

Draft of paper published in Philosophical Studies, symposium on Andy Clark's Supersizing the Mind, 152(3), 417-25, 2011.

In Search of Clarity about Parity

Michael Wheeler

**Department of Philosophy
University of Stirling
Stirling
FK9 4LA
UK**

m.w.wheeler@stir.ac.uk

1. The Parity Argument for Cognitive Extension

Andy Clark's *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (Clark 2008) is, among other things, a characteristically bold and timely defence of the extended mind hypothesis (Clark and Chalmers 1998). According to this hypothesis, which Clark here calls EXTENDED, the physical mechanisms of mind (the material vehicles that realize cognition) sometimes extend beyond the traditional boundaries of skull and skin, such that "actions and loops through nonbiological structure [sometimes count] as genuine aspects of extended cognitive processes" (p.85).¹ In the brief treatment that follows I cannot hope to engage with everything that is worthy of discussion in Clark's rich and exciting text, so I shall content myself with exploring and assessing a central thread in his argument for EXTENDED. That thread revolves around what is called *the parity principle*. Here is how that principle is formulated in *Supersizing the Mind* (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 is this: 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 already qualify as the proper parts of a cognitive trait (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.

So how exactly do we apply the parity principle? Here is a suggestion, one that has often crept into the debate, even though, as we shall see, it does not reflect Clark's (or indeed Chalmers') own understanding of the idea: first we fix the benchmarks for what it is to count as a proper part of a cognitive trait by identifying the details of the causal contribution made by the brain (at some appropriate level of algorithmic abstraction); then we look to see if any external or extended elements meet those benchmarks. This suggestion invites scepticism

¹ All page numbers refer to *Supersizing the Mind* (Clark 2008), unless otherwise indicated.

about EXTENDED, since the disbeliever will point to certain features of some cognitive trait as standardly (internally) conceived that are not shared (or perhaps not shared in the right way) by any extended arrangement that might plausibly be thought to perform the same cognitive task, and then argue that since the parity principle is not satisfied, EXTENDED is false. Moreover, there is a way of developing this argument which makes the initial prospects for EXTENDED appear even bleaker. Consider: psychologists tell us that standard internal psychological memory exhibits phenomena such as the generation effect (subjects gain a mnemonic advantage by generating their own meaningful connections between paired associate items to be learned), negative transfer interference effects (past learning interferes with the learning and recall of new paired associations), and recency and primacy effects (we are better at recalling the elements at the beginning and end of a list than we are at recalling the elements in the middle). But if we think about the kind of person-plus-notebook system that is often used as an example of a distributed memory system (with information stored in the notebook and accessed via sensory-motor control, rather than stored in the brain and accessed via purely neural recall mechanisms), there is no obvious reason to think that any of the effects just mentioned will be replicated. If one has stored a list in a notebook, and recall involves retrieving the relevant information from that notebook, typically in some fast, fluid and largely automatic way, in line with how organic memory functions, there is no compelling reason to think that, for example, items at the beginning and end of that list will be retrieved more reliably than items in the middle. So, if the purely inner memory system sets the benchmark for parity, then parity fails, taking with it the parity argument for cognitive extension. And notice that the pivotal appeal to data from empirical psychology here means that the sceptical challenge is arguably founded not on some un-argued-for pro-inner prejudice or unwarranted conservatism, but on a perfectly healthy respect for the methods and results of contemporary cognitive science.²

² *Something like* the argument I have just presented may, I think, be found in the anti-EXTENDED onslaughts of Rupert (2004) and Adams and Aizawa (2008), from whom the psychological examples mentioned in the main text are drawn. It is important to stress two things, however. First, to the extent that these critics do run such an argument, it is a single element in more sophisticated lines of reasoning. Second, and anyway, one might expect the critics in question to hesitate to endorse the argument in the bare form in which it appears here. That said, the bare form will suffice given my current goal, which is to unearth the real logic of the parity principle, not to respond to published criticisms of EXTENDED. For my stab at the latter, see e.g. Wheeler (forthcoming a, b).

How should the fan of the parity argument for EXTENDED respond? It seems abundantly clear that the kinds of functional differences highlighted above do exist, so resistance must, it seems, be based on the thought that they don't matter. One way of pursuing this thought begins with the observation that the appeal to the inner made by the parity principle is in fact *not* an appeal to the kind of fine-grained profile of the extant human inner that talk of primacy and recency effects and the like encourages. The parity principle 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. What happens in the thought experiment, then, is that certain external processes get shifted spatially, across the boundary of the skin, in an inwardly moving direction. Of course, we are not supposed to imagine that the relevant externally located physical elements themselves are grafted onto the brain. Rather, we imagine that exactly the same functional states and processes that are realized in the actual world by those externally located physical elements are now 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 are driven to 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 spatial shift alone is sufficient to result in a transition in the status of the external elements in question, from noncognitive to cognitive, he would, it seems, be guilty of begging the question against EXTENDED. Now notice that at no point in this explanation of how the appeal to the inner contained in the parity principle works have we been forced to use the fine-grained profile of the extant human inner in order to determine what counts as cognitive. In other words, the application of the parity principle does not itself set the benchmark for parity. Instead it acts as a heuristic device designed to free us from what Clark calls "biochauvinistic prejudice" (p.77).

So what does fix the benchmark for parity? Some functional differences will surely be relevant. A mechanism that failed to implement the context-sensitive storage and retrieval of information simply wouldn't be memory, *wherever* it happened to be located. What seems to be needed is some kind of scientifically informed *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 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 EXTENDED 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*.³ In effect, the parity principle, as I have interpreted it, is an appeal for equal treatment against an unbiased and theoretically motivated standard of what counts as cognitive. How could anyone complain?

2. Theories and Motleys

As it happens, Clark rejects a key aspect of the strategy that I have just recommended. A careful consideration of the following remarks will help us appreciate why.

The parity probe was thus 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. Thus understood, parity is not about the outer performing just like the (human-specific) inner. Rather, it is about equality of opportunity: avoiding a rush to judgment based on spatial location alone. 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

³ What I have called a 'theory-loaded, locationally uncommitted account of the cognitive' is tantamount to what Adams and Aizawa (2008) call a 'mark of the cognitive'. Although this is not the place to go into detail, my view is that Adams and Aizawa are right that EXTENDED needs a mark of the cognitive, but wrong about what that mark might be. This has implications for whether or not extended brain-body-world systems are likely to exhibit such a mark. And just so that no one ends up feeling cheated, I should say that nowhere in the present treatment do I specify the content of any theory-loaded, locationally uncommitted account of the cognitive. (For more on that issue, see Wheeler forthcoming c.) Here I am interested in the *structure* of the parity argument for EXTENDED.

digestion—but to do so without the pervasive distractions of skin and skull. (p.114)

As I understand it, this is an attempt to reject the fix-according-to-the-inner approach to parity and to replace it with an equal-treatment approach, while refusing to endorse the idea that equal treatment should be judged against some theory-loaded benchmark for parity. Clark's rejection of the theory-based option is revealed by the fact that he recommends an appeal to intuitive folk judgments about what belongs to the domain of the cognitive in order to provide the independent standard required by the equal-treatment approach. Such intuitive judgments stand in contrast to the theory-loaded account of the cognitive that I have advocated. But why does Clark shy away from the theory-based option? The explanation, I think, lies with his view that the range of underlying mechanisms that we collect together using terms such as 'mind' and 'cognition' is quite possibly too much of a mixed bag, too fundamentally disunified, too much of a *motley*, to be a scientific kind (p.95). The further thought would then be that such a motley is unlikely to reward any attempt to provide a systematic theory-loaded account of what it is to be part of the cognitive. Of course, to give up on the prospect of a theory-loaded account of the cognitive is not yet to give up on the idea of any sort of science of cognition. Thus Clark depicts the post-motley cognitive science that might be on the cards as needing "to embrace a variety of different explanatory paradigms whose point of convergence lies in the production of intelligent behaviour" (p.95), and elsewhere, in a more strident voice, as "a science of varied, multiplex, interlocking and criss-crossing causal mechanisms, whose sole point of intersection may consist in their role in informing processes of conscious reflection and choice" (Clark forthcoming, p.39). To his credit, then, Clark offers an alternative conception of cognitive science that does without any systematic theory-loaded account of what it is to be part of the cognitive. It seems to me, however, that such an alternative is the preferred option only if the motley-based argument against any theory-loaded account goes through. What, then, of that argument?

The first thing to notice is that the scale of underlying disunity in the vehicles of cognition that Clark intends to convey by his talk of a motley is strikingly extreme – so extreme in fact that there will not even be "a family resemblance (at the level of actual mechanism) to hold [those vehicles] together" (p.95). This is not mere rhetoric. Clark needs the disunity to be this radical, if it is to undermine the attempt to provide a systematic theory-loaded account of what it is to be part of the cognitive. Of course, even a mild degree of disunity would rule out the chance of finding some small set of non-disjunctive necessary and sufficient

conditions for a trait to be cognitive. But all that shows is that one particular kind of account is unlikely to yield dividends. Admittedly, if we really did arrive at the view that the vehicles of cognition defy even a family resemblance story, then the game would be up, but, as we are about to see, it's genuinely hard to see how contemporary cognitive science delivers a mandate for such a verdict.

When Clark (pp.95-6) describes the patchwork of mechanisms that he takes to provide evidence for his motley, he identifies a number of dimensions of difference culled from the orthodox, and thus inner-oriented, psychological and neuroscientific literature. This temporary focus on purely inner mechanisms makes sense in the context of Clark's framework, since his "suspicion" (p.95) is that there is no reason to think that the space of extended mechanisms will be any more disunified than the space of purely inner mechanisms. This suspicion strikes me as correct. But one implication is that if the existence of a motley is not established by what scientific evidence tells us about the character of the inner, prior to any consideration of EXTENDED, then there is no reason to think that adding an extended dimension to the picture will produce a different outcome. Here, then, is some of the evidence that Clark submits (see Clark's text for supporting references). The mind as conceived by orthodox cognitive science seems to contain slow, conscious, controlled processes that degrade rapidly as cognitive load increases and that allow conscious interruption. However, it also seems to contain fast, automatic, uncontrolled processes that do not degrade with load or allow conscious interruption. It may well contain look-up-tables as well as more complex combinatorial representational formats. It seems to contain motor representations as well as non-motor representations, and these different kinds of representation may support different episodes of imagination (e.g. mentally rehearsing a golf swing versus imagining a sunset over the sea). It seems to contain object-based visual representations that are abstracted from the egocentric particulars of the related visual stimulation, plus action-relative visual representations underpinned by neuronal cells that respond to a stimulus only when that stimulus is the target of, say, a grasping arm action or a saccadic eye movement.

The existence of this range of states and mechanisms is, of course, interesting and important, but it is not enough to establish the claim that the vehicles of cognition are a motley, in the sense that Clark requires. To explain: despite the nature of the evidence that he himself produces, Clark needs the alleged disunity in the vehicles of cognition to run deeper than the presence of multiple differences in things like, speed, controllability, or accessibility to consciousness. Indeed, if we take differences in algorithmic organization and high-level

knowledge representation to be indicators of disunity, then even classical AI would count as a motley. Compare, for example, the STRIPS means-end analysis planning algorithm with the AQ11 incremental learning algorithm, or production rules with semantic nets. Surely the co-presence of these different symbol-processing algorithms and language-like representational structures in the vehicles of cognition would be no evidence of the kind of disunity in which we are interested, especially when all of them are presumably flag-bearers of Newell and Simon's (1976) famous claim that a suitably organized physical symbol system (roughly, a classical AI system) has the necessary and sufficient means for general intelligence. And issues such as differences in speed, controllability and accessibility to consciousness can certainly be worked in to the classical story about our cognitive architecture without disrupting the basic unity of that architecture. For example, classical natural language processing algorithms for analyzing syntax, when implemented in human brains, were presumably supposed to be fast and resistant to conscious control, whereas many expert systems were supposed to capture aspects of conscious and deliberate problem solving. Any disunity here is, then, superficial.

To see if we can generate the kind of disunity that Clark needs, let's add mainstream connectionism into the picture and suggest that the vehicles of cognition are a combination of classical systems and orthodox connectionist systems. Our hypothetical cognitive architecture now features neurally inspired distributed representations and global computation by patterns of spreading activation alongside classical elements. Nevertheless, the fact is that, at a fundamental level, all these different mechanisms, classical and connectionist, share a series of common deep assumptions about how cognition works, for example that it's essentially a matter of representational states transformed according to computational processes. So any motley-like qualities of the architecture remain essentially superficial. (For much more on the shared aspects of classical and mainstream connectionist thought, see Wheeler 2005).

Where next? Recent (and not-so-recent) work on the ways in which dynamical systems might underpin cognitive activity may well encourage us to be rather bolder in the way we construct our picture. I have argued previously (Wheeler 2005) that our cognitive architecture is a shifting arrangement of (i) noncomputational, nonrepresentational dynamical systems, (ii) noncomputational, representational dynamical systems, and (iii) computational, representational dynamical systems. Is this the motley that Clark needs? I don't think it can be, since there is clearly a fairly straightforward family resemblance

structure to this space of mechanisms: (i) and (ii) are noncomputational in character, (ii) and (iii) are representational.

One could, of course, keep looking for Clark's motley, but I'm willing to bet that however hard one tried, contemporary cognitive science would refuse to play ball. Cognitive science provides no compelling evidence for the existence of a radical or fundamental disunity in the vehicles of cognition. And if there is no such disunity, there is no pressure on us from this quarter to be sceptical about the prospects for a scientifically informed, theory-loaded, locationally uncommitted account of the cognitive.

3. (Don't) Ask the Audience

The credentials of the theory-loaded account are enhanced further once we reflect for a moment on just what our pre-theoretical intuitions tell us about the domain of the cognitive. For as far as I can see, our contemporary pre-theoretical understanding of that domain includes a presumption of the within-the-skin internality of cognition that makes that understanding an unlikely source for the kind of locationally uncommitted benchmark for parity that EXTENDED needs. Clark considers this possibility, but is "inclined... to dispute the claim that the Extended Mind Model runs so wildly contrary to common sense" on the grounds that any presumption of within-the-skin internality results from the fact that "we are already in the grip of a form of theoretically loaded neurocentrism" (p.105) and that the "folk grip on mind and mental states... is surprisingly liberal when it comes to just about everything concerning machinery, location, and architecture" (p.106). Of course, as Clark himself has persuasively argued in the past (Clark 1989), various theoretical views from philosophy and science may, over time, become incorporated into what counts as our folk or pre-theoretical understanding of cognition (the Freudian unconscious would be one example). So it is at least arguable that what began as a theoretically loaded neurocentrism has now become an integrated *part of* our folk understanding of the domain of the cognitive, presumably through its capacity to give modern currency to a phenomenal sense of internality that preexisted contemporary scientific views of the brain as the seat of cognition. What this indicates is that the argumentative weight in Clark's position is carried not by the allegation of theoretically loaded neurocentrism but rather by whatever positive case can be built for the liberality of our folk understanding of where the mental might be located. To develop such

a case, Clark (pp.105-6) appeals to an analysis of mental content and external representational media due to Houghton (1997).⁴

In his prescient paper, Houghton claims to find direct support for cognitive extension in examples of our ordinary, everyday folk practices of attributing intentional states. For instance, in an example that Clark cites approvingly, Houghton argues that the content of an architect's intentions regarding the detailed structural design of a future building need not, typically will not, and indeed sometimes could not, be fully internalized by the architect. Rather, it's the architect's drawings that give her intentions detailed content, by fixing that content even though her brain does not. On the basis of such examples, Houghton concludes that "it is undeniable that we commonly credit people with intentional states whose contents they themselves never fully internalize" (Houghton 1997, p.166). And later: to "deny that these are cases of genuine intentional content would be to use the notion of content in some way as yet to be explained which is clearly at odds with our ordinary attributions of intentional states" (ibid p.166).

The problem with Houghton's argument (and thus with Clark's appeal to it) is located just beneath the surface of these conclusions. At root, Houghton conflates EXTENDED (the claim that the material vehicles that realize cognition are sometimes partly externally located) with the philosophical position known as *content externalism* (roughly, the claim that the factors to which one sometimes needs to appeal in order to individuate mental states by their content are externally located). Although the practices of intentional-state attribution as embraced by the folk may well display an implicit endorsement of content externalism, it is far less clear that they display a similar endorsement of EXTENDED's *vehicle* externalism. Thus the way in which the folk naturally attribute intentional content to an architect may well reflect the view that the content of an architect's drawings fixes the detailed contents of her plans and

⁴Clark remarks (p.105) that although he believes our folk (intuitive, pre-theoretic) model of mind to be, at heart, locationally uncommitted, nothing in the appeal to that model in his argument for EXTENDED depends on establishing this liberality, because "all the argument requires is an appeal to some notion of the coarse (i.e., unscientifically visible) role associated with some mental state". I find this remark puzzling, since it seems to me that folk psychology had better be at least consistent with EXTENDED, if it is to play any sort of supporting role in the picture. Moreover, given my analysis in the main text of where the argumentative weight lies here, Clark's remark seems misplaced.

intentions regarding the nature of the building. And recognizing that fact has important philosophical consequences. But if we deliberately tried to prise apart our two forms of externalism, and so asked the folk a specifically vehicle-targeting question – for example, ‘Where in space are the relevant cognitive states of the architect realized?’ – I have no doubt that we would receive an internalistic, skin-side answer. Moreover, it seems, the ordinary practices of the folk reflect this answer. If an environmental protester had stolen the plans of Heathrow Terminal 5, the folk would most likely have been interested, and either supportive of the act or outraged by it, depending on what other beliefs were in play. But presumably none of these attitudes would be held because the folk were considering the whereabouts of (to speak loosely) part of Richard Rogers’ mind. A plausible explanation for this pattern is that our folk grip on the cognitive involves a presumption of the within-the-skin internality of cognition.

Notice that I am not disputing the claim that the realizing vehicles of the architect’s intentional states may *in fact* be extended over brain, body and world. That may or may not be true. My point is that if this is a case of cognitive extension, then the fact that it is so suggests a theory-loaded revision to our intuitive understanding of the domain of the cognitive, one that removes the presumption of the within-the-skin internality of cognition. This in turn implies that, pace Clark, our intuitive understanding of the cognitive is not apt to provide the locationally uncommitted benchmark for parity for which we are searching. That benchmark will need to be a theory-loaded construction.

4. The Devil and the Deep Blue Sea

I have argued that serious doubt may be cast on Clark’s claim that the vehicles of cognition are characterized by a fundamental disunity. That was the claim that was supposed to motivate us to look to our intuitive pre-theoretical judgments to provide a locationally uncommitted account of the cognitive as a benchmark for parity. And we have since discovered that, in any case, the presumption of the internality of cognition embedded in our folk understanding of the cognitive means that we have no good reason to think that our intuitive judgments about the domain of the cognitive are up to the task of generating such an account. Given his rejection of any theory-loaded alternative here, this leaves Clark without a locationally uncommitted account of the cognitive. In the absence of such an account, and given that the benchmarks for parity will need to come from somewhere, one might again be tempted by the direct analogical route from what psychologists tell us about our innards to the determination of what counts as part of the cognitive. But, as we saw earlier, this kind of strategy ultimately

leaves the parity argument for cognitive extension vulnerable to criticisms that exploit certain inevitable fine-grained functional disparities between wholly inner and extended solutions. Between the devil of the internalistic folk and the deep blue sea of extension by straightforward analogy with the extant inner lies the solution that Clark needs but unwisely dismisses: a benchmark for parity specified in terms of a scientifically informed, theory-loaded, locationally uncommitted account of the cognitive.

References

- Adams, F. and K. Aizawa. 2008. *The Bounds of Cognition*. Malden, MA and Oxford: Blackwell.
- Clark, A. 1989. *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, Mass.: MIT Press.
- Clark, A. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.
- Clark, A. Forthcoming. Memento's revenge: The extended mind, extended. To appear in Menary forthcoming.
- Clark, A. and D. Chalmers, D. 1998. The extended mind. *Analysis*, 58 (1): 7-19.
- Houghton, D. 1997. Mental content and external representations. *Philosophical Quarterly*, 47 (187): 159-177.
- 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.
- Rupert, R. 2004. Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101 (8): 389-428.
- Wheeler, M. 2005. *Reconstructing the Cognitive World: the Next Step*. Cambridge, Mass.: MIT Press.

Wheeler, M. Forthcoming a. In defence of extended functionalism. To appear in Menary forthcoming.

Wheeler, M. Forthcoming b. Minds, things, and materiality. To appear in *The Cognitive Life of Things: Recasting the Boundaries of the Mind*, C. Renfrew and L. Malafouris (eds.). Cambridge: McDonald Institute for Archaeological Research Publications.

Wheeler, M. Forthcoming c. Embodied cognition and the extended mind. To appear in *The Continuum Companion to Philosophy of Mind*, J. Garvey (ed.). London: Continuum.