Draft of piece published in the American Philosophical Association Newsletter on Philosophy and Computers, 8(1), 9-12, 2008.

The Fourth Way: a Comment on Halpin's "Philosophical Engineering"

Michael Wheeler
Department of Philosophy
University of Stirling
m.w.wheeler@stir.ac.uk

It is common these days to distinguish between three kinds of cognitive science or artificial intelligence: classical, connectionist, and (something like) embodied-embedded. Of course, all such attempts at neat-and-tidy categorization are undoubtedly guilty of over-simplification in one way or another. For example, researchers sometimes build models that combine aspects of more than one approach (e.g. when conventional connectionist networks are used as control systems for embodied agents). That noted, however, one method for separating out our three kinds of cognitive science, so as to understand more clearly their basic theoretical commitments, would be to identify, in a very general way, the sorts of machine that each takes to capture the fundamental character of intelligence. If we adopt this strategy, classicism will be defined by the manipulation of symbols using structure-sensitive processes, connectionism by unfolding patterns of activity in neurally inspired networks of simple processing units, and embodied-embedded thinking by complete autonomous robots engaged in perceptually guided motor activity. One of the many fascinating claims in Harry Halpin's strikingly original article "Philosophical Engineering: Towards a Philosophy of the Web" (Halpin 2008) is that the Web constitutes a fourth conceptual anchor for the notion of mind as machine. Halpin's view, in short, is that the Web provides a general model of a computational machine that compels us to rethink the notion of representation, while simultaneously radicalizing our conception of cognition through a vindication of the idea that minds may be realized partly by factors located beyond the skin. In this comment on Halpin's article, I shall engage briefly with just some of the issues that confront us once we take this fourth way.

With apologies for the immediate whiff of self-centredness, I shall begin by considering an argument from Halpin's paper that responds explicitly to some of my own previous work. I have been known to claim (e.g. Wheeler

2005) that any adequate account of representational explanation in cognitive science must have the consequence that while certain inner (within-the-skin) elements count as representations, most external (beyond-the-skin) elements don't. The justification for this restriction is largely methodological: it seems likely that *neural* states and processes do something that is, for the most part, psychologically distinctive, and we expect the concept of representation to help us explain how that something comes about. Thus the constraint at issue may be specified more carefully as what I call the *neural assumption*. The neural assumption states that if intelligent action is to be explained in representational terms, then whatever criteria are proposed as sufficient conditions for representation-hood, they should not be satisfied by any extra-neural elements for which it would be unreasonable, extravagant, or explanatorily inefficacious to claim that the contribution to intelligent action made by those elements is representational in character. For if such illegitimate external factors qualified as representations, the claim that some neural state has a representational character would fail to single out what was special about the causal contribution to intelligent behaviour made by that state. Notice that the neural assumption, as formulated, is liberal enough to allow *some* external factors to qualify as representations in the sense that is relevant for cognitive-scientific explanation. However, it is clear that representations construed this way will remain largely inside the head.

Halpin distances himself from this approach to representation, arguing that once our intellectual goal becomes a philosophy of the Web, as opposed to a philosophical account of how representation figures as an explanatory primitive in cognitive science, any inner-focused account of representation (such as my own) will fail to deliver what theory demands. As he puts it, the Web is "nothing if not a robustly representational system, and a large amount of research on the Web focuses on how to enable increasingly powerful and flexible forms of representations" (Halpin 2008, p.6). Thus "[w]hat we need is a notion of what a representation is, a definition that applies to both "internal" and "external" representations, not conditions for a representational explanation in cognitive science" (Halpin 2008, p.7). In other words, what a philosophy of the Web requires is a suitably generic, locationally uncommitted account of representation that, in principle, applies equally to internal representations (those located within the skin) and external representations (those located outside the skin, such as those on the Web). Without such an account, we will be unable to make sense of the Web as a representational system. In the light of this analysis, Halpin proceeds to sketch a proposal for what it is for something to be a representation. Here he draws, in part, on Smith's (1996) notion of representation via *registration*, according to which the distinction between subject and object, and thereby between representation and represented, emerges from the dynamics of

certain physical processes in which one region of space-time tracks the behaviour of another.

I will be concerned not with the plausibility of Halpin's positive proposal, but rather with the alleged need for any unitary, locationally uncommitted account of representation. For it seems to me that, from the present perspective, although we need *a* concept of representation that illuminates the character of representational explanation in cognitive science, plus *a* concept of representation that makes sense of external representations (and thereby of the Web as a representational system), there is no reason to think that it must be the *same* concept of representation in both cases. Indeed, there are considerations which suggest that theoretically significant differences are to be expected. For example, when external representations are used to guide intelligent behaviour, they do so via perception-action loops. Thus consider familiar cases of visual maps, whether paper or electronic. Such representations are able to direct behaviour because the agent *looks at* and *performs an embodied spatial manipulation of* the map-realizing elements (the atlas or the PDA). No such perception-action engagement with the behaviour-guiding representations are present when we use neurally realized internal maps (assuming there are such things) to navigate around the world. One might expect these sorts of differences to have an impact upon the nature of the representations in question. Moreover, Halpin himself identifies certain principles that (he argues) not only characterize the external representations used by the Web, but also perhaps explain the intelligence-facilitating effects of the Web. It is hard to see how these principles (universality, inconsistency, self-description, least power and the open world – see Halpin 2008, p.9, for the details) apply to neural representations.

Of course, given this pattern of divergence, we need some reason to conclude that what we call internal representations and what we call external representations are both genuine members of some overarching category of representational elements. For this, however, it is sufficient that (a) the alternative notions be linked by the vague pre-theoretical thought that, to be a representation, a state or process should play some sort of standing-in-for function, and (b) there should be some sort of family resemblance structure in play. Evidence for (b) may be found in the observation that familiar cases of external representations (e.g. mathematical symbols) plausibly share certain properties with neural representations, properties such as multiple realizability and being the bearers of consumed information.

That said, Halpin's nervousness about my inner-focussed account of representational explanation in cognitive science may have an alternative source. To see this, we need to plug in the relationship that, according to

Halpin, exists between the external representationalism of the Web and (what is sometimes called) the *extended mind hypothesis* (Clark and Chalmers 1998). In general terms, those who believe in the extended mind hold that there are conditions under which cognitive states, processes, mechanisms, architecture, and so on may be partly realized by material elements located beyond the skin. Halpin's view is that the universal information space of the Web supplies a dynamic and open-ended suite of such elements. In other words, the ways in which we store, retrieve, manipulate and transform representational structures on the Web mean that, under certain conditions, some of our cognitive traits are partly realized by those structures. Dramatic examples of such cognitive extension occur when multiple agents remotely access and update a shared map on the Web. In such cases, the "active manipulation of a representation lets [the two agents] partially share a dynamic cognitive state and collaborate for their greater collective success… via shared external representations that are universally accessible over the Web" (Halpin 2008, p.8).

This changes things. If we are to make sense of the Web not only as a representational system, but as a representational system whose elements may sometimes constitute part of an agent's cognitive architecture, then one might think that the pressure in the direction of a unitary account of representation increases. After all, given that certain representational structures on the Web are now to be granted cognitive status, it seems that an adequate account of representational explanation *in cognitive science* will need to apply not only to familiar inner elements such as neural states and processes, but also to those external structures. In other words, any purely inner-focused account of representation is now revealed as failing to deliver what *cognitive* theory demands. Although, to my mind, Halpin himself does not clearly separate out the present argument from the one with which we began (which doesn't turn on the putatively cognitive status of the external representational structures), it seems to be the present argument that offers the more compelling case for the view that we need a unitary, locationally uncommitted account of representation.

Halpin's analysis alerts us to the fact that once the idea of cognitive extension is on the table, the neural assumption (see above) needs to be separated out from what we might now dub the *global adequacy requirement* – the demand that we develop an account of representation suitable for the task of cognitive-scientific explanation. The latter is what Halpin (2008, p.7) calls the "conditions for a representational explanation in cognitive science". In my previous work I have been guilty of running together the global adequacy requirement and the neural assumption (as indicated by the discussion of the neural assumption included above). Once we pull these analytical structures

apart, however, we can see that the strategy of appealing to different notions of representation – the strategy that, as we saw, made sense of external representation under a non-cognitive interpretation – will also make sense of external representation under a cognitive interpretation. To see why this is, notice first that, depending on how one carves up nature into cognitive and non-cognitive regions, an account of representation that meets the neural assumption may not meet the global adequacy requirement. Consider: if externally located representations on the Web figure as genuine parts of cognitive processes, the global adequacy requirement will not be met by an account of representation that respects the neural assumption – or at least not by that account *on its own*. But that, of course, is the key point. For the global adequacy requirement may be met by a varied explanatory toolkit encompassing different notions of representation designed for different explanatory tasks, such as understanding how neural states contribute to intelligent behaviour, and illuminating how external representations may figure as genuine parts of cognitive processes. What this suggests is that, with the extended mind added to the mix, and the concept of cognitive representational space expanded to include external structures, that space may reflect the same pattern of similarities and differences between internal and external representations that we identified earlier. Thus, even under a cognitive interpretation, the unitary notion of representation that Halpin seeks may be no more than a philosophical chimera.

So far, I have been assuming that Halpin is right that, under certain circumstances, the Web forms part of our cognitive resources. I now want to interrogate that idea – not, I hasten to add, because I think it's obviously wrong, but because we need to be clear about what a good argument for that conclusion would look like. The first thing to note here is that the extended mind hypothesis is a view about the whereabouts of mind that is distinct not only from the position adopted by orthodox cognitive science (classical or connectionist), but also from the position adopted by any merely embodied-embedded view. To illustrate this point, we can adapt an example originally due to Rumelhart et al. (1986). Most of us solve difficult multiplication problems using pen and paper. The pen and paper system is a beyond-the-skin factor that helps to transform a difficult cognitive problem into a set of simpler ones, and to temporarily store the results of intermediate calculations. For orthodox cognitive scientists and for supporters of the merely embodied-embedded view, that pen and paper system is to be conceived as a non-cognitive environmental prop. It is an external tool that aids certain cognitive processes via embodied interaction, but is not itself a proper part of those processes. Of course, orthodox cognitive scientists and embodied-embedded theorists differ on how best to characterize the interactive arrangement of skin-side cognitive processes and external prop. In particular, the embodied-

embedded theorist is likely to count the bodily activity involved as itself a cognitive process, as opposed to a mere output of neurally located cognition, and to trace rather less of the source of the manifest complexity of the observed behaviour to the brain, and rather more to the structured embodied interactions with the external pen and paper system. For all that, however, both of these camps think of cognition as a resolutely skin-side phenomenon. By contrast, the extended mind theorist considers the causally coupled combination of pen-and-paper resource, appropriate bodily manipulations, and in-the-head processing to be a cognitive system in its own right. We can now pinpoint the right question to ask: does Halpin's analysis indicate that certain manipulations of the Web's universal information space constitute genuine cases of cognitive extension rather than merely embodied-embedded intelligence?

Halpin sometimes seem to suggest that cognitive extension results whenever an adaptive causal coupling between inner and outer elements produces an intelligent outcome. Thus recall his example of two agents whose intelligent behaviour is structured by shared remote access, via mobile telephones, to a web page containing a map. He implies that coupling considerations are sufficient for cognitive extension when he writes that "[s]ince [the two agents] are sharing the representation and their behavior is normatively successful based on its use, [they] can be said to partially share the same cognitive state" (Halpin 2008, p.8). A more sophisticated version of the coupling argument for cognitive extension emerges during Halpin's subsequent discussion of the ways in which the coupled combination of analogue organic processing with external digital computer memory enable human beings to succeed at cognitive tasks that are poorly tackled by unaided organic processing. This is a particularly striking example of the ways in which human cognition may be transformed through the integration of internal processing with external props and scaffolds that possess a different range of fundamental properties. Unfortunately, however, even given the transformative effects brought about by integrated bio-technological couplings, we don't yet have an argument for cognitive extension. As Adams and Aizawa (2008) forcefully point out, all coupling-based arguments for cognitive extension are dangerously insensitive to a crucial causal-constitutive distinction, that is, to the distinction between cognition being merely causally dependent on some factor, and to cognition being constituted by, or partly constituted by, that factor. The cognitive activities of Halpin's remote-map-using agents, as well as those of his digitally embedded brains, are surely causally dependent on external factors in ways to which traditional theorizing in cognitive science has been largely oblivious, but that is not enough to secure the cognitive status of those factors.

The main alternative to coupling-based arguments for cognitive extension is what, in the literature, is known as the *parity principle* (Clark and Chalmers 1998). Exactly how one should formulate the parity principle remains a matter of some dispute (Clark 2007; Wheeler forthcoming), but the general idea is that if there is functional equality with respect to governing 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 process, then there is no good reason to deny equivalent status – that is, cognitive status – to the relevant external elements. Halpin (2008, p.8) quotes Clark and Chalmers' original statement of the parity principle, but it is unclear to what extent he gives weight to parity considerations as opposed to issues of coupling and integration. However, because the parity principle appeals, at root, to the notion of functional equivalence and not mere coupling, it does not run roughshod over the causal-constitutive distinction. So provisionally at least, the parity-driven case for cognitive extension is on the firmer footing. In relation to Halpin's arguments, this prompts the following question: is it ever correct to say that there is functional parity between (i) the causal contributions to intelligent behaviour made by those inner factors that qualify as cognitive, and (ii) the causal contributions to intelligent behaviour made by structures on the Web?

As far as I can tell, the answer to this question depends on the specific criteria that one thinks need to be satisfied for a causal contribution to count as cognitive, what Adams and Aizawa (2008) call the *mark of the cognitive*. Such criteria are necessary because, in order to deploy the parity principle, one must be able to isolate just those functions that inner elements perform that mark out their contribution as cognitive (e.g. the functions involved in the context sensitive storage and retrieval of information that might plausibly define the cognitive trait of memory). It is parity with respect to the realization of these particular functional roles that will establish the cognitive status of certain external elements. This introduces a complex issue that certainly cannot be settled here. It is worth noting, however, that if the extended mind theorist adopts a weak or promiscuous enough mark of the cognitive, then it will be easy enough to secure the result that cognition is extended; but the price of this success will be to welcome into the domain of the cognitive all kinds of wildly unlikely cases in a manner that ultimately casts doubt on the ability of the proposed mark to latch onto only what might be thought of as the proper objects of cognitive science. What this aspect of Halpin's project still needs, it seems, is a mark of the cognitive that allows certain external representations on the Web (such as remotely accessible maps just as they guide online intelligent behaviour) to count as cognitive, while denying that same status to 'wildly unlikely cases' (such as books in a home

library or standing mobile telephone access to an Internet search engine meaning that one might dispositionally believe everything on the Web). Put in a more generic way, the problem is to find a path between the dual dangers of a kind of disproportionate elitism (excluding from the domain of the cognitive certain genuinely cognitive traits, just because they happen to be externally located) and a kind of excessive liberality (welcoming in to the domain of the cognitive certain unwanted interlopers, as a side-effect of making conceptual room for extended cognition). Halpin is not alone in facing this problem. Extended mind theorists in general have perhaps failed to realize just how much hangs on it. Nevertheless, it is a problem for Halpin, and one that, I think, he cannot ignore.

My response to Halpin's arguments has necessarily been selective. I could have written another comment purely on the issues that Halpin explores towards the end of his discussion, when he turns his attention to the relationship between bio-technological intelligence and the specific case of the Semantic Web. What I hope to have made manifest, however, is the rich vein of thought that runs through Halpin's paper. For while the power of the Web as a technological innovation is now beyond doubt, the potential power of the Web to have a conceptual impact on cognitive science remains under-appreciated. The second of these contributions is what I have called the fourth way, an intellectual path innovatively revealed by Halpin's article. My critical comments here do no more than point to twists and turns that, in my view, remain to be navigated as we explore that trail. The fourth way may well be the next way.

References

Adams, F. and Aizawa, K. 2008. *The Bounds of Cognition*. Malden, MA and Oxford: Blackwell.

Clark A. 2007. "Curing Cognitive Hiccups: a Defense of the Extended Mind." *Journal of Philosophy* 104, 163-192.

Clark, A. and Chalmers, D. 1998. "The Extended Mind."*Analysis* 58 (1), 7-19.

Halpin, H. 2008. "Philosophical Engineering: Towards a Philosophy of the Web." *APA Newsletter on Philosophy and Computers*, 7(2): 5-11.

Rumelhart, D.E., Smolensky, P., McClelland, J.L. and Hinton, G. 1986. "Schemata and Sequential Thought Processes in PDP models." In *Parallel Distributed Processing: Explorations In The Microstructure Of Cognition, Vol.*

2: Psychological And Biological Models, J.L. McClelland and D. Rumelhart eds. Cambridge, Mass.: MIT Press, 7-57.

Smith, B. 1996. *On the Origin of Objects*. Cambridge Mass.: MIT Press.

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

Wheeler, M. Forthcoming. "Minds, Things, and Materiality." 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.