Draft of paper to appear in Garvey, J. (ed.), Companion to the Philosophy of Mind, Continuum.

Embodied Cognition and the Extended Mind

Michael Wheeler

1. The Flight from Cartesianism

There is a seductive image of intelligent action that sometimes gets labelled Cartesian. According to this image, as I shall present it here, the psychological understanding of the operating principles by which an agent's mind contributes to the generation of reliable and flexible, perceptually guided intelligent action remains conceptually and theoretically independent of the details of that agent's physical embodiment. Less formally, one might say that, in the Cartesian image, the body enjoys no more than a walk-on part in the drama of intelligent action. Whether or not the Cartesian image is Cartesian in the sense that it ought to be attributed to Descartes himself is a matter that demands careful exegetical investigation (see e.g. Wheeler 2005 for an analysis which concludes that, by and large, it should). In general, positions that are currently identified as Cartesian may not map directly or completely onto Descartes' own views. This potential mis-match is an example of a widespread phenomenon and should come as no surprise. Were Karl Marx with us today, he might well express serious misgivings about some of what has been said and done in the name of Marxism. In Descartes' case, his views have been handed down to us via a rich intellectual history of contested interpretations and critical debate. Inevitably, perhaps, some ideas that now bear the stamp Cartesian will have as much to do with that intervening process as they have to do with Descartes himself. Anyway, for now, I intend to ignore the question of provenance. What is crucial in the present context is that the two views of intelligent action with which I shall be concerned in this chapter – the hypotheses of embodied cognition and of the extended mind – may be understood as different stop-off points in a flight from the image in question.

To bring all this into better view, we can adapt an example due to Clark (1997, pp.63-4) of some different ways in which an intelligent agent might solve a jigsaw puzzle. Here is a strategy suggested by the Cartesian image. On the basis of perceptual information about the problem environment (the unmade jigsaw), the agent solves the entire puzzle 'in her head', using some combination of mental imagery, judgment, inference, reasoning, and so on. The solution arrived at in this way is then executed in the world, through a series of movement instructions that are dispatched from the mind, to the hands and arms. Things may not always go according to plan, of course, but any failures experienced during the execution phase act as nothing more than perceptual prompts for some newly initiated in-the-head planning. Now, it is quite obvious that the puzzle-solving mind at the core of this activity needs a body to execute the movement instructions generated by that mind; and nothing in the account on offer suggests that there could be minds without brains. (Substance dualism is not the issue.) Nevertheless, in this Cartesian scenario, the fact is that the body makes only an impoverished contribution to the intelligence on display. The nature of this impoverishment becomes clear once a second vision of jigsaw competence is placed on the table. According to this new vision, certain bodily acts – such as picking up various pieces, rotating those pieces to help pattern-match for possible fits, and trying out potential candidates in the target position – are deployed as central aspects of the agent's problem-solving strategy. In the unfolding of this alternative plot, the details of the thinker's embodiment, in the guise of the specific embodied manipulative capacities that she deploys, plays an essential supporting role in the story of intelligent action. This is an example of embodied cognition.[1]

Notice that problem-solving strategies which essentially involve bodily acts will often encompass a richer mode of environmental interaction than is present in Cartesian contexts. Thus in our Cartesian jigsaw-completing scenario, the physical environment is arguably no more than a furnisher of problems for the agent to solve, a source of informational inputs to the mind (via sensing), and a stage on which sequences of pre-specified actions, choreographed in advance by prior neural processes, are simply executed. By contrast, in the alternative, embodied cognition scenario, although the physical environment remains a furnisher of problems and a source of informational inputs, it has also been transformed into a readily available external resource which is exploited by the agent, in an ongoing way, to restructure the piece-finding problem and thus reduce the information processing load being placed on the inner mechanisms involved. Indeed, the external factors in play – in particular, the geometric properties of the pieces themselves – participate in a kind of ongoing goal-achieving dialogue with the agent's neural processes and her bodily movements. In so doing, those external factors account for some of the distinctive adaptive richness and flexibility of the problem-solving behaviour. The embodied mind is thus also a mind that is intimately embedded in its environment.

Once one starts to glimpse the kind of environmental contribution to intelligent action ushered in by embodied solutions, it is but a small step, although one which is philosophically controversial, to the second of our target positions, namely the extended mind hypothesis (Clark and Chalmers 1998).[2] According to this hypothesis, there are actual (in this world) cases of intelligent action in which thinking and thoughts (more precisely, the material vehicles that realize thinking and thoughts) are spatially distributed over brain, body and world, in such a way that the external (beyond-the-skin) factors concerned are rightly accorded cognitive status. In other words, 'actions and loops through nonbiological structure [sometimes count] as genuine aspects of extended cognitive processes' (Clark 2008b, p.85). So, if the extended mind hypothesis is true, it is not merely the case that thinking is sometimes (and perhaps sometimes essentially) causally dependent in complex and intricate ways on the bodily exploitation of external props or scaffolds. Indeed, bare causal dependence of thought on external factors is not sufficient for genuine cognitive extension (a point rightly emphasized by Adams and Aizawa 2008). Rather, if the extended mind hypothesis is true, thought must sometimes exhibit a constitutive dependence on external factors. This is the sort of dependence indicated by talk of beyond-the-skin factors rightly being accorded cognitive status. Stretching our thesbian metaphor beyond reasonable limits, this is the twist in the tale of intelligent action where the scenery and the props get a mention in the cast list.

In one short chapter, I cannot hope to give a comprehensive field guide to embodied cognition and the extended mind. So my goal will be more modest. I shall endeavour to cast light on a specific issue which lies at the very heart of the contemporary debate, namely the character of, and the argument for, the transition from embodied cognition to cognitive extension (see also, e.g., Clark 2008a, b; Wheeler 2010, forthcoming a; Rowlands forthcoming). Here, then, is where I am going. In section 2, I shall present some empirical research from cognitive science which illuminates the embodied cognition hypothesis, henceforth EmbC. In section 3, I shall suggest that once one has accepted the resulting picture of intelligent action, there remains a philosophical choice to be made over how to conceptualize the role of the body in the action-generation process, a choice between what Clark (2008a) identifies as a radical body-centrism and a newly interpreted functionalism. In section 4, I shall explore the connection between the second of these options and the extended mind hypothesis, henceforth ExM. My suggestion will be that the basic character of one of the central philosophical arguments for ExM, the argument from parity, makes that functionalist option more attractive. In section 5 I shall seek to strengthen the emerging picture by showing how a key element of the argument from parity may be secured.

2. Body Matters

As I shall use the term, orthodox cognitive science encompasses the bulk of research in both classical cognitive science (according to which, roughly, the mind recapitulates the abstract structure of human language, in that it is characterized by a combinatorial syntax and semantics) and mainstream connectionism (according to which, roughly, the mind recapitulates the abstract structure of the biological brain, in that it is organized as a distributed network of interconnected simple processing units). Although I shall not give a full defense of the claim here, it is arguable (see e.g. Wheeler 2005) that the Cartesian image of an explanatorily disembodied and disembedded mind has been a core feature of orthodox cognitive science and of the sort of scientifically oriented philosophy of mind that rides shotgun with that science.

This is not to say that no orthodox cognitive scientist has ever expressed the view that bodily acts in close interaction with environmental structures might play a crucial and active part in generating complex behaviour. Simon famously discussed the path followed by an ant walking on a beach in order to make precisely this point (Simon 1969; for discussion, see Boden 2006, pp.429-30, and Haugeland 1995/1998, pp.209-11). Moreover, the conceptual geography in this vicinity demands careful mapping. For one thing, orthodox connectionism takes its basic inspiration from a psychologically crucial part of the organic body, namely the brain. Indeed, the much recorded ability of orthodox connectionist networks to perform cognitively suggestive feats of graceful degradation, flexible generalization, fluid default reasoning, and so on, can, in many ways, be identified as a natural consequence of that nod to embodiment. So the claim that the disembodied aspect of the Cartesian image has been at work in this area of orthodox cognitive science needs to be backed by some sort of evidence (more on that soon). In addition, as we shall see later, the language-like compositional structures of the classical framework and the distributed network-style structures of connectionism may be rendered fully compatible with ExM, so it is not as if those structures must necessarily be associated with the Cartesian image. Nevertheless, it remains true, I think, that the Cartesian image has historically held sway as part of the received orthodoxy in cognitive science.

All that said, things are on the move. Over the past two decades, cognitive-scientific models generated from the EmbC perspective have become increasingly common. And to the extent that such models provide illuminating, compelling and fruitful explanations of intelligent action, EmbC as a paradigm garners empirical support. It is in this context that it will serve our current purpose to make a brief visit to the sub-discipline of contemporary artificial intelligence known as situated robotics. Roboticists in this camp shun the classical cognitive-scientific reliance on detailed internal representations (although they don't necessarily shun all forms of representation). The case for this scepticism about representational control often turns on the thought that where the adaptive problem faced by an agent involves integrating perception and action in real time so as to generate fast and fluid behaviour, detailed representations are just too computationally expensive to build and maintain. So situated roboticists favour an alternative model of intelligent action in which the robot regularly senses its environment (rather than checks an internal world model) to guide its actions. It is this commitment that marks out a robot as situated (Brooks 1991). One of the key lessons from research in this area is that much of the richness and flexibility of intelligence is down not to centrally located processes of reasoning and inference, but rather to integrated suites of special-purpose adaptive couplings that combine neural mechanisms (or their robotic equivalent), non-neural bodily factors, and environmental elements, as 'equal partners' in a behaviour-generating strategy. Unsurprisingly, then, the field of situated robotics is a rich storehouse of examples of embodied cognition.

To illustrate just how explanatorily powerful the appeal to embodiment may be in cognitive science, consider the following challenge. Clark and Thornton (1997) claim that there are certain learning problems – so-called type-2 problems – where the target regularities are inherently relational in nature, and so are statistically invisible in the raw input data. Type-2 problems are thus to be contrasted with type-1 problems, which involve non-relational regularities that are visible in that data. According to Clark and Thornton, this leaves cognitive science with a serious difficulty, because empirical testing suggests that many of the most widely used, 'off-the-shelf' artificial intelligence learning algorithms (e.g. connectionist back-propagation and cascade-correlation, plus others such as ID3 and classifier systems) fail on type-2 problems, when the raw input data is presented. This fact would, of course, be no more than a nuisance for cognitive science if such learning problems were rare; but, if Clark and Thornton are right, type-2 problems are everywhere – in relatively simple behaviours (such as approaching small objects while avoiding large ones), and in complex domains (such as grammar acquisition). Clark and Thornton proceed to argue that the solution to this difficulty involves the internal presence of general computational strategies that systematically re-represent the raw input data so as to produce a non-relational target regularity. This output re-representation is then exploited by learning in place of the initial input coding. In effect, the process of re-representation renders the type-2 learning problem tractable by transforming it into a type-1 problem.

So where do embodiment and situated robotics come in? Scheier and Pfeifer (1998) demonstrate that a type-2 problem may be solved by a process in which a mobile agent uses autonomous bodily motion to actively structure input from its environment. Once again the strategy is to transform an intractable type-2 problem into a tractable type-1 problem, but this time there is no need for any computational inner re-representation mechanism. The test case is the type-2 problem presented by the task of avoiding small cylinders while staying close to large ones. Scheier and Pfeifer show that this problem may be solved by some relatively simple, evolved neural network robot controllers. Analysis demonstrated that most of these controllers had evolved a systematic circling behaviour which, by inducing cyclic regularities into the input data, turned a hostile type-2 climb into a type-1 walk in the park. In other words, adaptive success in a type-2 scenario (as initially encountered) was secured not by inner re-representation, but by an approach in which the agent, 'by exploiting its body and through the interaction with the environment ... can actually generate ... correlated data that has the property that it can be easily learned' (Scheier and Pfeifer 1998, p.32).

Scheier and Pfeifer's canny and frugal solution to Clark and Thornton's challenge shows how being an embodied agent (of a mobile kind) can yield dividends in the cognitive realm, and thus how a proper sensitivity to what we might call 'gross embodiment' has an impact on cognitive science. A different, but equally important, perspective on how embodiment may shape our understanding of cognition comes into view if we switch scale, and concentrate instead on the detailed corporeal design of biological systems. Once again, as we shall see, situated robotics provides an experimental context in which an appeal to embodiment may be developed and tested.

As the flip-side of its claim to biological plausibility, mainstream connectionism tends to promote a vision of biological brain processes as essentially a matter of electrical signals transmitted between simple processing units (neurons) via connections (synapses) conceived as roughly analogous to telephone wires. However, as Turing once remarked, '[i]n the nervous system chemical phenomena are at least as important as electrical' (Turing 1950, p.46). The factoring out of brain-based chemical dynamics by mainstream connectionist theorizing thus indicates another dimension along which the embodiment of cognition is sidelined by orthodox cognitive science. So what happens when such chemical dynamics are brought into view?

Reaction-diffusion (RD) systems are distributed chemical mechanisms involving constituents that are (a) transformed into each other by local chemical reactions and (b) spread out in space by diffusion. There is evidence that RD systems explain how some unicellular organisms (e.g. slime molds) manage to co-ordinate biosignalling between spatially distributed sensors and actuators (Yamada et al. 2007). By explaining this sort of co-ordination, RD systems may help to explain the kind of behaviour in some unicellular organisms that researchers in the field of artificial life often describe as minimally cognitive, behaviour such as distinguishing between different relevant environmental factors, adapting to environmental change, and organizing collective behaviour.[3] Many of the molecular pathways present in unicellular organisms have been conserved by evolution to play important roles in animal brains, so an understanding of the ways in which RD systems may generate minimally cognitive behaviour will plausibly help us to explain the mechanisms underlying higher-level natural cognition. Against this background, Dale and Husbands (2010) show that a simulated RD system (conceived as a one-dimensional ring of cells within which the concentration of two coupled chemicals changes according to differential equations governing within-cell reactions and between-cell diffusion) is capable of intervening between sensory input (from whiskers) and motor output (wheeled locomotion) to enable a situated robot to achieve the following minimally cognitive behaviours: (i) tracking a falling circle (thus demonstrating orientation); (ii) fixating on a circle as opposed to a diamond (thus demonstrating discrimination); (iii) switching from circle fixation behaviour to circle avoidance behaviour on the presentation of a particular stimulus (thus demonstrating memory). As Dale and Husbands (2010, p.17) put it, a range of robust minimally cognitive behaviours may be exhibited by a 'seemingly homogenous blob of chemicals', a revision to our understanding of how cognition works that is inspired by our taking seriously the details of biological corporeal design.

In this section I have highlighted two important examples of the way in which embodiment may have an impact on cognitive theory. In the next section I shall address a further question: in the light of the examples of corporeal impact to which I have drawn attention, how, in general terms, are we conceptualize the fundamental contribution of the body to cognitive phenomena?

3. Two Kinds of Embodiment

Clark (2008a) observes that there are two different, although often tangled, strands of thinking at work within contemporary accounts that stress embodiment. In the following passage, he unravels those strands for us.

> One.... depicts the body as intrinsically special, and the details of a creature's embodiment as a major and abiding constraint on the nature of its mind: a kind of new-wave body-centrism. The other depicts the body as just one element in a kind of equal-partners dance between brain, body and world, with the nature of the mind fixed by the overall balance thus achieved: a kind of extended functionalism (now with an even broader canvas for multiple realizability than ever before). (Clark 2008a, pp.56-57)

In order to see this division of ideas in its proper light, one needs to say what is meant by functionalism, as that thesis figures in the debate with which we are concerned here. The final emphasis is important, because although Clark does not address the issue, the kind of functionalism plausibly at work in the transition from EmbC to ExM is not the kind most usually discussed by philosophers, although I think it is the kind most usually assumed in cognitive psychology. To bring our target version of functionalism into view, we can exploit McDowell's (1994) distinction between <u>personal-level</u> explanations, which are those concerned with the identification and clarification of the constitutive character of agency (roughly, what it is to competently inhabit a world), and <u>subpersonal</u> explanations, which are those concerned with mapping out the states and mechanisms (the parts of agents, as it were) that causally enable personal-level phenomena. Functionalism, as I shall understand it here, is a subpersonal causal-enabling theory. It is not, as it is in its more common philosophical form, a way of specifying constitutive criteria for what it is to undergo types of personal-level mental states. Depending on one's account of the relationship between personal and subpersonal levels of explanation, one might be a subpersonal functionalist while rejecting functionalism at the personal level. In this paper I shall say nothing more about personal-level functionalism. My concern is with the subpersonal version of the view, i.e., with the claim that what matters when one is endeavouring to identify the specific contribution of a subpersonal state or process *qua cognitive* is not the material constitution of that state or process, but rather the functional role which it plays in the generation of personal-level cognitive phenomena by intervening between systemic inputs, systemic outputs and other functionally identified, intrasystemic, subpersonal states and processes.

With that clarification in place, let's return to the division of ideas recommended by Clark. In the present context, it will prove useful to re-draw that division in terms of a closely related distinction between two kinds of materiality, namely *vital materiality* and *implementational materiality* (Wheeler 2010). The claim that the materiality of the body is vital is tantamount to the first strand of embodied thought identified by Clark, i.e., that the body makes a special, non-substitutable contribution to cognition, generating what, elsewhere, Clark (2008a, p.50) calls 'total implementation sensitivity'. On the other hand, if the materiality of the body is 'merely' implementational in character, then the physical body is relevant 'only' as an explanation of how mental states and processes are instantiated in the material world. The link between implementational materiality and functionalism becomes clear when one notes that, on any form of functionalism, including the subpersonal one presently on the table, *multiple realizability* will be at least an in-principle property of the target states and processes. Because a function is something that enjoys a particular kind of independence from its implementing material substrate, a function must, in principle, be multiply realizable, even if, in this world, only one kind of material realization happens to exist for that function. And since the multiple realizability of the mental requires that a single type of mental state or process may enjoy a range of different material instantiations, the specific material embodiment of a particular instantiation cannot be a major and abiding constraint on the nature of mind. Put another way, the implementational materiality of the mental (or something akin to it) is plausibly necessary for mental states and processes to be multiply realizable. And this remains true when one's functionalism – and thus the level at which the behaviour-generating causal states and processes *qua cognitive* are specified – is pitched at a subpersonal level. By contrast, where the materiality of the body is vital, multiple realizability is, if not ruled out altogether, at least severely curtailed (see e.g. Shapiro 2004, especially p.167).

Armed with the conceptual distinction just made, how are we to conceptualize the role of the body in each of our two flagship examples of embodied cognition – as a case of vital materiality (supporting a new wave body-centrism) or as a case of implementational materiality (supporting a functionalist picture)? My immediate answer to this question might come as something of a surprise. For, as far as I can see, each of our examples might be interpreted according to either vision of embodiment. Here's why.

To see Scheier and Pfeifer's cylinder discriminating robots as an instance of vital materiality, one might begin with the observation that Clark and Thornton's appeal to an inner process of re-representation exemplifies a computational information processing approach to solving the problem. One might then suggest, with some plausibility it seems, that the way in which Scheier and Pfeifer's robots exploit gross bodily movement in their specific circling behaviour provides us with a radical alternative to computational information processing as a general problem-solving strategy, an alternative available only to agents with bodies of a certain kind. To see Dale and Husbands' minimally cognitive RD system as an instance of vital materiality, one might interpret that system as an example of what Collins calls embrained knowledge. For Collins, knowledge is embrained just when 'cognitive abilities have to do with the physical setup of the brain,' where the term 'physical setup' signals not merely the 'way neurons are interconnected', but also factors to do with 'the brain as a piece of chemistry or a collection of solid shapes' (Collins 2000, p.182). Embrained knowledge so defined is an example of total implementation sensitivity and thus establishes vital materiality. And the evidence from Dale and Husbands that the spatio-temporal chemical dynamics of RD systems, as plausibly conserved in animal brains, may generate minimally cognitive behaviour surely provides an example of cognitive abilities being to do with the physical setup of the brain, that is, of embrained knowledge.

Now let's look at things from a different angle. To see Scheier and Pfeifer's robots as providing an instance of implementational materiality, one might argue that the restructuring of the learning problem achieved by their bodily movements is functionally equivalent to the restructuring of that problem effected by Clark and Thornton's inner re-representation strategy. In both cases, a type-2 learning problem (intractable to standard learning algorithms as it stands) is transformed into a type-1 problem (and so rendered tractable). Thus one might think in terms of alternative material realizations of a single multiply realizable, functionally specified contribution (the transformation of the statistical structure of the target information), a contribution that may be performed by inner neural mechanisms or by bodily movements. To see Dale and Husbands' RD system as an instance of implementational materiality, one need note only that the experiments described briefly above are designed explicitly as (something close to) replications, using an RD system, of experiments in minimally cognitive behaviour carried out originally by Beer

(1996, 2003; Slocum et al. 2000) using continuous recurrent neural networks (CNNs). RD systems thus emerge as one kind of vehicle for functionally specified mechanisms of orientation, discrimination and memory, mechanisms that could in principle be realized in other ways, such as by CNNs.

One might worry here that RD systems and CNNs are not alternative realizations of certain functionally specified mechanisms, but rather alternative ways of achieving certain minimally cognitive behaviours without there being any more specific functional unity in terms of processing architecture. And indeed, one might well analyze RD systems as examples of Collins' embrained knowledge, and thus of vital materiality (see above), while analyzing CNNs as a dynamically richer form of connectionism, and thus as a kind of microfunctionalist theorizing (Clark 1989) that demands an implementational notion of materiality.[4] But any such uncertainty in how to interpret the case is arguably grist to my mill, since it will be an illustration of the very issue of under-determination that I have set out to highlight.

As things stand, we seem to confront something of an impasse in our attempt to understand the fundamental contribution of embodiment to cognitive theory. To escape from this impasse, it seems to me, we have no option right now but to look beyond the thought that the understanding we seek may be directly read off from the available science. In the next section I shall present, analyze and briefly defend one of the central philosophical arguments for ExM, namely the argument from parity. I shall then explain why that argument forges a link with the functionalist perspective on embodiment. Given that vital materiality is inconsistent with functionalism, this suggests a consideration in favour of the view that the fundamental contribution of the body to cognitive theory is a matter of implementational materiality. At the very least, if the argument from parity is indeed sound, then the implementational view of embodiment is correct.

4. From the Parity Principle to Extended Functionalism

According to ExM, there are actual (in this world) cases of intelligent action in which thinking and thoughts (more precisely, the material vehicles that realize thinking and thoughts) are spatially distributed over brain, body and world, in such a way that the external (beyond-the-skin) factors concerned are rightly accorded cognitive status. To see how one might argue philosophically for this view, we need to make contact with what, in the ExM literature, is called the parity principle. Here is how that principle is formulated by Clark (2008b, p.77, drawing on Clark and Chalmers, 1998, p.8):

> If, 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.

The general idea here seems clear enough: if there is functional equality with respect to governing intelligent behaviour (for example, in the way stored information is poised to guide such behaviour), between the causal contribution of certain internal elements and the causal contribution of certain external elements, and if the internal elements concerned qualify as the proper parts of a cognitive system (state, process, mechanism, architecture…), then there is no good reason to deny equivalent status to the relevant external elements. Parity of causal contribution mandates parity of status with respect to the cognitive. But if the general idea of the parity principle is clear enough, the details of how to apply it are not, so we need to pause here to get clear about those details (for a similar analysis, see Wheeler forthcoming b).

One interpretation of the parity principle is suggested by the way in which it is applied by Clark and Chalmers themselves to the near-legendary (in ExM circles) case of Inga and Otto (Clark and Chalmers 1998). In this imaginary scenario, Inga is a psychologically normal individual who has committed to her purely organic (neural) memory the address of the New York Museum of Modern Art (MOMA). If someone asks her the location of MOMA, she deploys that memory to retrieve the information that the building is on 53rd Street. Otto, on the other hand, suffers from a mild form of Alzheimer's, but compensates for this by recording salient facts in a notebook that he carries with him constantly. If someone asks him the way to MOMA, he automatically and unhesitatingly pulls out the notebook and, without a hint of any critical scrutiny of the information stored within, retrieves the relevant fact, viz. that the museum is on 53rd Street. Clark and Chalmers claim that there is a functional equivalence between (i) the behaviour-governing causal role played by Otto's notebook, and (ii) the behaviour-governing causal role played by the part of Inga's brain that stores the same item of information as part of her purely organic memory. By the parity principle, then, Otto's memory turns out to be extended into the environment. Moreover, argue Clark and Chalmers, just as, prior to recalling the information in question, Inga has the non-occurrent dispositional belief that MOMA is on

53rd Street, so too does Otto, although while Inga's belief is realized in her head, Otto's is realized in the extended, notebook-including system.

If we reflect on precisely how the parity principle is intended to work in this particular case, we would be forgiven for thinking that the benchmark for parity (the set of conditions that the Otto-plus-notebook system would need to meet in order to count as cognitive) is fixed by whatever Inga's brain does. But although Clark and Chalmers' text sometimes leaves rather too much room for this reading of the parity principle, it would be a tactical disaster for the advocates of ExM if that really were what was meant. As Menary (a fan of ExM, but not of the parity principle), drawing on work by Sutton (ditto), observes:

> [o]nly at the grossest level of functional description can [the claim of equivalence] be said to be true. Otto and his notebook do not really function in the same kind of way that Inga does when she has immediate recall from biological memory. There are genuine and important differences in the way that memories are stored internally and externally and these differences matter to how the memories are processed. John Sutton has pointed out that biological memories stored in neural [i.e., connectionist] networks are open to effects such as blending and interference (see Sutton [2006] for discussion). The vehicles in Otto's notebook, by contrast, are static and do no work in their dispositional form (Sutton [2006]). (Menary 2007, p.59)

Other critics of the parity principle have appealed to the psychological data on various extant inner cognitive capacities, as delivered by cognitive science, in order to construct similar failure-of-parity arguments (see e.g. Adams and Aizawa 2008 on primacy and recency effects in organic memory; for discussion, see e.g. Wheeler 2010, forthcoming a, b). The general version of the worry, however, is this: if (i) the relatively fine-grained functional profiles of extant inner cognitive systems set the benchmark for parity, then (ii) any distributed (over brain, body and world) systems that we might consider as candidates for extended counterparts of those cognitive systems will standardly fail to exhibit full functional equivalence, so (iii) parity will routinely fail, taking with it the parity argument for cognitive extension.

Right now things might look a little bleak for a parity-driven ExM, but perhaps we have been moving too quickly. Indeed, it seems to me that the kind of anti-parity argument that we have been considering trades on what is in fact a misunderstanding of the parity principle. To see this, one needs to think more carefully about precisely what the parity principle, as stated above, asks us to do. It encourages us to ask ourselves whether a part of the world is functioning 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. So we are encouraged to imagine that exactly the same functional states and processes which are realized in the actual world by certain externally located physical elements are in fact realized by certain internally located physical elements. Having done this, if we then judge that the now-internal but previously external processes count as part of a genuinely cognitive system, we must conclude that they did so in the extended case too. After all, by hypothesis, nothing about the functional contribution of those processes to intelligent behaviour has changed. All that has been varied is their spatial location. And if the critic were to claim that that being shifted inside the head is alone sufficient to result in a transformation in the status of the external elements in question, from noncognitive to cognitive, he would, it seems, be guilty of begging the question against ExM.

To apply this understanding of the parity principle to the case of Otto and Inga, one must start with the functional contribution of Otto's notebook in supporting his behaviour, and ask whether, if that functional contribution were to be made by an inner element, we would count that contribution, and thus its realizer, as cognitive. If the answer is 'yes', then we have a case for ExM. Crucially, at no point in this reasoning have we appealed to Inga's organic memory (the relevant extant human inner) in order to determine what counts as cognitive. And while rather more would need to be said about the precise functional contribution of Otto's notebook, our reconceived argument from parity does not succumb to criticisms that turn on any lack of fine-grained functional equivalence between the target distributed system and some extant example of inner human cognition.

It is of course possible to conduct a debate that revolves around the functional contributions of certain elements, without that being an issue that concerns functionalism as such (cf. Chalmers 2008). So what is the link between the parity principle and functionalism? The parity principle is based on the thought that it is possible for the very same type-identified cognitive state or process to be available in two different generic formats – one non-extended and one extended. Thus, in principle at least, that state or process must be realizable in either a purely organic medium or in one that involves an integrated combination of organic and non-organic structures. In other words, it must be multiply realizable. So, if we are to argue for cognitive extension by way of parity considerations, the idea that cognitive states and processes are multiply realizable must make sense. As we have seen, functionalism provides one wellestablished platform for securing multiple realizability. Moreover, although functionalism has standardly been developed with respect to what is inside the head (e.g. the brain of some nonhuman entity may be wired up differently, or it may be silicon-based rather than carbon-based, without that affecting the rights of that entity to be judged a cognizer), there isn't really anything in the letter of functionalism as a generic philosophical outlook that requires such an internalist focus (Wheeler 2010, forthcoming a). According to (subpersonal) functionalism, when one is endeavouring to identify the specific contribution of a subpersonal state or process qua cognitive, it is not the material constitution of that state or process that matters, but rather the functional role which it plays in the generation of personal-level cognitive phenomena by intervening between systemic inputs, systemic outputs and other functionally identified, intrasystemic, subpersonal states and processes. There is nothing in this schema that requires multiple realizability to be a between-the-ears phenomenon. So functionalism allows, in principle, for the existence of cognitive systems whose boundaries are located partly outside the skin. It is in this way that we arrive at the position that, following Clark, I shall call extended functionalism (Clark 2008a, b; Wheeler 2010, forthcoming a).

We have seen already that there will be functional differences between extended cognitive systems (if such things exist) and purely inner cognitive systems. So, if extended functionalism and the parity principle are to fly together, what seems to be needed is some kind of theory that tells us which functional differences are relevant to judgments of parity and which aren't. To that end, here is a schema for a theory-loaded benchmark by which parity of causal contribution may be judged (Wheeler 2010, forthcoming a, b). First we give a scientifically informed account of what it is to be a proper part of a cognitive system that is fundamentally independent of where any candidate element happens to be spatially located. Then we look to see where cognition falls – in the brain, in the non-neural body, in the environment, or, as ExM predicts will sometimes be the case, in a system that extends across all of these aspects of the world. On this account, parity is conceived not as parity with the inner simpliciter, but rather as parity with the inner with respect to a scientifically informed, theory-loaded, locationally uncommitted account of the cognitive. So the parity principle now emerges not as the engine room of the extended mind, but as an heuristic mechanism that helps to ensure equal treatment for different spatially located systems judged against an unbiased and theoretically motivated standard of what counts as cognitive. It is a bulwark against what Clark (2008b, p.77) calls 'biochauvinistic prejudice'.

This idea of a scientifically informed, theory-loaded, locationally uncommitted account of the cognitive is tantamount to what Adams and Aizawa (e.g. 2008) call a mark of the cognitive. In the interests of expository elegance, I shall default to Adams and Aizawa's term. The most obvious next step in this dialectic would be for me to specify the – or, given the possibility that the phenomena in question will reward a disjunctive account, a – mark of the cognitive. In the next section I shall make a tentative proposal.[5]

5. A Mark of the Cognitive

Newell and Simon, two of the early architects of artificial intelligence, famously claimed that a suitably organized 'physical symbol system has the necessary and sufficient means for general intelligent action' (Newell and Simon 1976, p.116). As anyone familiar with cognitive science will tell you, a physical symbol system is (roughly) a classical computational system instantiated in the physical world, where a classical computational system is (roughly) a system in which atomic symbols are combined and manipulated by structure sensitive processes in accordance with a language-like combinatorial syntax and semantics. I shall take it that the phrase 'means for general intelligent action' points to a kind of cognitive processing. More specifically it signals the sort of cognitive processing that underlies 'the same scope of intelligence as we see in human action… in any real situation behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity' (Newell and Simon 1976, p.116). What we are concerned with, then, is a human-scope cognitive system. Notice that the concept of a human-scope cognitive system is not a species-chauvinistic notion. What matters is that the system exhibit roughly the same degree of adaptive flexibility we see in humans, not that it have our particular biological make-up, species ancestry or developmental enculturation.

Against this background, Newell and Simon's physical symbol systems hypothesis may be unpacked as the dual claims that (a) any human-scope cognitive system will be a physical symbol system, and (b) any physical symbol system of sufficient complexity may be organized so as to be a human-scope cognitive system. In effect, then, the hypothesis is equivalent to the claim that being a suitably organized physical symbol system is the mark of the (human-scope) cognitive. To unpack that claim: the physical symbol systems hypothesis advances a scientifically informed, theory-loaded account of the (human-scope) cognitive, one that supports a computational form of functionalist theorizing. But can it tick all our boxes by being a locationally independent account too? The answer, it seems, is yes. For while classical cognitive scientists in general thought of the symbol systems in question as being realized inside the head, there is nothing in the basic concept of a physical symbol system that rules out the possibility of extended material implementations. Indeed, as I shall now argue, the idea of an <u>extended physical symbol system</u> has much to recommend it.

In a series of compelling treatments that combine philosophical reflection with empirical modelling studies, Bechtel (1994, 1996; see also Bechtel and Abrahamsen 1991) develops and defends the view that certain human-scope cognitive achievements, such as mathematical reasoning, natural language processing and natural deduction, are the result of sensorimotor-mediated interactions between internal connectionist networks and external symbol systems, where the latter feature various forms of combinatorial syntax and semantics. It is useful to approach Bechtel's suggestion (as he does himself) by way of Fodor and Pylyshyn's (1988) well-known claim that connectionist theorizing about the mind is, at best, no more than a good explanation of how classical states and processes may be implemented in neural systems. Here is a brief reminder of Fodor and Pylyshyn's key argument. It begins with the empirical observation that thought is systematic. In other words, the ability to have some thoughts (e.g. that Elsie loves Murray) is intrinsically connected to the ability to have certain other thoughts (e.g. that Murray loves Elsie). If we have a classical vision of mind, the systematicity of thought is straightforwardly explained by the combinatorial syntax and semantics of the cognitive representational system. The intrinsic connectedness of the different thoughts in question results from the fact that the processing architecture contains a set of atomic symbols alongside certain syntactic rules for recombining those symbols into different molecular expressions. Now, Fodor and Pylyshyn argue that although there is a sense in which connectionist networks instantiate structured states (e.g. distributed connectionist representations have active units as parts), combinatorial structure is not an essential or a fundamental property of those states. This leaves connectionist networks inherently incapable of explaining the systematicity of thought, and thus of explaining thinking. What such systems might do, however, is explain how a classical computational architecture may be implemented in an organic brain.

Bechtel agrees with Fodor and Pylyshyn on two key points: first, that where systematicity is present, it is to be explained by combinatorially structured representations; and secondly, that connectionist networks fail to instantiate combinatorial structure as an essential property of their internal representational organization. He does not need to endorse Fodor and Pylyshyn's claim that all thought is systematic, however. For his purposes, all that is required is that some cognitive activities (e.g. linguistic behaviour, natural deduction, mathematical reasoning) exhibit systematicity.

Bechtel's distinctive next move is to locate the necessary combinatorial structure in systems of representations that remain external to the connectionist network itself. Given the idea that our inner psychology should be conceived in connectionist terms, this is tantamount to saying that the necessary combinatorial structure resides not in our internal processing engine, but rather in public systems of external representations (e.g. written or spoken language, mathematical notations). As Bechtel (1994, p.436) himself puts it, the 'property of systematicity, and the compositional syntax and semantics that underlie that property, might best be attributed to natural languages themselves but not to the mental mechanisms involved in language use'. (Notice that, for Bechtel, the mental is restricted to the inner. This is an issue to which we shall return.)

For this interactive solution to work, it must be possible for the natural sensitivity to statistical patterns that we find in orthodox connectionist networks to be deployed in such a way that some of those networks, when in interaction with specific external symbol systems, may come to respect the constraints of a compositional syntax, even though their own inner representations are not so structured. Bechtel's studies suggest that this may be achieved by exploiting factors such as the capacity of connectionist networks to recognize and generalize from patterns in bodies of training data (e.g. large numbers of correct derivations in sentential arguments), plus the temporal constraints that characterize real embodied engagements with stretches of external symbol structures (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). The conclusion is that 'by dividing the labor between external symbols which must conform to syntactical principles and a cognitive system which is sensitive to those constraints without itself employing syntactically structured representations, one can perhaps explain the systematicity… of cognitive performance' (Bechtel 1994, p.438).

How should we interpret the distributed solutions that Bechtel favours – as examples of embodied cognition or as instances of cognitive extension? Bechtel himself stops short of the extended option. Thus, as we have just seen, he tellingly describes systematicity as a feature of 'cognitive performance' rather than as a property of the cognitive system, and states that the compositional syntax and semantics 'might best be attributed to natural languages themselves but <u>not to the mental mechanisms involved in language use</u>' (my emphasis). What this indicates is that, for Bechtel, the genuinely cognitive part of the proposed solution remains skin-side. Let's see what interpretation we get, however, once we apply the parity principle. If the envisaged system of syntax-sensitive processes and combinatorially structured symbols were all stuffed inside the agent's head, we would, I think, have no hesitation in judging the symbol structures themselves to be bona fide parts of the agent's cognitive architecture. Equality of treatment therefore seems to demand that the external symbol structures that figure in the functionally equivalent distributed version of that solution also be granted cognitive status. On the strength of the parity principle, then, what we have here are models of extended cognition.[6]

Of course, the foregoing direct appeal to parity considerations takes us only part of the way toward ExM. As we have seen, parity-based arguments remain inconclusive until they receive backing from some mark of the cognitive that sets the benchmark for parity. It's at this point that we see the potential impact of the physical symbol systems hypothesis, as interpreted above. Let's accept, for the sake of argument, that being a sufficiently complex and suitably organized physical symbol system is at least <u>a</u> mark of the cognitive. I suggest that both the wholly inner and the environment-involving versions of the Bechtel-style network-plus-symbol-system architecture exhibit that mark, which means that both are cognitive systems, and the latter is an extended cognitive system. Given the functionalist character of the physical symbol systems hypothesis, such considerations strengthen further our reasons for thinking that the fundamental contribution of the body to cognitive theory is to be conceived in terms of implementational materiality, not vital materiality. One way to appreciate the plausibility of this picture is to reflect on the most obvious objection to it.

In response to the view just sketched, many cognitive scientists will want to complain that the kinds of pattern-matching and pattern-completion processes realized by connectionist networks are not equivalent to the syntactic rules present in classical systems, implying that the analysis of the Bechtel architectures as extended physical symbol systems is suspect. With all due respect this is, I think, a failure of the imagination. It is of course true that the network processes concerned are not explicitly rule-driven in a classical sense, but two considerations strongly indicate that this is not the end of the matter. First, the keystone of Bechtel's model is the thought that the networks involved are genuinely sensitive to the constraints of a compositional syntax. Thus, pending good arguments to the contrary, one might insist that Bechtel's networks implicitly realize the rules in question, at least in the minimal sense that, in this case (although not in others), classical-style rules will provide a perfectly reasonable, high-level, idealized description of the network's processing activity. (The fact that there is idealization here should not concern anyone. For one thing, idealization is part of scientific explanation. For another, as we have seen, orthodox connectionist models are themselves abstract idealizations of real brains.) Secondly, and from a more radical perspective, it may be that the classical rules are not implicitly realized in the neural network alone. If we think of those rules as principles that govern the skilled embodied manipulations of certain external material symbols, it might be more accurate to think in terms of dynamic subpersonal vehicles that include not just neurally-implemented connectionist elements, but also non-neural bodily factors, including physical movements. On either analysis of how the rules in question are realized, the objection under consideration would fail.

6. A Parting of the Ways

In exploring the relationship between embodiment and cognitive extension, I have presided over a parting of the ways between, on the one hand, ExM, understood as involving an extended functionalist commitment to a kind of open-ended multiple realizability, and, on the other, a particular strain of EmbC that depicts the organic body as, in some way, intrinsically special in the generation of cognitive phenomena. At root this fork in the theoretical road may be traced to a fundamental disagreement over how philosophy and cognitive science should conceive of the materiality of the body – as just one implementing substrate among possible others, or as a vital and irreplaceable determinant of cognitive life. I have presented a case for thinking that we should follow the ExM path to implementational materiality. But in this fast moving and complex debate, wrong turns and dead ends will abound. Under such circumstances, drawing up a road map will always be a hazardous task, and I expect there to be many moments of disorientation and puzzlement along the way, before we arrive at a detailed theory of embodied cognition and the extended mind.

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 Langauge. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bechtel, W. and Abrahamsen, A. (1991), Connectionism and the Mind: an Introduction to Parallel Processing in Networks. Oxford: Basil Blackwell.

Beer, R.D. (1996), 'Toward the evolution of dynamical neural networks for minimally cognitive behavior', in P. Maes, M. Mataric, J. Meyer, J. Pollack and S. Wilson (eds.), From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior, Cambridge, Mass.: MIT Press, pp. 421-429.

Beer, R.D. (2003), 'The dynamics of active categorical perception in an evolved model agent'. Adaptive Behavior. 11, (4), 209-243.

Boden, M. A. (2006), Mind As Machine: A History of Cognitive Science (two volumes). Oxford: Oxford University Press.

Brooks, R. A. (1991), 'Intelligence without reason', in Proceedings of 12th International Joint Conference on Artificial Intelligence, San Mateo, California: Morgan Kauffman, 569-95. Reprinted in Brooks' Cambrian Intelligence: the Early History of the New AI, chapt. 8, pp.133-186, Cambridge, Mass.: MIT Press.

Chalmers, D. (2008), Foreword to Andy Clark's Supersizing the Mind. In Clark 2008b, pp.ix-xvi.

Clark, A. (1989), Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing. Cambridge, Mass.: MIT Press.

Clark, A. (1997), Being There: Putting Brain, Body, And World Together Again, Cambridge, Mass.: MIT Press.

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. New York: Oxford University Press.

Clark, A. and D. Chalmers, D. (1998), 'The extended mind', Analysis, 58. (1). 7-19.

Clark, A. and Thornton, C. (1997), 'Trading spaces: computation, representation and the limits of uninformed learning', Behavioral and Brain Sciences, 20, 57-90.

Collins, H. (2000), 'Four kinds of knowledge, two (or maybe three) kinds of embodiment, and the question of artificial intelligence', in M. Wrathall and J. Malpas (eds.), Heidegger, Coping and Cognitive Science: Essays in Honor of Hubert L. Dreyfus, vol.2, Cambridge, Mass.: MIT Press, pp.179-195.

Dale, K. and Husbands, P. (2010), 'The Evolution of Reaction-Diffusion Controllers for Minimally Cognitive Agents', Artificial Life, 16.1: 1-19.

Fodor, J.A. and Pylyshyn, Z. (1988), 'Connectionism and cognitive architecture: a critical analysis'. *Cognition* 28, 3-71.

Haugeland, J. (1995/1998), 'Mind embodied and embedded', in Having Thought: Essays in the Metaphysics of Mind, chap. 9, pp. 207-237. Cambridge, Mass.: Harvard University Press.

Hurley, S. L. (1998), Consciousness in Action. Cambridge, Mass.: Harvard University Press.

Lakoff, G and Johnson, M (1980), Metaphors We Live By, Chicago: University of Chicago Press.

Lakoff, G and Johnson, M (1999), Philosophy in the flesh: the embodied mind and its challenge to western thought. New York: Basic Books.

McDowell, J. (1994), 'The content of perceptual experience'. The Philosophical Quarterly, 44, (175), 190-205.

Menary, R. (2007), Cognitive Integration: Mind and Cognition Unbounded. Basingstoke: Palgrave Macmillan.

Menary, R. (ed.) Forthcoming. The Extended Mind. Cambridge, Mass.: MIT Press.

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.

Noe, A (2004), Action in Perception, Cambridge, Mass.: MIT Press.

O'Regan, J.K. and Noe, A, (2001), 'A sensorimotor approach to vision and visual consciousness', Behavioral and Brain Sciences. 24, (5), 939–973.

Rowlands, M. (1999), The Body in Mind. Cambridge: Cambridge University Press.

Rowlands, M. (2003), Externalism: Putting Mind and World Back Together Again. Chesham, Bucks.: Acumen

Rowlands, M. Forthcoming. The New Science of the Mind: From Extended Mind to Embodied Phenomenology. Cambridge, Mass.: MIT Press.

Rupert, R. (2004), 'Challenges to the hypothesis of extended cognition', Journal of Philosophy 101, (8), 389-428.

Scheier, C. and Pfeifer, R. (1998), 'Exploiting embodiment for category learning', in R. Pfeifer, B. Blumberg, J.-A Meyer and S. W. Wilson eds., From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior. Cambridge, Mass.: MIT Press, pp.32-37.

Shapiro, L. (2004), The Mind Incarnate. Cambridge, Mass.: MIT Press.

Simon, H. (1969), The Sciences of the Artificial. Cambridge, Mass.: MIT Press.

Slocum, A.C., Downey, D.C. and Beer, R.D. (2000), 'Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention', in J. Meyer, A. Berthoz, D. Floreano, H. Roitblat and S. Wilson (eds.), From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior. Cambridge, Mass.: MIT Press, pp. 430-439.

Sutton, J. (2006), 'Distributed cognition: domains and dimensions'. Pragmatics and Cognition, 14, (2), 235-247.

Turing, A.M. (1950), 'Computing machinery and intelligence', *Mind* 59, 433-460. Reprinted in M. A. Boden (ed.), (1990), The Philosophy of Artificial Intelligence, Oxford University Press, Oxford, pp. 40-66. Page numbers in the text refer to the reprinted version.

Wheeler, M. (2005), Reconstructing The Cognitive World: The Next Step. Cambridge, Mass.: MIT Press.

Wheeler, M. (2010), 'Minds, things, and materiality', in L. Malafouris and C. Renfrew (eds.), The Cognitive Life of Things: Recasting the Boundaries of the Mind. Cambridge: McDonald Institute Monographs.

Wheeler, M. Forthcoming a, 'In defense of extended functionalism', to appear in Menary forthcoming.

Wheeler, M. Forthcoming b, 'In search of clarity about parity', to appear in Philosophical Studies, book symposium on Andy Clark's Supersizing the Mind (Clark 2008b).

Wilson, R. A. (2004), Boundaries of the Mind: the Individual in the Fragile Sciences. Cambridge: Cambridge University Press.

Yamada, H., Nakagaki, T., Baker, R. E., and Maini, P. K. (2007), 'Dispersion relation in oscillatory reaction-diffusion systems with self-consistent flow in true slime mold', Journal of Mathematical Biology, 54, (6), 745-760.

---

[1] More precisely, it's an example of the version of embodied cognition with which we shall be concerned here. There are other versions of the view. For instance, some embodied cognition theorists concern themselves with the way in which embodiment has an impact on our understanding of perceptual experience (e.g. O'Regan and Noë 2001, Noë 2004). Others argue that our embodiment structures our concepts (Lakoff and Johnson 1980, 1999). This sample is not exhaustive.

[2] The classic presentation of what I am calling the extended mind hypothesis is by Clark and Chalmers (1998). See Menary (forthcoming) for a recent collection of

papers. Rather confusingly, the view has always traded under a number of different names, including close variants of the original monicker, such as the *hypothesis of extended cognition* (Rupert 2004) and the *extended cognition hypothesis* (Wheeler forthcoming a), but also *active externalism* (Clark and Chalmers 1998), *vehicle externalism* (Hurley 1998; Rowlands 2003), *environmentalism* (Rowlands 1999), and *locational externalism* (Wilson 2004).

[3] Many thanks to Xabier Barandiaran for discussion of unicellularity, slime molds and RD systems. Any misunderstandings that remain in the text are mine.

[4] It has often been noted that connectionist networks may be analysed in terms of cognitively relevant functions which need to be specified at a finer level of grain than those performed by classical computational systems (e.g. using mathematical relations between units that do not respect the boundaries of linguistic or conceptual thought), hence Clark's term 'microfunctionalism'.

[5] Two comments: First, although this is not the place to launch into a critique of the details of Adams and Aizawa's position, my view is that while they are right that ExM needs a mark of the cognitive, they are wrong about what that mark might be. Secondly, my appeal to a scientifically informed, theory-loaded mark of the cognitive will, in some quarters, be controversial. For example, Clark (2008b) suggests that the domain of the cognitive should be determined by our intuitive folk-judgments of what counts as cognitive. His supporting argument is (roughly) that our intuitive understanding of the cognitive is essentially locationally uncommitted, while the range of mechanisms identified by cognitive science is in truth too much of a motley to be a scientific kind, and so will thwart any attempt to provide a scientifically driven, theory-loaded account of the cognitive – locationally uncommitted or otherwise. I disagree with this assessment. I hold out for a locationally uncommitted account of the cognitive that is scientifically driven and theory-loaded, on the grounds (roughly) that our intuitive picture of the cognitive has a deep-seated inner bias, while Clark's argument for the claim that there is a fundamental mechanistic disunity in cognitive science is far from compelling (Wheeler forthcoming b).

[6] In previous ExM treatments of Bechtel's logical reasoning studies, Rowlands (1999, pp.168-171) and Menary (2007, also pp.168-171) rely at root not on parity considerations to justify the claim of cognitive extension, but rather on the integration of inner connectionist processing with external symbol systems in order to complete a cognitive task that could not ordinarily be achieved by the

inner networks alone. My own view is that the mere fact that an external resource is necessary to complete a cognitive task is not sufficient to establish cognitive extension, as opposed to a compelling case of embodied-embedded cognition.