

Breaking the Waves: Beyond Parity and Complementarity in the Arguments for Extended Cognition

Michael Wheeler

1. Andy Clark – He’s One of our Own

According to the hypothesis of extended cognition (henceforth ExC), the physical machinery of mind sometimes extends beyond the skull and skin. More precisely, there are actual (in this world) cases of intelligent thought and action, in which the material vehicles that realize the thinking and thoughts concerned are spatially distributed over brain, body and world, in such a way that certain external (beyond-the-skin) elements (paradigmatically, technological devices) are rightly accorded fundamentally the same status (i.e., cognitive status) as would ordinarily be accorded to a subset of your neurons. So, if ExC is true, then sometimes your mobile phone isn’t just an external information storage device that saves your poor old brain the trouble of storing all those phone numbers (although it is that), it is also literally part of your memory, and thus part of your mind, in the sense that it’s part of your mnemonic machinery.

ExC was given its ceremonial launch in a now-famous co-authored paper by Andy Clark and David Chalmers (1998), although it was Clark who really ran with the idea afterwards (see especially Clark 2008), and it is Clark who has done more than anyone to amplify, clarify and promote the view.¹ Over the years, ExC has generated a lively debate between its advocates and its gainsayers (see e.g. the developments, defences and critiques collected in Menary 2010), with the advocates of extended cognition dividing into several factions, each with its own favoured ways of expressing and arguing for the extension claim. Arguably the most prominent of these pro-ExC groups are the two that Sutton (2010) has helpfully dubbed the *first wave* and the *second wave* of ExC theorists. There are also third-wave (Kirchhoff 2012)

¹ Readers who are familiar with the original Clark and Chalmers treatment might worry that my opening paragraph mischaracterizes the position on offer, by riding roughshod over a distinction between cognition and mind that plays an organizing role in that treatment. After all (someone might complain), Clark and Chalmers first argue for extended cognition and then, *additionally*, for the extended mind, so glossing extended cognition as a view about the physical machinery of *mind* cannot reflect their conception of ExC. However, Clark and Chalmers (1998, 12) actually draw a distinction between cognitive processing (with the emphasis explicitly on the word ‘processing’) and on ‘truly mental states’ such as ‘experiences, beliefs, desires [and] emotions’. There are, in fact, many distinctions that might be in play here – including at least those between processes and states, the non-conscious and the conscious, the affective and (a narrow notion of) the cognitive, and the non-doxastic and the doxastic. Any attempt to regiment these different distinctions so as to fix a global distinction between cognition and mind looks to be (at best) artificially stipulative. Suffice to say (i) that Clark typically abandons the distinction between cognition and mind in later discussions of ExC, and (ii) that treating ‘mind’ and ‘cognition’ as interchangeable terms that pick out the same group of states and processes – the psychological ones – is standard practice in cognitive science.

and fourth-wave (Fritzman and Thornburg 2016) versions of the view, but, as I interpret them anyway, these positions don't have the same intimate relationship with Clark's own work that I shall be exploring here.

First-wave ExC is standardly characterized (by Sutton among others) as emphasizing and defending the kinds of arguments for extended cognition that were to the fore in the original Clark and Chalmers paper. Almost all of the attention here is concentrated on the so-called *parity principle*. Clark's more recent formulation of this principle is as follows: '[i]f, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process' (Clark 2008, 77, drawing on Clark and Chalmers 1998, 8). As has been repeatedly pointed out (e.g. by Menary 2007; by Wheeler 2011a, b; and by Clark himself – Clark 2008), the parity principle has sometimes been misunderstood in the literature, so let's tread carefully. The parity principle asks us to start by considering a distributed (over brain, body and world) system, one that generates some psychologically interesting outcome in such a way that an important functional contribution (such as information storage or information processing) is made by certain beyond-the-skin elements. We are then encouraged to imagine a hypothetical, functionally equivalent, but wholly inner system. In this second system, the specific functional contribution that is made in the distributed case by the previously identified external element is made by an internal one. Having taken this imaginative step, if we then judge that the internal element in the latter, hypothetical case counts as a genuine part of that agent's mental machinery, then we ought to conclude that the very same status – that is, cognitive status – should be granted to the highlighted external element in the environment-involving case with which we began. To do otherwise would be to succumb to neural chauvinism, which is tantamount to begging the question against ExC.

Let's follow Sutton in taking adherence to the parity principle to be the mark of first-wave ExC. By contrast, second-wave ExC theorists reject, or at least downplay significantly, the parity principle, in favour of considerations of either *complementarity* (Sutton 2010, Kiverstein and Farina 2011) or, in a closely related vein, *cognitive integration* (Rowlands 1999, Menary 2007). According to second-wave ExC, functional equivalence between outer and inner elements, as expressed in terms of the parity principle, is not necessary for extended cognition, so focussing on that notion may in fact mislead us as to the most important, exciting and illuminating features of extended cognitive systems. Put crudely, it's difference not sameness that matters. Sutton (2010, 194) expresses these points as follows: 'in extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes', so 'different components of the overall (enduring or temporary) system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting'. Adding a further dimension to, or perhaps making explicit an existing dimension of, complementarity, the integrationists emphasize the processes by which internal and external elements with different properties may be combined into a single (and thus integrated) cognitive whole. Thus while maintaining the second-wave emphasis on the importance of functional differences and complementarity, integrationists also foreground factors such as the completion of cognitive tasks through the skilled manipulation of external elements (especially cases in which the task couldn't be completed without the

external elements in question), the transformation of our cognitive abilities through the learning of such manipulative skills, and the application of norms of manipulation with a distinctively cognitive character (Menary 2007). Henceforth, and with apologies to any integrationists who are irritated by the move, I shall use complementarity in a way that is meant to subsume integration.

On the face of things, it looks as though even if first-wave and second-wave ExC aren't formally inconsistent (more on this as we go along), the two groups of theorists must at least disagree about the best way to argue for ExC. As Menary (2007, 57) puts it, a 'major difference between [first-wave] extended mind style arguments and [second-wave] cognitive integration is that the latter does not depend upon the parity principle'. And Sutton (2010, 205) interprets the parity principle as, at best, 'a temporary indicator of cognitive extension, a place-holder for fuller, more inclusive sciences of the interface'. But now here's an intriguing fact. First-wave theorists claim Clark as one of their own (he was, after all, one of the architects of the parity principle), *but so too do second-wave theorists* (see e.g. Sutton, 2010, 205, who focusses on a piece by Clark that I shall discuss in the next section, and Menary 2007, 57). If these attributions are all correct, Clark is simultaneously a first-wave (parity-based) and a second-wave (complementarity-based) theorist. But if the tension between first-wave and second-wave ExC is genuine, this position is, at best, an uncomfortable one for Clark to occupy. The goal of the rest of this paper is to explore this apparent discomfort.

2. Walking Both Sides of the Street

During an illuminating reply to some commentators, Clark (1998) considers an early challenge to (what we now know as) first-wave ExC, a challenge mounted by O'Brien (1998). O'Brien focusses on a specific example of external information storage that Clark and Chalmers (1998) judge, largely by appeal to parity considerations, to be part of an extended cognitive system, namely the linguistic inscriptions in a reliably and uncritically accessed notebook. These external elements allegedly store the content of one of the agent's dispositional beliefs. For present purposes, the only thing that matters about this example (which many readers will recognize immediately as the contested case of Otto) is the form of information storage instantiated by the notebook. Roughly, O'Brien complains that, if we take the notebook as a good example of external information storage, and if we take connectionist networks as our best model for inner (neural) psychological information storage, then we are forced to conclude that internal and external elements store and organise information in very different ways, since the linguistic inscriptions in the notebook are akin to a classical database and so will fail to ground capacities such as fluid generalization, default reasoning and graceful degradation that we associate with connectionist forms of representation. As a result, functional equivalence (parity) fails, and, along with it, the first-wave case for ExC.

Clark could have responded to O'Brien's objection by pointing out that it trades on a misreading of the parity principle. As mentioned earlier, according to that principle, we should start with the distributed, notebook-deploying system, and then imagine a functionally equivalent element to the notebook operating inside the head of an agent, as part of a wholly inner system. As far as the parity principle is concerned, then, the

requirement is not that the mode of information storage under the spotlight should replicate the functionality of any extant inner system (such as, using O'Brien's chosen model, that of a connectionist network), but rather that it should replicate the functionality of the external notebook system (plausibly that of a classical database). Indeed, Clark (2008, 114) explains that:

The parity probe was... meant to act as a kind of veil of metabolic ignorance, inviting us to ask what our attitude would be *if currently external means of storage and transformation were, contrary to the presumed facts, found in biology* [my emphasis]. Thus understood, parity is not about the outer performing just like the (human-specific) inner.

But this technical knock-out against O'Brien would only have opened the door to a new challenge, one in which the critic would no doubt have pressed the point that the fan of first-wave ExC still owes us an account of why we should treat the notebook-equivalent inner element as having cognitive status, that is, she owes us an account of the basis for parity judgments. Simply pointing to the fact that the element in question is now inside the head would of course be a profoundly dubious response to an entirely reasonable request. For one thing, it would imply that the fan of first-wave ExC understands being inside the head to be sufficient, although not necessary, for cognitive status, and while that's not an incoherent position to adopt, it certainly jars with the idea that parity is 'about equality of opportunity: avoiding a rush to judgment based on spatial location alone' (Clark 2008, 114). More damagingly, perhaps, if simply being inside the head is sufficient for an element to have cognitive status, then one could presumably apply the parity principle to any old external element that shapes adaptive behaviour, no matter what its functional contribution to action or its degree of integration with the rest of the system, in order to deliver the conclusion that it has cognitive status. And that looks like a recipe for what those of us in the ExC business call 'cognitive bloat', a to-be-avoided outcome in which one is forced to concede cases of extended cognition that are wildly counter-intuitive.

Clark himself suggests that the basis for parity judgments is the intuitive adjudication of common sense. As he puts it, 'the isomorphism [what I am calling 'functional equivalence'] is said to hold only in respect of the explanatory role of the external elements in a commonsense account of the agent's behaviour' (Clark 1998, 98). And, in a later treatment: the 'Parity Principle was meant to engage our rough sense of what we might intuitively judge to belong to the domain of cognition – rather than, say, that of digestion – but to do so without the pervasive distractions of skin and skull' (Clark 2008, 114). The idea, then, is that the inscriptions in the target notebook contribute functionally to the kinds of broad patterns of actual and counterfactual behaviour (navigating to, or being disposed to navigate to, particular locations, and so on) that everyday commonsense psychology locks onto as cognition-laden. That's why the information stored in the inscriptions forms the contents of extended dispositional beliefs (believing where things are, and so on), and why the notebook counts as part of the relevant mental machinery. In effect, then, O'Brien appeals to cognitive science to ground parity (to fix the conditions for functional equivalence), and satisfies himself that he has an objection to ExC, whereas Clark appeals to folk psychology to do the same job, and satisfies himself that he has an argument for, or at least has made adequate conceptual space for, ExC.

So, does Clark's strategy work? I think we should be sceptical, on the grounds that our contemporary pre-theoretical understanding of the domain of the psychological plausibly includes a presumption of the within-the-skull-and-skin internality of cognition. This is a claim that requires more in the way of a defence than I can give it here (for such a defence, see Wheeler 2011b). In lieu, however, here is a quick and dirty example that provides a partial motivation. Imagine that Darwin were alive today. If a radical creationist cell stole the notebook in which he sketched the tree of life and added his famous comment 'I think'², the folk might well be interested in that act, and either supportive of it or appalled by it. But whatever attitudes the folk may adopt here, what's driving their thinking is not an interest in the whereabouts of part of Darwin's cognitive machinery. A plausible explanation for why the folk's attitudes here aren't driven by such an interest, but rather by whatever views they happen to hold about creation and the idea of a last universal common ancestor, is that modern commonsense psychology takes Darwin's mind to be realized wholly inside Darwin's head, so, by appeal to the authority of commonsense, the whereabouts of the beyond-the-skin notebook cannot coincide with the whereabouts of part of Darwin's cognitive machinery. Putting the point another way, if Darwin's notebook provides a case of cognitive extension, then, pace Clark, we will need to call on something other than our intuitive understanding of the cognitive to explain why.

It's an intriguing fact about Clark's reply to O'Brien that, having lodged the idea that functional equivalence should be understood by way of commonsense psychology (an idea that, as we have seen, he pursues in later work), Clark promptly drops it, on the (as far as I can tell, unexplained) grounds that, in his view, it 'only gets us so far' (Clark, 1998, 99). He continues: '[a] better response to O'Brien's critique is... to see it as identifying a potential tension between two components of the extended mind story itself' (ibid., 99). Those two components are functional equivalence (i.e. parity) and complementarity, so the potential tension that Clark identifies is between arguments for ExC based on parity and those based on complementarity. It is at this moment, then, that Clark establishes, within his own work, the conceptual basis for the subsequently drawn distinction between first-wave and second-wave ExC. Moreover, and perhaps surprisingly, Clark explicitly gives priority to *second-wave* arguments. He writes:

...it is best to see functional isomorphism as at most part of a sufficient condition for cognitive extension, rather than as a necessary feature. The *more interesting and plausible argument* [my emphasis], I feel, is the one which describes the seepage of mind into the world by stressing that "the brain's brief is to provide complementary facilities that will support the repeated exploitation of operations upon the world [and] to provide computational processes (such as powerful pattern completion) that the world, even as manipulated by us, does not usually afford". (Embedded quotation from Clark 1997, 68)

For Clark, then, functional equivalence is not necessary for extended cognition. But now notice that an absence of functional equivalence, as it relates to this claim, cannot be

² See http://darwin-online.org.uk/EditorialIntroductions/vanWyhe_notebooks.html, last accessed 25 May 2017.

expressed as a failure of parity, *if* we adopt the correct understanding of the parity principle. After all, it is an assumption of the parity-based argument that the target distributed system with which we begin and the hypothetical wholly inner system that we imagine are functionally equivalent. The question that the parity principle helps us answer is not whether the two systems are functionally equivalent, but rather whether the two functionally equivalent systems are both cognitive or both non-cognitive. If the two systems are not functionally equivalent, one doesn't have a failure of parity; one simply hasn't set things up properly. So the only way to make sense of Clark's claim that functional equivalence is not necessary for extended cognition is in O'Brien's terms, that is, in terms of a failure of parity between (a) the functional contributions of certain external elements to some psychologically interesting behaviour and (b) the functional contributions of certain extant inner elements to broadly similar behaviour (e.g. an external classical database versus a neurally realized inner connectionist network).

With that thought in hand, one can ask the million-dollar question: is there a good argument for extended cognition that is driven not by functional equivalence, but by complementarity? My own view is that the answer is 'no'. To see why, we need to reflect on the fact that, according to the second-waver, it is precisely the differences between certain internal and certain external elements that explain how many cognitive tasks are performed. For example, following Bechtel (1994, 1996), one might explain how some examples of linguistic behaviour, natural deduction and mathematical reasoning are accomplished, by citing a complementary combination of externally located combinatorial symbol systems and internally located pattern-sensitive connectionist networks, where the latter do not themselves realize combinatorial structure. In such an arrangement, the capacity of connectionist networks to recognize, and to generalize from, patterns in training data, plus the temporal constraints that characterize real embodied engagements with strings of external symbols (e.g. different parts of the input will be available to the network at different times, due to the restrictions imposed by temporal processing windows) are harnessed to allow those networks to be appropriately sensitive to the structural properties of the external symbol systems. Adding in some integrationist-friendly features, one might even highlight the fact that the skilled embodied manipulation of the external symbols will proceed according to learned normative rules that, when mastered, transform what we can do.

As far as I can see, there is nothing to point to in the undoubtedly important phenomenon of complementarity, even when developed in terms of integration, that mandates ExC. Or at least there isn't, as long as one accepts (what I take to be) the pretty much unassailable thought that vehicle internalism (the view that the machinery of mind is always entirely inside the head) constitutes the default position in cognitive science (due to its status as the orthodox view in the field with a well-documented history of empirical success), and therefore that one should give up vehicle internalism only when it can no longer be reasonably maintained in the face of (perhaps some combination of) unaccounted-for data, differential explanatory power, or cogent philosophical argument. If that is indeed right, then the second-wave emphasis on the existence of theoretically significant differences between the internal and the external elements in question means that complementarity fails to deliver ExC, because a vehicle internalist reading of the brain-body-world systems in question, according to which the performance of inner cognitive mechanisms is causally

scaffolded by non-cognitive external elements, remains eminently available. As Rowlands (2010, 90) puts it at the culmination of a related line of reasoning, 'given that there are significant differences between internal cognitive processes and external processes involved in cognition, why not simply suppose that the latter are part of the extraneous scaffolding in which the real, internal cognitive processes are embedded?'.³ If one looks at the Bechtelian systems described above through this lens, then the internally located pattern-sensitive connectionist networks are the genuinely cognitive machinery, whereas the externally located combinatorial symbol systems are 'merely' performance-boosting, but non-cognitive, environmental props and scaffolds. This is precisely the reading to which, one assumes, O'Brien would be tempted, and that Bechtel himself seems to favour (see e.g. Bechtel 1994, 36; for discussion, see Wheeler 2011b).

The Bechtelian combination of inner connectionist networks and external symbol systems will figure again later in our discussion. For the moment, let's return to Clark's analysis of parity and complementarity. Clark's next move is to attempt to fuse these two elements into a single, integrated picture of extended cognition. He writes:

The argument for the extended mind thus turns primarily on the way disparate inner and outer components may co-operate so as to yield integrated larger systems capable of supporting various (often quite advanced) forms of adaptive success. The external factors and operations, in this model, are most unlikely to be computationally identical to the ones supported directly in the wetware – indeed, the power of the larger system depends very much on the new kinds of storage, retrieval and transformation made possible by the use of extra-neural resources... These new operations, however, may often be seen as performing kinds of tasks which, were they but done in the head, we would have no hesitation in labelling cognitive. This is because they contribute to behavioural success by for example storing and manipulating information, and by reconfiguring problem spaces. This kind of higher-level functional isomorphism is, I think, quite compatible with the idea... that there exist deep and important differences between e.g., active biological and passive symbolic modes of storage and retrieval.

There are really two different questions being addressed here. The first is: as phenomena, are functional equivalence and complementarity compatible with each other? The answer to this question is surely 'yes'. As Clark observes, inner and outer resources may be functionally different when judged against one standard (e.g. the precise computational operations performed), but functionally equivalent when judged against another (e.g. a general capacity for storing and manipulating information). Of course, given what we have learned about Clark's understanding of the basis for parity, the fact that he introduces the parity principle into this passage indicates that, for him, the standard for functional equivalence is ultimately set by the way in which the commonsense adjudications of folk

³ Although Rowlands' (1999) earlier work is rightly identified as one important source for integrationism, and thus as one wellspring of second-wave ExC, his more recent position (e.g. Rowlands 2010) has seen him argue that parity considerations (properly understood) and complementarity-integrationist thinking have equal weight in the justification for cognitive extension.

psychology patrol the cognitive/non-cognitive boundary. So the idea is that the category of the cognitive, as implicitly understood by the folk, is coarse-grained enough to paper over the fine-grained differences between inner and outer computational operations.

The second question in play in the target passage, perhaps in a more subterranean register, is this: does recognizing the peaceful co-existence of complementarity and Clark-style functional equivalence deliver extended cognition? Here, I think, the answer must be 'no'. As argued above, neither commonsense-driven parity nor an argument driven solely by complementarity can deliver extended cognition *on its own*: the internalist bias of contemporary folk psychology makes it an inappropriate source for an ExC-compliant standard of functional equivalence, and complementarity considerations fail to destabilize vehicle internalism. But might the two sets of considerations when combined be jointly sufficient for extended cognition? That doesn't seem likely. If we approach things from one direction, the recognition of commonsense-driven parity alongside complementarity would need to counteract the inability of the latter to destabilize vehicle internalism. But that thought is undermined by the vehicle-internalist bias of commonsense psychology. From the other direction, the recognition of complementarity alongside commonsense-driven parity would need to counteract the vehicle-internalist bias of commonsense psychology. But that thought is undermined by the result that complementarity is unable to destabilize vehicle internalism. In sum, neither Clark-style functional equivalence nor complementarity produces a compelling argument for extended cognition, and neither does their simple conjunction. Where do we go from here?

3. The Missing Ingredient

Here is a proposal: what a good argument for extended cognition needs is a suitable standard for functional equivalence, one that doesn't exhibit the kind of vehicle-internalist bias manifested by modern commonsense psychology, but which allows for the phenomenon of complementarity to exist at some appropriate level of analysis. Let's see if such a standard can be found.

The concept of the *mark of the cognitive* is most readily associated with those hardy critics of ExC, Adams and Aizawa (e.g. 2008). However, the notion itself is apt to contribute to pro-ExC as well as anti-ExC arguments. To see why, we need to bring out a distinction that is sometimes obscured in the literature. In the present vicinity, there are two ways of applying the notion of the mark of the cognitive. The first focusses on the claim that some theoretical account of what counts as cognitive is needed, in order to determine where in the physical world our mental machinery is located. The second focusses on what specific theoretical account of the cognitive one ought to adopt. Call the first of these the *slot-level* application of the idea (there's a slot that needs to be filled), and the second the *filler-level* application (what should we put in the slot?).

To bring this distinction into better view, consider first an example given by Adams and Aizawa themselves (Adams and Aizawa 2008, 86-7). Imagine someone sincerely suggesting that there are good reasons to believe that crystals are alive. Any genuine resolution of this issue would surely turn on a scientifically informed account of what it is for an entity to be alive. As one might put it, *what is needed is a mark of the living*. Notice that, so far, all we've

said is that there's a hole in our picture – a slot that needs to be filled, if we are to decide whether certain entities, namely crystals, are alive. We haven't yet said precisely what it is for an entity to be alive. Similarly, in the face of the claim that our cognitive machinery extends beyond the skin, one might say, along with Adams and Aizawa, that what is needed to decide this issue is a mark of the cognitive (there's a slot to be filled), without saying precisely what it is for a systemic element to be cognitive (without filling that slot), and thus without saying whether, in fact, ExC is true. At the slot level, then, and using my favoured formulation (Wheeler 2010a, b, 2011a, b), a mark of the cognitive is a scientifically informed account of what it is to be a proper part of a cognitive system that, so as not to beg any questions, is fundamentally neutral with respect to where any candidate element might happen to be spatially located.

The mark-of-the-cognitive slot is occupied once we have a scientifically informed, locationally neutral account of what it is for a systemic element to be cognitive in character. That's when a filler-level mark of the cognitive is in place. So, when one says, correctly, that, for Adams and Aizawa, the mark of the cognitive is twofold, in that, for them, the cognitive is individuated by the presence of (i) non-derived representations and (ii) the specific kinds of information processing mechanisms identified by human cognitive psychology, one is explicating their proposal for a filler-level mark of the cognitive. This is not the end of the working day, however. Further philosophical and scientific leg-work will still be required, in order to find out precisely 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. As Adams and Aizawa put it, when discussing their own proposal: 'We do not maintain that non-derived representations must be found in the head. That cognition involves non-derived representations is one empirical hypothesis; that non-derived representations are to be found in some particular regions of spacetime is another' (Adams and Aizawa 2008, 55). They conclude, however, that their filler-level mark of the cognitive ultimately licenses vehicle internalism (for critical discussion, see e.g. Clark 2008, 89-92).

As a slot-level structure, the concept of the mark of the cognitive has been endorsed by a number of pro-ExC theorists, including Rowlands (2010), Walter (2010) and Wheeler (2010a, b, 2011a, b), as making an indispensable contribution to the debate surrounding extended cognition. I am about to add extra noise to that clamour. Of course, this endorsement leaves open the question of whether a specific account of the cognitive can be found that, pace Adams and Aizawa, can fill that slot in an ExC-compliant way. I shall have something to say about that too.

It might seem that the mark of the cognitive (slot-level application) is just what we need, in order to plug the gaps that, if I am right, have opened up in Clark's argument for ExC. Firstly, the dependence on science and the demand for locational neutrality mean that the mark-of-the-cognitive route avoids the worry about the internalist bias of commonsense psychology. Secondly, if an external element bears a true mark of the cognitive, then any differences between the functional contribution of that element and the functional contribution of our extant neural resources (at some level of description) will provide no mandate for relegating that element to the status of non-cognitive external scaffolding. And

thirdly, a mark of the cognitive will supply a standard by which judgments of parity may be made. Game over.

Clark disagrees. He argues that any attempt to argue for ExC on the basis of a mark of the cognitive is 'doomed to failure [because] the shape of any such scientific theory of legitimate vehicles will surely be determined, in large part, by what we take as central examples of real-world realizers of cognitive processes in the first place' (Clark 2011, 452). The suggestion here is that the appeal to science in determining a mark of the cognitive will end up undermining any effective sense of locational neutrality. Thus say that cognitive science settles on a fundamentally unified account of cognition, in the strong sense that all psychological phenomena embody some deep principle of organization. Clark's worry is that this account will fail to deliver ExC, because it will be 'heavily biased towards its own origins' (ibid. 453). To illustrate this worry, Clark focusses on contemporary predictive processing models of cognition – models in which, very roughly indeed, the brain is conceived as a system of hierarchically organized subsystems, each of which predicts its inputs and then corrects for any errors in those predictions in order to minimize informational surprise. Clark notes that the predictive processing approach provides a unifying account of a vast range of psychological phenomena such as perception, learning, inference and attention (the list is expanding daily). At this point, Clark's worry comes to the fore, and it is, I think, best expressed as a dilemma: either (i) predictive processing is the (filler-level) mark of the cognitive, in which case the fact that it has been developed as an account specifically of neural activity will skew things towards vehicle internalism, or (ii) the fact that predictive processing has been developed as an account specifically of neural activity will mean that we need different intellectual resources in order to establish that external elements count as genuine parts of our cognitive machinery. In Clark's own terms, this is the difference between (a) '[regarding] the neural kernel, the common neural mechanisms for the progressive reduction of prediction error, as limning the space of the cognitive', or (b) 'allow[ing] that genuinely cognitive processes can also become hybridized, so that *their* effective mechanisms include not just the neural elements but span brain, body, and world' (ibid. 454). In sum, the result is either a bias towards vehicle internalism or a failure to answer the key question.⁴

So, is it true that a filler-level mark of the cognitive formed within the heartland of internalism must inevitably deliver us into the jaws of Clark's dilemma? I don't think so. Here I shall redeploy an example that I have used previously (Wheeler 2011a, 2014, 2015). Recall that Bechtel (1994, 1996) defends a view according to which a number of high-end cognitive achievements such as linguistic behaviour, natural deduction and mathematical reasoning are often the result of sensorimotor-mediated causal interactions between certain external representational systems (e.g. logical and mathematical notations) and internal connectionist networks, interactions in which atomic symbols are combined and manipulated according to the principles of an externally realized compositional syntax and semantics. As also mentioned earlier, Bechtel himself seems to hold that the genuinely

⁴ This is not to say, of course, that predictive processing is necessarily inconsistent with extended cognition, or that it couldn't figure in an argument for extended cognition, but only that predictive processing cannot figure in such an argument by providing a mark of the cognitive. For more on predictive processing and ExC, see e.g. (Clark 2017).

cognitive part of the proposed distributed solution here remains skin-side. But there is an alternative view available.

Newell and Simon (1976, 116) once claimed that a suitably organized ‘physical symbol system has the necessary and sufficient means for general intelligent action’. A physical symbol system (henceforth PSS) is (roughly) a material system in which atomic symbols are automatically combined and manipulated by structure-sensitive processes, according to the principles of a compositional syntax and semantics. Although Newell and Simon adopted what we might call an unrestricted form of this hypothesis (i.e., all cognition is the result of a suitably organized PSS), one might reasonably adopt a more restricted version. For example, let’s proceed – as many classical computational psychologists studying the human mind manifestly have – by holding that a suitably organized PSS has the sufficient means for certain high-end cognitive achievements. There’s no doubt that Newell and Simon intended their proposal to deliver an account of what the brain does – more precisely, to provide an account of the cognition-realizing computational processes that are implemented in the brain. By Clark’s reasoning, this ought to drive us into the jaws of the dilemma highlighted above. But, I submit, it doesn’t. Rather, Bechtel’s distributed architecture of an inner connectionist network coupled to an external symbol system qualifies as a PSS. Put more carefully, the different-but-complementary inner and outer elements (the inner pattern-sensitive connectionist processes and the outer symbols on which those processes are targeted) together comprise a hybrid, distributed realization of a PSS. So, given the theoretical resources of a slot-level appeal to the need for a mark of the cognitive, and a filler-level unpacking of that notion in terms of our restricted PSS hypothesis, the Bechtelian architecture in question emerges not as a hybrid arrangement of cognitive and non-cognitive elements, but as a distributed (over brain, body and world) system in which each element counts as cognitive. In other words, if being a PSS is a mark of the cognitive, then our Bechtelian architecture is not only an extended PSS, it is also an extended cognitive system.

Of course, more would need to be said to drive home this idea. For example, one would have to establish that some subset of, or some combination of, inner connectionist structures and non-neural bodily factors (where the latter might include the physical movements by which the external symbols in question are manipulated) could reasonably be interpreted as realizing the classical-style rules of a PSS (for discussion, see Wheeler 2011a). But let’s assume that any such concerns can be met. What we have brought into view is a mark of the cognitive that an area of established human cognitive psychology (classical computational psychology focussed ultimately on explaining what the brain does) takes to be theoretically important. And yet, contra Clark, that very approach seems to support ExC, in that, under the right circumstances, some of the systemic elements in question (the symbol structures) may be externally located.

To be clear, I am not claiming that the PSS hypothesis is true. But then I don’t need that claim for the point at issue. I have sought to demonstrate only that a candidate mark of the cognitive forged within the heartland of vehicle internalism is able to meet the locational neutrality requirement and thus potentially play a pivotal role in the debate over ExC. Put another way, and pace Clark, the fact that a particular mark of the cognitive has been

extracted from an account designed to explain inner (neural) activity does not necessarily skew things towards vehicle internalism or leave ExC with a justificatory deficit.

4. Snakes and Ladders

Where Clark sees a snake, I see a ladder. For Clark, appealing to a mark of the cognitive condemns the advocate of ExC to a relentless slide into a helpless oscillation between vehicle internalism and more work to be done. By contrast, I have argued that just such an appeal furnishes the advocate of ExC with a sturdy ladder that traverses the gaps left in those arguments for extended cognition that are based on parity, complementarity, or the combination of the two. But perhaps ‘traverse’ is the wrong word. Maybe ‘by-pass’ is better. For it is arguable that once we have access to a mark of the cognitive, we simply don’t need to concern ourselves with notions such as parity and complementarity. As Walter (2010, 295) puts it, during a discussion of the former, ‘[o]nce we have at hand a mark of the cognitive, then if some extended process has it, it is cognitive, and if not, then not, regardless of any parity reasoning’. Something similar might be said in relation to complementarity. Here, however, we need to recall the distinction between the arguments and the phenomena. As Clark’s treatment of the issues nicely indicates, there seems little doubt that functional equivalence and complementarity (co-existing when different standards for functional equivalence are utilized) are phenomena that will figure in extended cognitive systems, if any such systems exist, so these notions will continue to play an important part in our *understanding* of extended cognition, even if their roles in the *argument* for ExC have been called into question.

Given these reflections, perhaps a better description of how things stand is that the best argument for extended cognition, the one that relies centrally on the concept of a mark of the cognitive, does not fit comfortably into either wave of the movement that was founded in and shaped by Clark’s seminal work. That said, and given that the mark-of-the-cognitive argument for extended cognition might reasonably be interpreted as a call for systems with various material instantiations to receive equal treatment when judged against a spatially unbiased and theoretically motivated standard of what counts as cognitive, that argument is perhaps most naturally seen as a development of first-wave (parity-based), rather than of second-wave (complementarity-based), thinking. Issues of pigeonholing aside, however, the key conclusion of our investigation into parity and complementarity remains this: in spite of Clark’s arguments to the contrary, the hypothesis of extended cognition needs a mark of the cognitive.

Acknowledgments

The argument of this chapter was improved thanks to some very useful editorial comments on an earlier draft. A few passages of text were adapted with revision from (Wheeler 2013, 2014).

References

Adams, F. and Aizawa, K. (2008) *The bounds of cognition*. Malden, MA and Oxford: Blackwell.

- Bechtel, W. (1994) Natural deduction in connectionist systems. *Synthese*, 101, 433-463.
- Bechtel, W. (1996) What knowledge must be in the head in order to acquire language. In B. Velichkovsky and D. M. Rumbaugh (Eds.), *Communicating meaning: the evolution and development of language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clark, A. (1997) *Being there: putting brain, body, and world together again*. Cambridge, Mass.: MIT Press.
- Clark, A. (1998) Author's response. Review symposium on Andy Clark's *Being there*. *Metascience*, 7, 95-103.
- Clark, A. (2008) *Supersizing the mind: embodiment, action, and cognitive extension*. New York: Oxford University Press.
- Clark, A. (2011) Finding the mind. *Philosophical Studies*, symposium on Andy Clark's *Supersizing the mind*, 152(3), 447-461.
- Clark, A. (2017) How to knit Your own Markov blanket: resisting the second law with metamorphic minds. In T. Metzinger and W. Wiese (Eds.), *Philosophy and predictive processing*, <http://predictive-mind.net/papers>, last accessed 26 May 2017.
- Clark, A. and D. Chalmers, D. (1998) The extended mind. *Analysis*, 58(1): 7-19.
- Fritzman, J. M. and Thornburgh, K. (2016) "I is someone else": constituting the extended mind's fourth wave, with Hegel. *Essays in Philosophy*, 17(2), 156-90.
- Kirchhoff, M. D. (2012) Extended cognition and fixed properties: steps to a third-wave version of extended cognition. *Phenomenology and the Cognitive Sciences*, 11(2), 287-308.
- Kiverstein, J. and M. Farina. (2011) Embraining culture: leaky minds and spongy brains. *Teorema*, 32(2), 35-53.
- Menary, R., ed. (2010) *The extended mind*. Cambridge, MA: MIT Press.
- Menary, R. (2007) *Cognitive integration: mind and cognition unbounded*. Basingstoke: Palgrave Macmillan.
- Newell, A. and Simon, H. A. (1976) Computer science as empirical inquiry: symbols and search. *Communications of the Association for Computing Machinery*, 19(3), 113-126.
- O'Brien, G. (1998) Review symposium on Andy Clark's *Being there*. *Metascience*, 7, 78-83.
- Rowlands, M. (1999) *The body in mind*. Cambridge: Cambridge University Press.

- Rowlands, M. (2010) *The new science of the mind: from extended mind to embodied phenomenology*. Cambridge, MA: MIT Press.
- Sutton, J. (2010) Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (Ed.) *The extended mind* (pp.189-225). Cambridge, Mass: MIT Press.
- Walter, S. (2010) Cognitive extension: the parity argument, functionalism, and the mark of the cognitive. *Synthese*, 177, 285-300.
- Wheeler, M. (2010a). In defense of extended functionalism. In R. Menary (Ed.) *The extended mind* (pp.245-70). Cambridge, Mass: MIT Press.
- Wheeler, M. (2010b) Minds, things, and materiality. In L. Malafouris and C. Renfrew (Eds.), *The cognitive life of things: recasting the boundaries of the mind* (pp. 29-37). Cambridge: McDonald Institute Monographs. Reprinted in J. Schulkin (Ed.) (2011), *Action, perception and the brain* (pp.147-63). Basingstoke: Palgrave-Macmillan.
- Wheeler, M. (2011a). Embodied cognition and the extended mind. In J. Garvey J (Ed.), *The Continuum companion to philosophy of mind* (pp.220-238). London: Continuum. Reprinted in paperback version published as *The Bloomsbury Companion to Philosophy of Mind* (pp.220-38), 2015.
- 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. (2013) Is cognition embedded or extended? The case of gestures. In Z. Radman (Ed.), *The hand, an organ of the mind: what the manual tells the mental* (pp.269-301). Cambridge, Mass.: MIT Press.
- Wheeler, M. (2014). Revolution, reform, or business as usual? The future prospects for embodied cognition. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp.374-83). Abingdon and New York: Routledge.
- Wheeler, M. (2015) A tale of two dilemmas: cognitive kinds and the extended mind, in C. Kendig (Ed.), *Natural kinds and classification in scientific practice* (pp.175-85). London: Routledge.