2008a, b, Wheeler 2010a, b, 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. Of course, historically, the assumption (in philosophy and cognitive science) has tended to be that, when it comes to human beings and other, minded, biological life on earth, the functionally identified states and processes that matter for cognition will be realised by the nervous system. In truth, however, there isn't anything in the letter of functionalism that automatically mandates this internalism (Wheeler 2010a, b, 2014). 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

⁴ For the maiden voyage of ExC, see (Clark and Chalmers 1998). For a field-defining collection of papers on the topic, see (Menary 2010).

boundaries of the functionally identified system of interest may in principle fall beyond boundaries of the organic sensory-motor interface.

What this tells us is that functionalism makes extended cognition a possibility. The way in which it does this can, however, be specified more carefully. The fan of ExC will standardly want to hold onto the 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. Consider, for example, mathematical reasoning: 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 on only my organic resources. Now, if we are to describe these alternative problem-solving routines as two realisations 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, one that is spread out over brain, body and world, and one that is wholly neural. In other words, the target cognitive phenomenon must be multiply realisable. But if the advocate of ExC requires multiple realisability, she will be tempted by functionalism. For while functionalism may not be necessary for multiple realisability, one of the things that we all learn in introductory philosophy of mind classes is that it is routinely thought to be sufficient. That's why functionalism is what protects the very idea of alien minds from the parochial prejudices of geocentric chauvinism.

Of course, just as accessing a philosophical view according to which it is possible for Martians, Klingons and the Ood to have cognitive lives is not the same as establishing that such creatures actually exist, accessing a philosophical view according to which it is possible for cognition to be extended is not the same as establishing that cases of extended cognition actually exist. For the latter, and for reasons that will be discussed later (section 3), we need to augment (what we might call) raw extended functionalism with some extra conceptual machinery, in the form of what Adams and Aizawa (2008) have famously dubbed a *mark of the cognitive*, a scientifically informed account of what it is for a material element to be a proper part of a cognitive system that, so as not to beg any questions, is independent of where any candidate element happens to be spatially located. Of course, once a mark of the cognitive (so conceived) is specified, it still takes further philosophical and empirical leg-work to tell us where cognition (so conceived) falls – in the brain, in the non-neural body, in the environment, or, as ExC predicts will sometimes be the case, in a system that extends across all of these aspects of the world.⁵

⁵ I have made the point that extended functionalism needs to be augmented by a mark of the cognitive, in order to deliver more than the conceptual possibility of cognitive extension, a number of times in published pieces over recent years (Wheeler 2011a, 2013, 2014, and for the same point minus the language of the mark of the cognitive, see Wheeler 2010a, b). It is therefore surprising to find Hutto et al. (2014, 5) describing what they clearly (but mistakenly) take to be a shortfall in my treatment of the issue, when they write that 'Wheeler's analysis sets the stage for the extended mind debate, but without a substantive theory of cognition... empirically based functionalism only allows for the logical possibility that minds might extend – it does nothing more to promote the fortunes of extended functionalism'. Of course, to complete a functionalist argument for the existence of extended cognition in the actual world, one would need to provide, among other things, a suitable mark (or some suitable marks) of the cognitive. I will not attempt to do that here. That's one reason why I shall speak of extended functionalism as being 'appropriately poised' to deliver

Functionalism enjoys a number of conceptual and historical links with computationalism, where computationalism is the view that cognitive processes are a subset of computational processes. Thus Putnam's (1967) classic presentation of functionalism famously uses a Turing machine as its illustrative example, indicating that the set of computational systems is plausibly conceived as a subset of the set of functionally defined systems.⁶ In addition, computationalism is a key assumption in much of cognitive psychology. (This is one way of developing the thought that functionalism is the house philosophy of mind in cognitive science.) Add in the defensible, if not unassailable, thought that there is no computation without representation (e.g. Fodor 1981, Wheeler 2005, O'Brien and Opie 2009), and we have a line of reasoning which explains why functionalists standardly maintain a commitment to some variety of representationalism. Extended functionalists are no different, although of course the representationalist palette is now expanded to encompass both internal and external representations. This already tells us something illuminating, namely that, on the face of things, the extended functionalist has no theoretical interest in pursuing antirepresentationalism. And it is at this point that the radical enactivist's hackles are well and truly raised. Indeed, as I shall explain, it is, according to Hutto et al. (2014), precisely because extended functionalism endorses representationalism that (they claim) it ultimately fails to deliver externalism (for related considerations, although without the specific focus on extended functionalism, see Chemero 2011). Before considering the arguments, however, we need to bring radical enactivism itself into view.

In the most general terms, a position is enactivist if it pursues some version of the claim that cognition unfolds (is enacted) in looping interactions between an active organism and its environment. For the enactivist, then, cognition depends on a tight and dynamic relationship between perception and action. Moving beyond this broad characterization, it is increasingly common to find analyses distinguishing between three varieties of the view: sensorimotor enactivism, autopoietic enactivism and radical enactivism. Very roughly, sensorimotor enactivism (e.g. O'Regan and Noë 2001, Noë 2004) is rooted in the thought that perceptual experience is constituted by implicit knowledge of so-called sensorimotor contingencies – the law-like effects that either my movement or the movements of objects in my sensory field have on the sensory input that I receive. By contrast, autopoietic enactivism (e.g. Varela et al. 1991, Di Paolo 2005, Thompson 2007) is based on the idea that cognition is a process of sense-making by adaptively autonomous systems, where an autonomous system is a network of interdependent processes whose recurrent activity (a) produces and maintains the very boundary that determines the identity of that network as a unitary system, and concurrently (b) defines the ways in which that system may encounter perturbations from what is outside it while maintaining its organization. A system is adaptively autonomous when it is able to alter its behaviour in response to changes in its environment, in order to improve its situation. Although both sensorimotor and autopoietic enactivism are influential and productive views

externalism. (This will be important later, at the end of section 3.) For a completed functionalist argument for ExC, see e.g. (Wheeler 2011a, 2014). I should note that the need for a mark of the cognitive, and indeed the basic coherence of the notion, is a matter of debate, even *amongst* extended functionalists. For a relevant exchange of views on this topic, see (Clark 2008b, Wheeler 2011b, Clark 2011).

⁶ Not everyone thinks that the historical linking of functionalism and computationalism is a good thing. For example, Piccinini (2004) argues that they are logically independent doctrines.

that are worthy of discussion in their own right, our concern here will be primarily with radical enactivism.

Radical enactivism (Hutto and Myin 2013, Hutto et al. 2014) is marked out by its steadfast commitment to antirepresentationalism (to a no-content view) about *basic* cognition, where basic cognition is to be equated not with ‘simple’ cognition but with ‘the most fundamental kinds of minds’ (Hutto et al. 2014, 1). Such minds are phylogenetically and ontogenetically basic, but they may be organisationally and behaviourally sophisticated. An example of basic cognition given by the radical enactivists themselves (Hutto and Myin 2013, 42-3), namely cricket phonotaxis (the capacity of female crickets to find mates by tracking an auditory signal produced by the male), will feature later in this paper. Further proposed instances include the fine-grained motor activities that characterise skilled human manual know-how (Hutto and Myin 2013, 46-50).⁷ Importantly, the radical enactivist claim that basic cognition is nonrepresentational in character should not be taken as implying that there is no such thing as representational cognition. Rather, given that basic cognition is nonrepresentational in character, contentful thought emerges as a special achievement of minds that are scaffolded through structured and complex manipulations of external representational resources. The radical enactivist commitment to antirepresentationalism about basic cognition is allied with a strongly enactivist interpretation of the idea that cognition is embodied. For the functionalist, and thus for the extended functionalist, the physical body is relevant ‘only’ as an explanation of how cognitive states and processes are implemented in the material world. Of course, the extended functionalist is more liberal than her neurocentric internalist cousins about which parts of the body might count among the vehicles of cognition, since non-neural bodily structures and movements will be part of the extended realising machinery, but that outcome does not herald a fundamental change in our understanding of the relationship between cognition and embodiment (Wheeler 2014). By contrast, as Hutto and Myin (2013, 5-6) explain, ‘for enactivists, embodiment is not defined with reference to an intuitive, everyday understanding of bodies and their boundaries, but in terms of wide reaching organismic sensorimotor interactions that are contextually embedded’. Enactively speaking, such wide reaching sensorimotor interactions are held to be literally constitutive of cognition, which suggests at the outset that enactive embodied cognition has some sort of externalist credentials. As Hutto and Myin (2013, 7) put it, ‘mentality – with the possible exception of phenomenality... – is in *all* cases concretely constituted by, and thus literally consists in, the extensive ways in which organisms interact with their environments, where the relevant ways of interacting involve, but are not exclusively restricted to, what goes on in brains’. The term ‘extensive’ is not an idle cog in this passage. Rather, it expresses what the radical enactivists are keen to present as a nonrepresentational alternative to, and thus as an improvement upon, the functionalist notion of cognitive extension. The concept of cognitive extensiveness will exercise us at length later in this paper.

It is worth pausing here to comment on the following issue. I can imagine the radical enactivists being more than a little tetchy about my decision to treat them as externalists, *in the precise sense in which I have defined the term*, that is, as making a claim about the *material vehicles* of cognition. Any such tetchiness will be traceable to the (in my view, unnecessarily restrictive) thought, regularly expressed by the radical enactivists, that talk of vehicles already buys into a representationalist perspective that they have sworn to reject.

⁷ For an analysis which concludes that the radical enactivists do not have the theoretical resources to delineate the class of basic minds, see (Froese 2014). Froese argues that the necessary correctives are to be found in autopoietic enactivism.

Thus: ‘if basic minds lack content then it is a mistake to talk of vehicles in this domain; the vehicle/content distinction has no purchase here’ (Hutto and Myin 2013, 151). And similarly: ‘[t]he fate of vehicles depends on the fate of content’ (Hutto and Myin 2013, 20). So let me stress that I am using the term ‘material vehicle’ in a maximally liberal way, one that makes no assumption that the cognitive phenomena that we can usefully speak of as being carried by such vehicles are representational in form. This is just another way of saying that I am using the phrase ‘the material vehicles of cognition’ to mean the same as the phrase ‘the material realisers of cognition’. So, if, or where, there are representations, then there are vehicles of content and realisers of content-involving states and processes, while if, or where, there are no representations, then there are material realisers or vehicles of other kinds of states and processes, such as, perhaps, certain sorts of nonmodular self-organising architectures characterised by high degrees of entangled context-sensitive continuous reciprocal causation.⁸ Given that, on this view, there is no necessary connection between content and vehicles, there is no good reason to hold that the wide-reaching nonrepresentational sensorimotor interactions that the radical enactivist takes to be literally constitutive of cognition do not have material vehicles. The fate of vehicles does not depend on the fate of content, and the radical enactivists qualify as aspiring vehicle externalists. I shall tidy up this idea in section 4. Right now, a different issue warrants our attention.⁹

According to the radical enactivists, many of their more conservative enactivist colleagues have been inclined to talk the antirepresentational talk while failing to walk the antirepresentational walk. For example, Hutto and Myin (2013, chapter 2) charge sensorimotor enactivists such as Noë with failing to rule out the idea that knowledge of sensorimotor contingencies involves representations, in the form of behaviour-mediating content-involving inner states with the character of knowledge-that (e.g. knowledge that if I move in a particular way, my sensory input will change in a certain law-like manner). Of course, if this is correct as a reading of Noë, it signals not only a retreat into representationalism on the part of sensorimotor enactivism, but also a slide into internalism. It is thus a revolution betrayed.¹⁰ It is also the moment where something that is especially

⁸ For the same view of vehicles, see (Hurley 2010). For arguments that nonmodular self-organising architectures characterised by high degrees of entangled context-sensitive continuous reciprocal causation fail to reward a representational understanding of their dynamics, see e.g. (Wheeler 2005, Chemero 2011).

⁹ I could, of course, have developed the point about vehicles another way. I could have agreed with the radical enactivists that vehicle-talk entails content-talk, by taking a vehicle to be the kind of realiser where what is realised is contentful. Then it would have turned out that extended functionalism is externalist about both realisers and vehicles, while radical enactivism, by virtue of its antirepresentationalism, is externalist about realisers but eliminativist about vehicles, at least in the case of basic minds. This way of setting things up would have been less likely to cause radical enactivist tetchiness, but it wouldn’t have changed anything of substance about the arguments. However, since I see no good theoretical reason to adopt the restricted notion of a vehicle, and since the term, in its liberal sense, already has currency in the extended mind literature (see note 8), I have followed a different path.

¹⁰ Both sensorimotor enactivism and autopoietic enactivism have complex and troubled relationships with both antirepresentationalism and externalism. For just a small sample of texts that explore the relevant issues, see (di Paolo 2009, Froese et al. 2013, Hutto and Myin 2013, Noë 2004, 2009, Silverman 2014, Thompson 2011, Thompson and Stapleton 2009,

important for the concerns of this paper comes out of the shadows. As mentioned earlier, and as now indicated by the tabled objection to Noë, the radical enactivists hold that the prospects for rejecting internalism and the prospects for rejecting representationalism are intimately connected. It is here that things start to get really interesting.

3. The Prospects for a Representational Externalism

In general terms, the radical enactivist argument against extended functionalism that I intend to examine in this section unfolds like this: extended functionalism is standardly developed in a representationalist register; but representationalism invites internalism; so extended functionalism will be unable to repel the most powerful of the internalist counter-moves to which it is subjected. In other words, according to the radical enactivists, extended functionalism fails to deliver externalism, and that failure is to be traced ultimately to an endorsement of representationalism. As it happens, the idea that there is an argument with precisely this shape to be found in the radical enactivist literature is, perhaps, not immediately obvious, so an essential part of the task ahead is to dig around in the core texts to reveal its presence.

The first premise of the target argument is that extended functionalism is standardly developed in a representationalist register. In relation to this claim, it is worth noting that Hutto et al. (2014, 5-7) present a detailed analysis which concludes that unless the kind of empirically oriented functionalism that plausibly supplies the generic framework for extended functionalism endorses representationalism, it has no way of individuating the cognitive. If that analysis is correct, extended functionalism will need to be representational in form. Of course, this is a stronger claim than we require, but, assuming that most extended functionalists are rational individuals who are swayed by good arguments, it would not only entail the first premise as stated. In any case, we have agreed already that extended functionalism is standardly representational in character, so this is uncontentious ground. But now what about the second premise – the claim that representationalism invites internalism?

Here is an illuminating passage of radical enactivist text: ‘if basic cognition is not inherently contentful then defenders of internalism lose their most compelling reason for supposing that cognition – at least in its primary phylogenetic and developmental forms – is an “inner” business (where “inner” is understood in the sense of implying a cerebral location of the vehicles of content)’ (Hutto et al. 2014, 2).¹¹ At first sight, it might seem that what’s being

Varela et al. 1991, Ward 2012, Wheeler 2010b, 2011c, 2015a, b). If the arguments of the present paper are right, radical enactivism has its own problems delivering externalism.

¹¹ It is worth pausing to deflect a possible objection to the way in which I shall interpret the radical enactivists. An unsympathetic reader might complain that when I say that ‘the radical enactivists hold that representationalism invites internalism’, what I really ought to say is only that ‘the radical enactivists hold that, *for the internalist*, representationalism invites internalism’ and so on. In other words, all claims regarding the relationship between content and the whereabouts of our cognitive machinery should be relativised to whichever side in the debate over ExC makes them, and should not, *in any circumstances*, be attributed to the radical enactivist as well. Followed through, the implication of this relativisation would be that the anti-ExC argument that I am attributing to the radical enactivists couldn’t reliably be attributed to them. The interpretative issue here is brought into focus by the phrase ‘their most compelling reason’ in the quotation currently under consideration in the main text. Does this mean ‘the most compelling reason that internalists have’, implying that the radical

claimed here is not that representationalism invites internalism, but rather the opposite, that internalism invites representationalism. After all, the text states that if cognition is not representational, then internalism is threatened, which transmutes neatly into the claim that representationalism is necessary for internalism – no representations, no internalism. But, if we reflect for a moment on the passage in question, it becomes clear that the claim that representationalism is necessary for internalism simply *wouldn't do enough work*. Indeed, for all that claim says, representationalism might be necessary for externalism too, in which case a conception of cognition as necessarily representational might well contribute to a mark of the cognitive that the extended functionalist would find attractive. But if all that were right, then merely endorsing representationalism shouldn't incline us *one way or the other* in the dispute between internalism and externalism, and so couldn't be the 'most compelling reason' that one could have for endorsing internalism. Therefore, it cannot be correct to interpret the passage under consideration as claiming *only* that representationalism is necessary for internalism. It must be interpreted as claiming that some sort of comparative advantage is bestowed on the internalist by the fact that a shared commitment to representationalism is in play. In other words, an endorsement of representationalism must in some way *invite* internalism, even if such a commitment is not, on its own, sufficient for internalism.¹²

enactivists agree that it is indeed the most compelling reason that the internalists have? Or does it mean only 'the reason that the internalists themselves take to be the most compelling', which is neutral with regard to what the radical enactivists think? As far as I can tell, the suggestion that claims like this should be relativised, here to the internalist, is exegetically off-track, *for certain statements that appear in the radical enactivist texts*, because it can make no sense of passages such as the following:

Ultimately, we agree with O'Brien and Opie (2009). In the absence of an appeal to content there is no obvious alternative way to ground claims about what constitutes the cognitive in a scrupulous scientific manner. As we aim to show, one consequence of this is that without appeal to a notion of content to supply the mark of the cognitive – or an adequate replacement notion that can play that role – there is no principled way to advance the claim that cognition is neurally based. How else, other than by appeal to content, might the claim that cognitive processes are "contained in the brain" be supported? How else could it be established that, as a matter of fact, cognition is always and necessarily brainbound? (Hutto et al. 2014, 2)

It would, I think, be bordering on the perverse to read this passage as simply a collection of reports of 'what the internalist thinks'. Of course, it is perfectly consistent for the radical enactivists to hold, for example, both (i) that representationalism invites internalism and (ii) that, for the internalist, representationalism invites internalism.

¹² Here we need to head off another potential interpretative challenge. Consider the following passage from Hutto and Myin (2013, 135-6):

Nearly all discussions of [ExC] to date have proved unproductive, typically ending in stalemate. This is because those involved in these debates have not satisfactorily addressed the root issue. Their focus has typically been on special cases in which internal cognitive activity is augmented by certain external resources. Do these hybrids result in extended cognitive processes or systems in which the external resources become partly constitutive of mind? Parties on either

If representationalism invites internalism, and given that the extended functionalist is not inclined (and perhaps not able) to follow the radical enactivists into the nonrepresentationalist land of milk and honey, then extended functionalism will fail to deliver externalism. That's our target argument put in place. But there's something not quite right. For if the extended functionalist can, does, and perhaps needs to appeal to the notion of representation as (part of) her mark of the cognitive, how exactly could representationalism invite internalism? After all, the bald contention that to accept representational content as a component of one's mark of the cognitive favours internalism looks, on the face of things, as if it simply begs the question against the extended functionalist who routinely counts external representational elements, as well as internal representational elements, as genuine parts of cognitive states and processes.

The resolution to this worry is to be found in the fact that, in the just-discussed passage from Hutto et al., basic cognition is described not simply as contentful, but as *inherently*

side of the issue do not question that cognition should be understood in terms of the processing of representational or informational content... By fully breaking faith with unrestricted [representationalism,]... [radical enactivism] enables the [extended mind] debate to move ahead.

This passage begins with an explicit suggestion that debates over ExC typically end in a stalemate between externalism and internalism, a stalemate explained by a shared commitment to representationalism. But this doesn't sound at all like a situation in which representationalism invites internalism. For if representationalism explains a stalemate, then representationalism should invite neither internalism nor externalism. This would be worrying for my suggested interpretation, were it not for the fact that, as far as I can tell, the idea that there is a stalemate between internalism and ExC is undermined by the radical enactivists' own reasoning.

Let's return to the passage above, the very passage in which the alleged stalemate is introduced. There, it is linked to representationalism via what Hutto and Myin later dub the *default internal mind assumption* (DIM), the assumption that basic minds are 'fundamentally brain-bound and... are extended only in exceptional cases' (Hutto and Myin 2013, 137). Now, first assume that I am right that, for Hutto and Myin, representationalism invites internalism. If representationalism invites internalism, then it is unsurprising that advocates of ExC who endorse representationalism nevertheless hold that mind is, in the first instance, internal. Now assume that I am wrong, and that the radical enactivist view is only that representationalism is necessary for internalism. Then the connection between DIM and representationalism that Hutto and Myin surely take to exist would be a complete mystery, since (for the reasons explained in the main text), there would be no pressure to believe that being a representationalist should make one at all 'internally inclined'. Given that, as we shall see later, an adherence to DIM is a key feature that distinguishes the functionalist notion of the extended mind from the radical enactivist's notion of the extensive mind, that part of the picture needs to hang together. For their own well-being, then, the radical enactivists surely need to be interpreted as holding that representationalism invites internalism. But that means that, by the radical enactivist's own reasoning, there is no stalemate between ExC and internalism. Here is what I think the radical enactivists, in light of their own reasoning, ought to say: with representationalism in place, the counter-revolutionary forces of internalism win; antirepresentationalism is necessary to defeat those forces.

contentful. Up to now I have ignored this qualifier, but it is time to do it justice. Consider another, related passage:

[T]he standard, and strongest, move internalists can make to motivate their position is to appeal to a notion of narrow or intrinsic content... The appeal to mental content features [crucially] in internalist arguments because it is needed to provide a principled “mark of the cognitive” – one which backs up and gives definition to demarcation claims about what is constitutive of, as opposed to merely causally supportive of, cognition. (Hutto et al. 2014, 2)

According to the radical enactivist, then, it is a notion of intrinsic (or inherent) representational content, rather than the notion of representational content *simpliciter*, that provides the basis for the ‘standard, and strongest, move internalists can make to motivate their position’.¹³ To appreciate what effect this clarification has on the balance of power, we need to understand what intrinsic content is, and get clearer about precisely how it might provide the basis for a mark of the cognitive that fixes the causal-constitutive boundary in an internalist-friendly manner.

Recall that a mark of the cognitive is a scientifically informed account of what it is for a material element to be a proper part of a cognitive system that, so as not to beg any questions, is independent of where any candidate element happens to be spatially located. Although the idea remains controversial in some quarters, the claim that a mark of the cognitive is needed to distinguish genuinely constitutive contributions to cognition from merely causal contributions is deployed by a range of internalist and externalist thinkers, including extended functionalists, in order to draw the distinction between embedded cognition and extended cognition (see e.g. Adams and Aizawa 2008, Rowlands 2010, Wheeler 2010a, b, 2011a, 2013, 2014). According to the embedded approach, intelligent thought and action are regularly, and perhaps sometimes necessarily, causally dependent on the bodily exploitation of certain beyond-the-skin props or scaffolds. For the embedded theorist, even if it is true that a particular problem could not have been solved without the support of some external resource, nevertheless that resource remains a noncognitive aid to some internally located thinking system. It does not qualify as a constitutive part of the cognitive architecture itself. But how are we to decide whether a particular causally distributed system is a case of internalist cognitive embeddedness or externalist cognitive extension? What we need, it

¹³ Eagle-eyed readers will have noticed two things. First, in the target passage, Hutto et al. write of the internalist appealing to a notion of *narrow or* intrinsic content. Secondly, I ignore the first element of the disjunction, despite the fact that ‘narrow content’ and ‘intrinsic content’ are different concepts. I think we can disregard the mention of narrow content. For one thing, when Hutto et al. go on to develop the point in question, they focus entirely on the notion of intrinsic content. Moreover, and remembering that we are thinking of internalism and externalism in their vehicular forms, there’s surely no obvious route from narrow content to internalism. Narrow content is content that is fixed without reference to the environment within which the system of interest is situated. So if, as the externalist will insist is the case, the system of interest is an extended system, then it is possible that all the content borne by states within that system may be fixed without reference to the environment of *that* system, without there being any adverse implications for externalism. (Clark (2010b) makes a brief remark about narrow content that points us in this direction.) So, framing the internalist case in terms of narrow content wouldn’t suit the aims of the internalist or the radical enactivist. Many thanks to Nico Orlandi and Andy Clark for discussion of this issue.

seems, is a mark of the cognitive. Then we can determine which parts of the causal machinery are the elements that enjoy cognitive status. On the one hand, internalism would be served by a mark of the cognitive that ended up relegating all beyond-the-skin elements to the status of causal scaffolds for the wholly internal cognitive machinery. On the other hand, externalism would be served by a mark of the cognitive that ended up counting some beyond-the-skin elements, along with various inner factors, as constitutive parts of our cognitive machinery.¹⁴

So where does the causal-constitutive boundary fall, if one adopts the view that intrinsic representational content is an essential component of the mark of the cognitive? To answer that question, we need to say more about the concept of intrinsic content, as that concept is understood by the radical enactivists. When they introduce the idea of intrinsic content (which, of course, they themselves reject, along with all other forms of content in relation to basic minds), Hutto et al. explicitly cite Adams and Aizawa, those stalwart internalist critics of ExC. Adams and Aizawa famously argue for a twofold mark of the cognitive. First, cognition ‘involves non-derived representations, representations that mean what they do independently of other representational or intentional capacities’ (Adams and Aizawa 2008, 31). Examples of nonderived representations include thoughts, experiences and perceptions. Second, cognitive states and processes are ‘individuated by specific kinds of information processing mechanisms’, namely those identified by human cognitive psychology (Adams and Aizawa 2008, 31). The second condition in Adams and Aizawa’s account won’t concern us here, but the relevance of the first becomes clear once one realises that nonderived representation is what Adams and Aizawa, and thus the radical enactivists, mean by intrinsic representation. The transition from the claim that intrinsic/nonderived representation is an essential component of the mark of the cognitive to the triumph of internalism is then effected by way of a contrast between nonderived representations and derived representations, with the latter being representations whose content ‘arises from the way in which items are handled or treated by intentional agents... who already have thoughts with meaning’ (Adams and Aizawa 2008, 32). Examples of such derived representations include road signs, ordinance survey maps and flags, items whose representational content depends on thoughts, experiences and perceptions, that is, on states with nonderived content. According to Adams and Aizawa, derived representation is not itself a mark of the cognitive. Moreover, in the kinds of distributed systems beloved of extended functionalists, all the nonderived content will be found in the brain, with any external representational elements

¹⁴ Given that, in the quoted passage beginning ‘the standard, and strongest, move internalists can make’, Hutto et al. claim only that internalism depends on a mark of the cognitive, without mentioning externalism at all, one might wonder whether the radical enactivist’s own view is that the concept of a mark of the cognitive is something of which the externalist should be suspicious. However, when push comes to shove, it is not at all easy to dismiss the need for such a notion, or at least for something very like it. For although it is true, as Hutto et al. claim, that, without a mark of the cognitive to settle the issue of whether the dependence of cognitive processing on certain beyond-the-skin elements is constitutive or merely causal in nature, there are no principled grounds for internalism, it is equally true, as Adams and Aizawa (e.g. 2008) have repeatedly stressed, that, without a mark of the cognitive to settle the same issue, there are no principled grounds for externalism either. Indeed, it is worth noting that Hutto et al. (2014, 2) are happy to characterize basic cognition as ‘constitutively world-involving’, which at least suggests an adherence to some sort of constitutive-nonconstitutive distinction, although admittedly not necessarily to the view that a mark of the cognitive is required to navigate it.

realising only derived content. Thus an environment-involving solution to long multiplication (such as shuffling mathematical symbols around on pen and paper) might well contain cognitive elements – elements that involve the manipulation and transformation of nonderived representations; but these representations will be located in the mathematician’s brain, not in the environment. The external representational factors in this problem-solving system are not themselves (we are told) cognitive, because they do not involve the right kind of content. They are derived representations which get their content from the nonderived representations in the mathematician’s brain. Therefore (at best) we have a case of embedded, rather than extended, cognition. In other words, if one accepts that nonderived representation is an essential component of the mark of the cognitive, then one is on a fast ride to internalism.¹⁵

At this juncture, one might be moved to complain that, whatever the internalist believes, the extended functionalist remains perfectly at liberty to adopt a different notion of representation, one that fails to introduce any such internalist bias.¹⁶ However, it is not at all straightforward for the extended functionalist to resist the thought that nonderived representation must be an essential component of any mark of the cognitive. For exactly how would such a dismissal go? The most obvious strategy would be to attack the very idea of nonderived representation. If there is no such thing as nonderived representation, then nonderived presentation cannot be an essential component of the mark of the cognitive. But if the upshot of rejecting nonderived representation altogether is, as it seems to be, that all representation is derived representation (all content is derived content), then a kind of pincer movement seems to make the strategy unappealing. On the one hand, the notion of nonderived representation is actually relatively undemanding (nonderived representations are simply representations that mean what they do *independently of other representational or intentional capacities*), and so may be satisfied by a range of off-the-shelf theories of content, such as Dretske’s (1988) indicator function semantics or Fodor’s (1987) asymmetric causal dependency theory (for discussion of these and other options, see Adams and Aizawa 2008 35-9). On the other hand, the idea that all representational content must be derived from other representational content that is itself derived from other representational content (and so on), smacks of an infinite regress (cf. Searle 1992). So, it looks as if by advocating representationalism at all, the extended functionalist ends up having to concede that nonderived content is an essential component of any mark of the cognitive. The fast ride to internalism may have been replaced by a slippery slope, but whether the extended functionalist gets there quickly or slowly, she ends up in the same place, a place where representationalism invites internalism and where extended functionalism fails to deliver externalism.

¹⁵ The externalist might try to block the present argument for internalism by maintaining that, under certain circumstances, external elements may themselves carry nonderived content. This might be an attractive move for the externalist to make, given that the notion of nonderived content may be given a relatively undemanding reading (see later in main text), and indeed it is a move is made by Clark (2010b). However, I have decided to grant, and then to respond to, the internalist’s best-case scenario, which is that if there is such a thing as nonderived content, then all of it will be found in brains. Thanks to Andy Clark and Bill Ramsey for discussion of this point.

¹⁶ In effect, this would be a way of arguing that the radical enactivists were correct all along when they claimed that internalism and extended functionalism are locked in a stalemate (see note 12 above).

How should the extended functionalist respond? It seems that she can allow both that there is a genuine distinction between nonderived (intrinsic) and derived representation, and that the bearing of nonderived content is sufficient for an element to achieve cognitive status. What she needs to deny is that the bearing of nonderived content is necessary for an element to achieve cognitive status. Indeed, it strikes me as eminently plausible that a genuinely cognitive structure or process may feature elements which carry only derived content, just so long as those elements function systematically alongside other elements of that structure or process which carry nonderived content. This is one case where a heuristic appeal to what we would intuitively say about certain imaginary innards helps to direct our thinking.¹⁷ Consider Clark's (2010a, b) example of an imaginary Martian whose internally located memory systems involve bit-mapped images of printed text which are created and internally stored, and then later retrieved in the form of bit-mapped signals sent to and interpreted by the Martian visual cortex. Pretheoretically, we appear to have no hesitation in counting the bit-map system as a genuinely cognitive mechanism and the bit-maps themselves as genuinely cognitive structures. Nevertheless, as Clark observes, it seems clear that the arrangement in question contains representations (the bit-map elements) that have only derived content. This casts doubt on the thought that bearing nonderived content is necessary for an element to achieve cognitive status. But if this is right, then the door is once more opened to extended functionalism, via the manifest possibility of extended cognitive systems that have the same mixed representationality as the Martian bit-map system, in that they contain both nonderived and derived representations, even if the latter do inherit their content from the former, and even if all the former are internally located.

On the face of things, Adams and Aizawa now face a choice: they can either give up the claim that derived representations are noncognitive elements, which would respect the intuition that the Martian bit-maps be granted cognitive status, but would constitute an invitation to extended cognition; or they could refuse to grant the contested Martian states cognitive status, which would be to hold the internalist party line, but would fly in the face of some seemingly strong intuitions about the states in question. Unsurprisingly, perhaps, it is the latter course that they pursue: 'these Martian representational states are not cognitive states' (Adams and Aizawa 2008, 49). However, if we look beyond the face of things, one might reasonably wonder whether the move of withholding cognitive status from the Martian bit-maps is strictly available to Adams and Aizawa. For when they explain their own mark of the cognitive, they portray it as requiring only that '*at least some* components of cognitive states require *some* non-derived content' (Adams and Aizawa 2008, 50, my emphases). So the problem, for Adams and Aizawa, is that their own interpretation of their own nonderived content condition permits a situation in which some components of some cognitive states may possess only derived content, just so long as there is also a nonderived content component. But that's precisely the condition under which the bit-map Martian system counts as cognitive and which, therefore, allows for extended cognition (Clark 2010a, b).

This is bad news for Adams and Aizawa, but it also leaves the radical enactivist in something of a dilemma. If she accepts what seems to be Adams and Aizawa's own interpretation of how to appeal to nonderived content in formulating a mark of the cognitive, then the link from nonderived representation to internalism is broken and nonderived content becomes a welcome addition to the conceptual toolbox of representational externalism. But if she rejects that interpretation, and insists on a stronger version of the nonderived content condition,

¹⁷ Aficionados of the debate over extended cognition will recognise what follows as an application of the famous (or infamous) parity principle (Clark and Chalmers 1998).

according to which the only states that enjoy cognitive status are those with nonderived content, then while she can reinstate the link from nonderived content to internalism, she risks insisting on a mark of the cognitive from which even the most strident internalists shy away. Moreover, she invites a charge of question-begging. To see this, let's return to Clark's Martian. The proposed move on the part of the radical enactivist would mean that the inner bit-map – a state which does not itself contain any nonderived content – is rendered noncognitive. That would certainly block the path to externalism. However, it is desperately unclear that there is any motivation, independent of simply preserving the link between representation and internalism, for dividing up the Martian's cognitive architecture in this way, rather than in a way that treats the bit-map as part of a cognitive structure that features both derived and nonderived content. And that, I think, is the final nail in the coffin of the radical enactivist's idea that anyone, including the extended functionalist, who endorses representationalism is driven headlong into the internalist's clutches. It is time to bury that idea.

Funereal metaphors aside, it might seem that, in the bigger scheme of things, the outcome of the foregoing analysis falls short of what was promised. For although I may have shown that representationalism doesn't invite internalism, what I clearly haven't shown is that representationalism invites externalism, which might be taken to imply that the extended functionalist cannot really be appropriately poised to deliver externalism, or at least not in a way that the internalist is not equally appropriately poised to deliver internalism.¹⁸ However, anyone who drew this deflationary conclusion would have failed to register two important asymmetries between internalism and extended functionalism. To explain: We have seen that the very thought that was supposed to establish an irresistible slide from representationalism to internalism – namely the thought that nonderived representation is an essential component of the mark of the cognitive – provides a perfectly good basis for the truth of ExC. And since the case for internalism that we have been critically examining was paraded not simply as one pro-internalist move among many, but rather as the 'strongest... move internalists can make to motivate their position' (Hutto et al. 2014, 2), the internalist, as interpreted by the radical enactivists, has played her best card – and lost. By contrast, and this is where we meet our first asymmetry, the extended functionalist, as I have portrayed her, does *not* claim that representationalism invites externalism, but only that representationalism provides a perfectly good basis for externalism. For all this claim says, the same simple endorsement of representationalism might provide a perfectly good basis for internalism too, which tells us that the disagreement between internalism and extended functionalism will need to be settled by other considerations, namely by whatever else is packed into a defensible mark of the cognitive – whether that be certain nonrepresentational factors or the distinctive features of a certain kind of representation – and by the subsequent empirical and philosophical legwork that will subsequently establish where in space such systems are located. The huge task of developing such a story in a pro-ExC way has not even been attempted here.¹⁹ Nevertheless, the more limited result that has been achieved still constitutes a significant strategic victory for the extended functionalist. To see why, we need to enlist the help of our second asymmetry, one that obtains between the logical structures of our two generic positions.

Recall that, for internalism to be true, *all* the material realisers of cognitive states and processes must be located *entirely* inside the skin, whereas, for externalism to be true, only

¹⁸ This would herald yet another resurrection of the official radical enactivist position that internalism and extended functionalism are locked in a stalemate (see notes 12 and 16 above).

¹⁹ See note 5 for more on this issue.

some of the material realisers of cognitive states and processes must include beyond-the-skin elements. In a sense, then, internalism is more demanding than externalism, since externalism can tolerate the existence of some wholly internal psychological realisers, while internalism cannot tolerate the existence of any extended realisers. Of course, this logical sense in which externalism is less demanding doesn't, in and of itself, make it more likely that externalism is true. But once we add in the additional weight of the conclusion that the internalist's best supporting argument has been found wanting, it seems right to say that the extended functionalist is appropriately poised to deliver externalism in a way that the internalist isn't appropriately poised to deliver internalism. Consider the following analogy. For there to be witches, some witches need to exist, although they may exist alongside non-magical folk. On the other hand, for all folk to be non-magical, there can be no witches. In and of itself, that logical asymmetry doesn't seem to increase the likelihood that there are witches. But now imagine coupling it to the additional thought that the very best reason that one could possibly have for not believing in witches has been removed. Although it might still turn out that there are no witches, since other reasons that one might conceivably have for dismissing all such talk might turn out to be good enough, surely, right now, as things stand, the pro-witch lobby would have justifiable cause for celebration. For 'pro-witch lobby', read 'advocates of extended functionalism'.

If the analysis presented in this section is correct, the radical enactivists are doubly wrong: representationalism does not invite internalism, and extended functionalism, with its representationalism fully intact, is, in the relevant sense, appropriately poised to deliver externalism. So far so good for extended functionalism. Perhaps, however, the victory celebrations are destined to be short-lived.

4. The Extensive Mind

Around now I expect the radical enactivists to be reaching for the following rejoinder: there is no representational content in basic minds, so, even if it is true, as I have argued, that there is no irresistible slide from representationalism to internalism, nevertheless representationalist extended functionalism is dead in the water. With the revolution in sight, this might be connected to a second line of argument: if one decides, for reasons that are essentially independent of the relationship between representationalism and internalism, to do away with the notion of content, at least in the case of basic minds, then externalism follows, or is at least strongly favoured. As Hutto et al. put it, 'once one goes non-representational there is no clear scientific rationale for, and no clear theoretical means of, thinking of cognitive activity as something smaller or shorter than world-involving extensive relational activity' (Hutto et al. 2014, 8; see also Hutto and Myin 2013, 151).

It may surprise some readers to learn that I am not going to respond to these thoughts by engaging directly with Hutto and Myin's argument that there is currently no naturalistically credible account of representational content (Hutto and Myin 2013, especially chapter 4). For it strikes me that even if Hutto and Myin are right about that, one thinker's trigger for a pessimistic meta-induction is another's occasion for an interesting research problem. Indeed, one might even adopt a vaguely Kuhnian tone, and argue that a well-established explanatory framework, even one that confronts acknowledged difficulties, should be replaced only when a new, competing framework offers robust evidence of its own explanatory and predictive superiority. In short, I take Hutto and Myin's negative arguments against the idea that basic minds are contentful to be inconclusive, even if they go through. What will finally decide the

issue is the positive character of radical enactivism as a conceptual and explanatory framework.

In what remains of this paper, I shall argue that radical enactivism currently fails to deliver the revolutionary transformation of cognitive-scientific explanation that is promised by its publicity machine, and thus that its positive character remains fundamentally unclear. More specifically, I shall argue that even if we grant the radical enactivist her antirepresentationalism, she does not currently have the theoretical resources to deliver externalism. In other words, just as the radical enactivist fails to secure the inference from representationalism to internalism in her critique of extended functionalism, she fails to secure the inference from antirepresentationalism to externalism in her account of the extensive mind. Time, then, to return to the concept of extensiveness itself and to ask the following question: does the revolution need a new 'E'?

As noted earlier, the notion of cognitive extensiveness is proposed by the radical enactivists as a nonrepresentational alternative to, and thus as an improvement upon, the functionalist notion of cognitive extension. So what precisely is the difference between functionalist cognitive extension and cognitive extensiveness, beyond the point that the former is standardly, and perhaps necessarily, developed in a representational register while the latter is supposed to be some sort of consequence of antirepresentationalism? As we saw previously, given that we are working explicitly with a liberal notion of a vehicle of cognition, one that is (a) equivalent to the notion of a realiser of cognition, and (b) stripped of any necessary connection with the notion of content, there seems to be no momentum behind the claim that radical enactivism is not a version of vehicle externalism alongside extended functionalism. So that cannot be the grounds for the distinction. Radical enactivism is simply a competing form of vehicle externalism.

To bring things into better focus, consider how Hutto et al. introduce the idea of extensiveness, in the context of the claim that extensiveness, in some sense, follows from antirepresentationalism, although here that claim is expressed in slightly weaker terms.

In all, from a [radical enactivist] perspective, to let go of the idea that basic cognition is necessarily contentful and representational in character is to remove a standardly assumed barrier to seeing at least this form of cognition as constitutively world-involving. Such a shift in perspective is tantamount to acknowledging that fundamental cases of perceiving and thinking are not necessarily grounded in, nor do they take the form of, representing aspects of the world or having contentful thoughts about those aspects. And to think this aligns perfectly with understanding cognitive processes as a form of wide reaching activity that is – at root – extensive and unbounded; thus extensive minds are not merely, occasionally and in special circumstances, extended. (Hutto et al. 2014, 2)

First we need to put aside the potentially distracting expression 'unbounded'. Whatever that term is supposed to signal here, it cannot mean that there are no spatial limits to the material realisers of cognitive states and processes, since in answer to the question 'Does this mean the extent of cognition has no limits?', Hutto et al. answer '[o]f course not' (Hutto et al. 2014, 10), although they stress that the spatial boundaries in question are 'messy' and 'rough-edged' (Hutto et al. 2014, 10). So the distinction between extended functionalism and extensive enactivism cannot be that the former takes our cognitive machinery to have spatial

limits that are located beyond the skin while the latter takes our cognitive machinery to have no spatial limits at all. And notice that talk of spatial boundaries that are ‘messy’ and ‘rough-edged’ doesn’t provide any new momentum to the idea that radical enactivism is not a form of vehicle externalism. It is a well-worn theme in the extended functionalist literature that extended cognitive systems are dynamic, soft-assembled structures whose untidy boundaries may shift in context-sensitive ways (Clark 2008b), and that doesn’t prevent extended functionalism from being a form of vehicle externalism. So, beyond the divergence over content, the differentiating observation must, I think, be this. For the advocate of extended functionalism as standardly understood, mind is a phenomenon whose primary ontological manifestation is inner, but which sometimes, in exceptional circumstances, spreads beyond the skin (Hutto and Myin 2013, 137; see also note 12 above). By contrast, for the radical enactivist, mind is ‘fundamentally, constitutively already world-involving’ (Hutto and Myin 2013, 137). In other words, mind is a phenomenon whose primary ontological manifestation is to be spread out over brain, body and world. This is in virtue of the fact that, for the radical enactivist, basic cognition is, as explained earlier, strongly embodied in an enactive sense, in that it is constituted by wide-reaching organismic sensorimotor interactions.

With this understanding of cognitive extensiveness in play, one argument that the radical enactivists offer in support of the extensive view is that it is justified by a correct philosophical account of our folk-psychological practices. Thus:

Arguably what Wittgensteinian and phenomenological investigations into our ordinary thought and talk about the mental reveal is that while the folk do conceive of cognition and perception as kinds of public activity, their spatial and temporal boundaries are messy, rough-edged and extensive... This is what justifies thinking of minds as naturally extensive – the idea that cognitive activity always already entangles embodiment, action, and world-involving resources and does not restrict itself only to what is inside the individual organism. (Hutto et al. 2014, 10)

Here we are on the edge of some murky philosophical issues concerning the relationship between, on the one hand, our philosophical analysis of the folk-psychological practices that regulate our commonsense attributions of mental states and processes and, on the other, our judgment about where the material realisers of those states and processes are spatially located. If the notion of extensiveness is in the business of grounding claims about the latter – and, to be a genuine competitor to the functionalist notion of cognitive extension, it surely has to be in that business – it is not at all obvious that one can simply read off the decisive considerations from the former. This remains true, I think, even if one wants to resist a simple naturalistic intervention according to which the latter is a matter for science, and science always trumps philosophy. The fact is that, without further argument, the fundamental entanglement of embodiment, action and world that allegedly emerges from the correct analysis of our folk-psychological practices may signal a constitutive entanglement of phenomena at the personal level (e.g. at the level of what it is for a person to perceive) without signalling a constitutive (or even a purely causal) spatial entanglement between brain, body and world at the level of the subpersonal realising machinery that underpins those personal-level phenomena. We can make these rather abstract worries more concrete by looking at a more specific feature that the radical enactivist takes to be a defining characteristic of extensive cognition, namely relationality.

The idea that cognition is an essentially relational phenomenon is implicit in the radical enactivist claim that basic cognition is constituted by wide-ranging embodied sensorimotor interactions that take the form of organism-environment or perception-action loops. Thus cognition is described as ‘world-involving extensive relational activity’ (Hutto et al 2014, 8). This relationality is, in fact, a common theme in the enactivist literature, and we can illuminate the current issue by looking at connected work. Thus, Froese et al. (2013, 1420, drawing on Di Paolo 2009 and Thompson and Stapleton 2009) claim that cognition is ‘a kind of adaptive relationship between an agent and its environment’. Moreover, in a structural echo of a key claim made by the self-identifying radical enactivists, Froese et al. depict this relationality as part of a view that, in contrast to extended functionalism, ‘takes the notion of an extended mind as its starting point, rather than as a curious exception’ (Froese et al. 2013, 1419). In other words, despite the variation in terminology, we are in the vicinity of the extensive mind. At this point in the proceedings, however, Froese et al. make a further claim that, on a certain understanding, is in tension with externalism – extensive, functionalist or otherwise. That claim is that cognition isn’t extended, or indeed wholly internal, since, in truth, it isn’t anywhere: ‘[i]f cognition 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). Now, I am not claiming here that Froese et al. and the other autopoietic enactivists just mentioned are themselves recommending that we should draw conclusions about the nature of our psychological machinery directly from the alleged relational character of cognition, but, if we did succumb to that inference, then what we would end up with would not be externalism. The relationality of cognition would be poised to secure not the distributed (in space) character of our cognitive machinery, but rather its blanket nonspatiality. In other words, the enactive rejection of internalism on grounds of the relationality of cognition would result not in the spatial repositioning, in an outward direction, of the boundary of the cognitive system (externalism), but rather in the jettisoning of the idea that there is a relevant spatial boundary to be repositioned.

This opens up the very issue about relations between explanatory levels that is our present concern, because the canvassed inference from the relationality of cognition to its nonspatiality is an inference from one explanatory level (the level of personal-level or organism-level phenomena) to another (the level of material realisers). Moreover, once one understands it as such, this inference is, it seems to me, deeply unconvincing (Wheeler 2011c, 2015a). 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 realising vehicle (or vehicles) are located in space. In other words, even though towing is constitutively a relational phenomenon, the machinery of towing is spatial.²⁰ Examples of this sort could be multiplied indefinitely. And the lesson surely generalises as far as cognition. As the autopoietic enactivist Thompson (2011, 218) notes, ‘it does not make sense to think of cognition as spatially located in the way that the “vehicles” enabling cognitive processes are spatially located’.

At first sight, this looks like a positive result for radical enactivism: if the relationality of cognition, as delivered by, say, an enactive personal-level analysis of folk-psychological practice, doesn’t entail the non-spatiality of the relation-realising machinery, then perhaps it does entail extensive externalism. However, for all that has been said so far, it is just as likely that the reason why the relevant notion of relationality doesn’t entail the relevant notion of nonspatiality is because analyses at the level of the former typically have no *decisive*

²⁰ Many thanks to Peter Sullivan for the towing example.

implications for how things stand at the level of the latter, even if it might have the beneficial heuristic effect of preemptively blocking any proposed inference from an internalist understanding of mind at the personal-level to an internalist account of our psychological machinery. In other words, even if the relationality of cognition, as identified by radical enactivism, is a genuine phenomenon, it is manifested at the wrong level of explanation to deliver extensive externalism. But if this is right, then not only are considerations of relationality poorly placed to support the extensive mind over the extended mind, they are equally poorly placed to support externalism regarding our cognitive machinery over internalism. Mere spatiality is consistent with all these options.

Faced with this challenge, the radical enactivist might try to provide further arguments to justify the level-crossing inference from personal-level enactive relationality to machinery-level extensive externalism. Alternatively, she might try to appeal directly to some kind of enactive relationality at the level of the underpinning machinery itself that would justify the extensive picture. For today, I intend to consider only the latter strategy.²¹

Hutto et al. (2014) refer approvingly to work in which psychological explanation is framed in terms of affordances – possibilities for action presented by the environment. The affordance concept is associated primarily with ecological psychology (see, canonically, Gibson 1979), but has recently found a second home within embodied cognitive science (see e.g. Chemero 2011, Rietveld and Kiverstein 2014). Affordances are, without doubt, relational structures, since what counts as a possibility for action will be fixed by the relations that hold between the embodied sensorimotor capacities of a particular organism and certain specific environmental structures. A chair affords sit-on-ability to me, but not to a blue whale. Affordances therefore point both ways, as Gibson (1979) once put it, toward the organism and toward the environment. Given also that there is a strong antirepresentationalist line of thought both in traditional ecological psychology (e.g. Gibson 1979) and in recent embodied cognitive science that makes use of the affordance concept (e.g. Chemero 2011), it seems reasonable to interpret such work on affordances in a manner that is extensive, in at least the sense that we are dealing with nonrepresentational structures that are constituted relationally. The question, then, is this: does psychological explanation in terms of affordances deliver externalism? In my view, the answer is no.

Here's the problem for the radical enactivist. It seems that the embedded internalist – the internalist who thinks that cognition is often causally dependent, although never constitutively dependent, on environmental scaffolding (see above) – is entitled to maintain her view, even though affordances must be relationally defined, and even if affordances are interpreted nonrepresentationally. Why is this? Affordances are possibilities for action presented to a creature by its environment. So, one might think, what cognitive science needs to identify are the mechanisms by which various creatures realise what we might call affordance-sensitivity. Of course, the cognitive scientist won't be able to specify what the affordances in play are, independently of knowing what embodied sensorimotor capacities a creature possesses in a particular environment – that's the force of the relationality point – but that's consistent with the claim that the genuinely cognitive elements of the behaviour-generating process all remain internal. On this internalist view, relationally defined affordances are structures to which inner psychological states and processes are attuned or to which they enable sensitivity. Now, Hutto et al. readily concede that there is an orthodox

²¹ For an attempt to provide the first kind of argument in the narrower case of conscious experience, see (Ward 2012). For a critical response to that attempt, see (Wheeler 2015a).

cognitive-psychological interpretation of affordances according to which the affordance properties of various objects are neurally coded for, but, in the present context, they can legitimately disregard that interpretation on the grounds that it assumes the notion of representational content. However, it would be harder – indeed, I suggest impossible – for them to ignore work which productively pursued an internalist account of affordances and affordance-sensitivity while thinking of the inner mechanisms concerned as nonrepresentational in character. So is there any such work? Yes, there is. If there's one thing we all absorbed from the classic research in situated robotics from the 1990s (see, canonically, Brooks 1991), it's that affordance-sensitive behaviour may be generated nonrepresentationally, by exploiting a simple and distinctive environmental cue using a nonrepresentational special-purpose inner mechanism that responds to that cue alone. Moreover, one of the most widely discussed examples of this kind of adaptive strategy – Webb's (1994) robotic explanation of cricket phonotaxis – is one of Hutto and Myin's flagship examples of nonrepresentational basic cognition (Hutto and Myin 2013, 42-3). It will be instructive to revisit briefly the relevant details of the cricket example.

Female crickets find mates by tracking an auditory signal produced by the male. The male's signal thus affords mate-finding to the female cricket. Part of Webb's nonrepresentational explanation for the female cricket's success in responding to this affordance appropriately is that she has two specialized interneurons (one connected to each of her ears) whose decay rates are tightly coupled with the specific temporal pattern of the male's signal, such that signals with the wrong temporal pattern will simply fail to produce the right motor-effects. Of course, the observed affordance-sensitive behaviour here is dependent on the reliable presence of a particular environmental factor (the male's song with its particular structure), but that is entirely consistent with an embedded, rather than an externalist, account of cognition. In other words, one can have a nonrepresentational account of sensitivity to affordances that does not result in externalism, even where affordances are conceived as relational structures. Once again, it seems, the radical enactivist fails to secure her target inference from antirepresentationalism to externalism in her account of the extensive mind.

At this point in the dialectic, the radical enactivist might be tempted to protest that all they intended to claim was that antirepresentationalism removes a barrier to externalism about basic cognition, not that antirepresentationalism entails externalism about basic cognition, so the fact that one can give coherent examples of nonrepresentational internalism in that domain is neither here nor there. But this response won't fly. For one thing, it sits uncomfortably alongside the radical enactivist's contention that antirepresentationalism 'aligns perfectly' (see above) with externalism, since a perfect alignment seems to suggest that any alternative (i.e., internalist) alignments will be discernibly problematic; but they don't seem to be. For another thing, we are considering the radical enactivist's own highlighted example of extensiveness – affordance-based psychological explanation – and surely one ought to expect *that* to be resistant, to some significant degree, to a nonrepresentational internalist understanding; but it doesn't seem to be. Most importantly, however, representationalism can be a barrier to externalism only if representationalism somehow invites internalism; but we have seen already that it doesn't. In sum, there seems to be no good reason to believe that 'once one goes non-representational there is no clear scientific rationale for, and no clear theoretical means of, thinking of cognitive activity as something smaller or shorter than world-involving extensive relational activity' (Hutto et al. 2014, 8).

5. Prodnalog

Revolutions don't always work out as planned. After the Kronstadt rebellion of March 1921, and after years of devastating civil war in Russia, Lenin adopted the New Economic Policy. This permitted the existence of private enterprises (including some that were funded by full or partial foreign capital) and replaced forced grain requisition (*prodrazvyorstka*) with a tax on farmers (*prodnalog*). If the arguments of the present paper are correct, then, in relation to what may be achieved by extended functionalism or radical enactivism, the 4E revolution in cognitive science looks to be destined for a similar fate. The two revolutionary demands expressed by our paradigm-shifting alternative April Theses were (i) externalism and (ii) antirepresentationalism. Extended functionalism is appropriately poised to deliver the first of these demands, but is under no pressure to implement the second, while even if radical enactivism can implement the second, it is unable to deliver the first. The revolution, it seems, will not be optimised.²²

References

- Adams F, Aizawa K (2008) *The bounds of cognition*. Blackwell, Malden, MA and Oxford
- Brooks R (1991) Intelligence without representation. *Artificial Intelligence* 47:139-59
- Chemero A (2011) *Radical embodied cognitive science*. MIT Press, Cambridge, Mass.
- Clark A (2008a) Pressing the flesh: a tension in the study of the embodied, embedded mind? *Philosophy and Phenomenological Research* 76 (1): 37-59
- Clark A (2008b) *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford University Press, New York
- Clark A (2010a) Coupling, constitution and the cognitive kind: a reply to Adams and Aizawa. In (Menary 2010)
- Clark A (2010b) Memento's revenge: the extended mind, extended. In (Menary 2010)
- Clark A (2011) Finding the mind. *Philosophical Studies* 152 (3): 447-461
- Clark A, Chalmers D (1998) The extended mind. *Analysis* 58 (1): 7-19
- Cummins R (1989) Representation and covariation. In: Silvers S (ed) *Rerepresentation: readings in the philosophy of mental representation*. Kluwer, Dordrecht

²² Many thanks to Tom Froese and to an anonymous referee for perceptive critical comments on an earlier version of this paper that have enabled me to clarify and strengthen the arguments. For useful discussion and feedback, thanks also to audiences at the following workshops: *Varieties of Enactivism – A Conceptual Geography*, Goldsmiths University of London, April 2014; *Mental Representations – The Foundation of Cognitive Science?*, Ruhr-Universität Bochum, September 2015.

- Di Paolo E A (2005) Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4 (4): 429-52
- Di Paolo E A (2009) Extended life. *Topoi* 28: 9-21
- Dretske F (1988) *Explaining behavior*. MIT Press, Cambridge, Mass.
- Fodor J A (1981) *Representations: philosophical essays on the foundations of cognitive science*. MIT Press, Cambridge, Mass.
- Fodor J A (1987) *Psychosemantics: the problem of meaning in the philosophy of mind*. MIT Press, Cambridge, Mass.
- Froese T (2014). Review of *Radicalizing enactivism: basic minds without content* by Hutto, D D and Myin, E. *The Journal of Mind and Behavior*, 35(1-2): 71-82.
- Froese T, Gershenson C, Rosenblueth D A (2013) The dynamically extended mind: a minimal modeling case study. *IEEE Congress on Evolutionary Computation* 1419-1426, IEEE Press
- Gibson J J (1979) *The ecological approach to visual perception*. Houghton Mifflin, Boston
- Hurley S. (2010) The varieties of externalism. In (Menary 2010)
- Hutto D D, Myin E (2013) *Radicalizing enactivism: basic minds without content*. MIT Press, Cambridge, Mass.
- Hutto D D, Kirchhoff M D, Myin E (2014) Extensive enactivism: why keep it all in? *Frontiers in Human Neuroscience* 8 (706): 1-11, doi:10.3389/fnhum.2014.00706
- Menary, R. (ed.) (2010) *The extended mind*, MIT Press, Cambridge, Mass.
- Noë A (2004) *Action in perception*. MIT Press, Cambridge, Mass.
- Noë A (2009) *Out of our heads: why you are not your brain, and other lessons from the biology of consciousness*. Hill and Wang, New York
- O'Brien G, Opie J (2009) The role of representation in computation. *Cognitive processing* 10: 53–62.
- O'Regan J K, Noë A (2001) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24: 939-1031
- Piccinini G (2004) Functionalism, computationalism, and mental states. *Studies in the History and Philosophy of Science* 35 (4): 811-833
- Putnam H (1967) Psychological predicates. In: Capitan W H, Merrill D D (eds) *Art, mind and religion*. University of Pittsburgh Press, Pittsburgh

- Rietveld E, Kiverstein J (2014) A Rich Landscape of Affordances. *Ecological Psychology* 26 (4): 325-352
- Rowlands M (2010) *The new science of the mind: from extended mind to embodied phenomenology*. MIT Press, Cambridge, Mass.
- Rupert R (2009) *Cognitive systems and the extended mind*. Oxford University Press, Oxford
- Searle J R (1992) *The rediscovery of the mind*. MIT Press, Cambridge, Mass.
- Silverman D (2014) *The sensorimotor theory of perceptual experience*. PhD thesis, University of St. Andrews, <http://hdl.handle.net/10023/5544>
- Thompson E (2007) *Mind in life: biology, phenomenology, and the sciences of mind*. Harvard University Press, Cambridge, Mass.
- Thompson E (2011) Reply to commentaries. *Journal of Consciousness Studies*, special issue on consciousness and life: commentaries on Evan Thompson, Mind in life, Schlicht T (ed), 18 (5-6): 176-223
- Thompson E, Stapleton M (2009) Making sense of sense-making: reflections on enactive and extended mind theories. *Topoi* 28: 23–30
- Varela F, Thompson E, Rosch E (1991) *The embodied mind: cognitive science and human experience*. MIT Press, Cambridge, Mass.
- Ward D (2012) Enjoying the spread: conscious externalism reconsidered. *Mind* 121 (483): 731-751
- Webb B (1994) Robotic experiments in cricket phonotaxis. In: Cliff D, Husbands P, Meyer J-A, Wilson S W (eds), *From animals to animats 3: proceedings of the third international conference on simulation of adaptive behavior*. MIT Press, Cambridge, Mass.
- Wheeler M (2005) *Reconstructing the cognitive world: the next step*. MIT Press, Cambridge Mass.
- Wheeler M (2010a) In defense of extended functionalism. In (Menary 2010)
- Wheeler M (2010b) Minds, things, and materiality. In: Malafouris L and Renfrew C (eds) *The cognitive life of things: recasting the boundaries of the mind*. McDonald Institute Monographs, Cambridge
- Wheeler M (2011a) Embodied cognition and the extended mind. In Garvey J (ed) *The Continuum companion to philosophy of mind*. Continuum, London
- Wheeler M (2011b) In search of clarity about parity. *Philosophical Studies*, symposium on Andy Clark's *Supersizing the Mind*, 152(3): 417-25.

- Wheeler M (2011c) Mind in life or life in mind? Making sense of deep continuity. *Journal of Consciousness Studies*, special issue on consciousness and life: commentaries on Evan Thompson, Mind in Life, Schlicht T (ed) 18 (5-6): 148-68
- Wheeler M (2013) Is cognition embedded or extended? The case of gestures. In: Radman Z (ed) *The hand, an organ of the mind: what the manual tells the mental*. MIT Press, Cambridge, Mass.
- Wheeler M (2014) Revolution, reform, or business as usual? The future prospects for embodied cognition. In: Shapiro L (ed) *The Routledge handbook of embodied cognition*. Routledge, Abingdon and New York
- Wheeler M (2015a) Extended consciousness: an interim report. *Southern Journal of Philosophy* 53 (Spindel supplement): 155-75
- Wheeler M (2015b) Not what it's like but where it's like: phenomenal consciousness, sensory substitution and the extended mind. *Journal of Consciousness Studies*, special issue on embodied, embedded and extended accounts of phenomenal consciousness, Silberstein M, Chemero A (eds), 22 (3-4): 129-47