Draft of paper published in Kendig, C. (ed.), *Natural Kinds and Classification in Scientific Practice*, Routledge, London, 2015, pp.175-85.

A Tale of Two Dilemmas: Cognitive Kinds and the Extended Mind

Michael Wheeler

1. Extended Cognition and Natural Kinds

Most work in cognitive science and naturalistic philosophy of mind is unashamedly internalist in outlook, in at least the following sense: the parts of the physical world where psychological states occur and where psychological processes happen are held to be located entirely inside the head. One's first reaction to this sort of internalism about the mind might well be that it must be right. Indeed, given all those wonderful 'pictures of the brain thinking' that have been delivered over the past few years by contemporary neuroimaging techniques, where else could the material machinery of mind be? Enter the hypothesis of extended cognition (henceforth ExC).[1] If ExC is true, there are actual (in this world) cases of intelligent thought and action, in which the material machinery that realizes the thinking and thoughts concerned is spatially distributed over brain, body and world, in such a way that the external (beyond-the-skull-and-skin) factors concerned are rightly accorded cognitive status. Here, 'cognitive status' is just a place-holder for 'whatever status it is that we standardly grant the brain when explaining intelligent thought and action'.

What has ExC got to do with the topic of natural kinds? To answer this question, we need to say something about what natural kinds are. It seems to me that, in the present context, we can safely go along with Rupert (more on whom below) and start from the following thought. 'Natural kinds are simply the causal-explanatory properties and kinds of the successful sciences, or to be a bit more careful, the properties and kinds that our sciences attempt to identify. As such, they are the kinds or properties that ground successful induction… appear as relata in laws of nature… or play causal-explanatory roles.'[2] This is essentially a *commitment-light* characterization of natural kinds, in that, for all it says, natural kinds may or may not be cluster-based, may or may not be family-resemblance-based, and so on. This is deliberate, since the opening connection with ExC that I am about to describe is neutral on such matters of detail. We will, of course, be focussing our attention on those natural kinds that make up the basic metaphysical furniture of minds, what we might call the psychological natural kinds, or, as I shall say, the *cognitive kinds*. Assuming our commitment-light characterization of natural kinds, cognitive kinds are simply the causal-explanatory kinds that our psychological sciences attempt to identify.

Against this backdrop, ExC becomes linked to the question of cognitive kinds, and, as we shall see, to the more specific theme of taking scientific practice seriously in our understanding of such natural kinds, via a particular argument for ExC that takes its cue from the following claim by Clark and Chalmers: '[b]y using the 'belief' notion in a wider [i.e., extended] way, it picks out something more akin to a natural kind. The notion becomes deeper and more unified, and is more useful in explanation'[3]. Taking it for granted that the practice of explanation that is relevant here is explanation in cognitive science, this claim turns on the thought that a theoretical framework for cognitive science that endorses *extended cognitive kinds* – cognitive kinds whose physical instantiations are spread out over brain, body and world – will be explanatorily more powerful than one that does not. Add in the surely defensible thought that our best theoretical framework for cognitive science is also our most reliable guide to how nature is carved up into the cognitive and the non-cognitive, and we get what Rupert calls the *natural kinds argument for the extended mind* (henceforth NKA).[4]

As it happens, Clark and Chalmers themselves seem to believe that commonsense notions of psychological phenomena ought to play a regulative role in determining the cognitive kinds that are operative in scientific psychology. This opens the door to some thorny issues concerning the relationship between commonsense psychology and extended cognition,[5] but right now we can safely ignore such problems, since Rupert's interpretation of NKA takes commonsense psychology out of the picture. NKA, as Rupert understands it, attempts to establish ExC on the grounds that our best cognitive science will trade in extended cognitive kinds. Maybe these kinds will be identical to, or somehow influenced by, our commonsense psychological notions, or maybe they won't. In any case, it's the practice of science that ultimately calls the metaphysical shots. From now on, then, I am going to leave Clark and Chalmers almost entirely behind, and concentrate on the 'pure-science' interpretation of NKA proposed by Rupert.

2. The Rupert Dilemma

Here is Rupert's formulation of NKA:

> *Premise 1* If the most explanatorily powerful (known) framework for theorizing in a given domain presupposes a given taxonomy of states, we should at least tentatively accept the existence of states of the kinds in question.
> *Premise 2* The most explanatorily powerful (known) framework for theorizing about intelligent behavior presupposes kinds that, in fact, have a significant number of instances external to the human organism.

Conclusion Therefore, we should at least tentatively accept the extended view of human cognition.[6]

Premise 2 tells us that, for NKA to succeed, there would need to be causal-explanatory factors at work in our best known cognitive science that (i) count as cognitive kinds and (ii) have extended instances. Rupert claims that any attempt to find structures that satisfy these two necessary conditions runs aground on a dilemma. Let's call this the *Rupert Dilemma*.[7]

To set things up, Rupert observes that any cognitive kind will be either fine-grained or coarse-grained, and either benchmarked or non-benchmarked. A *fine-grained* kind is one individuated by the sorts of detailed psychological properties and dynamics with which practising cognitive scientists are often concerned. A *coarse-grained* kind is one that is insensitive to such detailed properties and dynamics. A cognitive kind is *benchmarked* if the theorist has begun by singling out an uncontroversial case of a cognitive kind in a recognized group of thinkers and then proceeded to identify further instances of the same kind by way of similarity with that paradigmatic example. Given that the only cognition we know much about is human cognition, the natural choice of benchmark-supplying group will be human thinkers. A cognitive kind is *non-benchmarked* if one's individuation procedure does not turn on any such uncontroversial case. Rupert then assesses the different ways in which these principles of individuation might be combined by the proponent of ExC in her attempt to justify premise 2 of NKA. His test case is a core cognitive phenomenon, namely memory.

Let's begin by pruning the options. Although Rupert doesn't explicitly open his own treatment of NKA with this move, his view, as I understand it, is that there are really only two combinations of the foregoing individuating principles that stand any chance of success. That's because, for Rupert, whether we are on the hunt for fine-grained or coarse-grained cognitive kinds, our theorizing has no viable point of departure other than the known instances of cognitive kinds that are ordinarily thought to be located in human brains. In light of this, any sort of non-benchmarked approach threatens to be a methodological disaster characterized by lost bearings and under-constrained speculations. So any prospect of justifying premise 2 of NKA by appeal to non-benchmarked cognitive kinds, whether fine-grained or coarse-grained, is summarily eliminated. We can simply ignore the two non-benchmarked options.

Two options remain. The first is to adopt a benchmarked, fine-grained approach to cognitive kinds, where the benchmark-supplying group is human beings. In considering this strategy, as applied to memory, Rupert first notes that the widespread view in cognitive science is that human memory is a massively diverse phenomenon involving many different mechanisms. This suggests that, from a fine-

grained perspective, memory isn't a cognitive kind at all, in which case necessary condition (i), as specified above, isn't satisfied. The fan of NKA might try to side-step this worry by taking the fine-grained causal-explanatory factors themselves to be the cognitive kinds. But then a different problem arises, namely that these kinds plausibly won't have extended instances, in which case condition (ii) isn't satisfied. To give just one illustrative example, there are psychological experiments which demonstrate that human organic memory is sensitive to what is called the generation effect, according to which subjects gain a mnemonic advantage by generating their own meaningful connections between paired associate items to be learned. Rupert argues that the generation effect will simply not occur in some candidates for extended memory systems (e.g., in a system according to which, during recall, the subject refers to a notebook in which the paired associates are accompanied by connection sentences produced by that subject during learning, but which were entered into the notebook by the experimenter). He concedes that it might occur in others (e.g., in a system according to which, during recall, the subject refers to a notebook in which the paired associates to be learned are accompanied by connection sentences produced and entered by that subject during learning). In the latter case, however, he suggests that the effect is an accidental feature, rather than an essential or defining dimension, of the candidate storage-and-retrieval system. The moral is this: if we individuate our memory-related cognitive kinds in a fine-grained, benchmarked way (the generation effect being our illustrative example of a fine-grained property of human memory), then they don't have extended instances. Condition (ii) is not satisfied, premise 2 remains unjustified, and NKA fails.

The second, and only remaining, option is to adopt a benchmarked, coarse-grained approach to memory. Here, Rupert targets the issue of coarseness of grain (so, in fact, his argument would apply equally to a non-benchmarked, coarse-grained approach, but that option has already been eliminated). Rupert claims that if we individuate memory in a coarse-grained way, for example as the context-sensitive storage and retrieval of information, then although we may secure a structure that does important work in organizing the explanatory practices of cognitive psychologists, what we don't get is a cognitive kind. Consider the following case[8], which may be thought of as an example of benchmarking, assuming our benchmark-supplying group can be stretched to include conceivable human beings. If a team of orthodox cognitive psychologists were introduced to an unusual human subject whose purely organic memory system didn't exhibit the generation effect, but who nevertheless continued to achieve the context-sensitive selective storage and retrieval of information, it seems clear that they wouldn't pronounce this subject to be lacking a memory. What this seems to indicate is that the generation effect is an accidental, rather than a defining, feature of human memory, and therefore that it would be a mistake to attend to such fine-grained features when individuating one's cognitive kinds. But now notice that if exhibiting the generation effect is not necessary for a system to instantiate the cognitive kind of (coarse-grained) memory,

then the failure of an extended storage-and retrieval system to exhibit that phenomenon is no barrier to it realizing the same cognitive kind. NKA is back in the game.

In response, Rupert argues that the genuine and important contribution that may be made by the notion of coarse-grained memory, in organizing and shaping the explanatory practices of cognitive psychologists, isn't sufficient for it to count as a genuine natural kind. Here Rupert draws an analogy with the notion of 'Bell Labs research', which is a convenient and useful notion for grouping together all the innovations that were developed in those labs, but one which carries no implication that the research thereby grouped together exhibits any fundamental unity. Similarly, 'memory' might be a convenient and useful term for grouping together some diverse processes that share certain behavioural similarities, but it carries no implication that the processes in question exhibit the kind of fundamental unity that would make memory a natural kind. So, how do we know when we are using an authentic coarse-grained natural kind term, rather than a merely organizational one? Rupert argues that what distinguishes a genuine coarse-grained scientific natural kind from a mere pragmatically useful grouping is that the various instances of a coarse-grained natural kind bear family resemblances to each other determined by overlaps between (a) the causal-explanatory elements that constitute those instances and (b) the relations between those elements. What this suggests is that if two phenomena are instances of a genuine coarse-grained natural kind, we will be able to get smoothly from the first to the second by, as Rupert puts it, 'tweaking and extending' the relevant scientific model (by adding a term, say, or by adjusting certain parameter values). This picture of natural kinds is mandated, argues Rupert, by the practice of scientists in taking an existing model of some recognized phenomenon and tweaking or extending that model in various ways in order to explain some new phenomenon, with the thought that if this can be achieved, then the new phenomenon is of the same kind as the first (that is, there is a coarse-grained kind that subsumes them both), whereas, if this cannot be achieved, the new phenomenon is of a different kind to the first (that is, there is no coarse-grained kind that subsumes them both).

At this point, one might complain that Rupert has just smuggled in an account of coarse-grained natural kinds that goes beyond the commitment-light notion of a natural kind that he endorses earlier in his treatment (see above). This is an issue to which we shall return. What is important right now is that if we generalize this part of Rupert's argument, we get the following result: plugging coarse-grained groupings of the sort envisaged – benchmarked or not – into NKA would deliver ExC, *if* those groupings were causal-explanatory in the right way so as to count as natural kinds; but they are not. Condition (i) is not satisfied, premise 2 remains unjustified, and NKA fails.

We can now formulate the Rupert Dilemma.

> *First horn*: If the fan of ExC individuates her cognitive kinds in a benchmarked, fine-grained way – say by appealing to the explanatory factors typically of interest in established human cognitive psychology – then the external elements in any candidate distributed system fail to count as cognitive kinds. Given the fact that, according to NKA, we find cognitive states and processes where, and only where, the cognitive kinds are, this means that the external elements in question fail to enjoy cognitive status (in the sense outlined earlier). So NKA fails.
>
> *Second horn*: If, by contrast, the fan of ExC attempts to individuate her cognitive kinds in a benchmarked, coarse-grained way – one that is insensitive to the fine-grained character of human psychology as revealed by established cognitive psychology – then the structures that she ends up with are not cognitive kinds at all (wherever they are located). Given the fact that, according to NKA, we find cognitive states and processes where, and only where, the cognitive kinds are, this means that the external elements in question fail to enjoy cognitive status (in the sense outlined earlier). So NKA fails.

Given the assumption that we have exhausted all the available options for individuating cognitive kinds, Rupert concludes that NKA should be rejected.

3. Extended Physical Symbol Systems

The first horn of the Rupert Dilemma would be defused, if there are benchmarked, fine-grained cognitive kinds with extended instances. I think it plausible that such kinds exist. Here is just one class of examples.

Bechtel[9] defends the view that high-end cognitive achievements such as linguistic behaviour, natural deduction and mathematical reasoning are often the result of sensorimotor-mediated interactions between internal connectionist networks (processing architectures inspired by the abstract organization of the brain) and certain external representational systems (e.g., mathematical languages, natural languages) in which atomic symbols are combined and manipulated according to the principles of a compositional syntax and semantics. The capacity of connectionist networks to recognize, and to generalize from, patterns in training data, plus the temporal constraints that characterize real embodied engagements with stretches of external symbol arrangements (e.g. different parts of the input will be available to the network at different times, due to the restrictions imposed by temporal processing windows) are harnessed to allow those networks to be appropriately sensitive to the structural properties of the external symbol system.

Bechtel himself seems to hold that the genuinely cognitive part of the proposed distributed solution here remains skin-side.[10] But there is an alternative view available.[11] Newell and Simon once claimed (famously) that a suitably organized 'physical symbol system has the necessary and sufficient means for general intelligent action'.[12] A physical symbol system (henceforth PSS) is (roughly) a materially instantiated, automatic compositional system. More precisely, it is a material system in which atomic symbols are automatically combined and manipulated by structure-sensitive processes, according to the principles of a compositional syntax and semantics. Although Newell and Simon adopted what we might call an unrestricted form of this hypothesis (i.e., all cognition is the result of a suitably organized PSS), one might reasonably adopt a more restricted version. For example, let's proceed – as many classical computational psychologists focussed on the human mind manifestly have – by holding that a suitably organized PSS has the sufficient means for certain high-end cognitive achievements. I suggest that Bechtel's distributed architecture of an inner processing network coupled to an external symbol system qualifies as an extended PSS. Of course, more would need to be said to drive home this idea[13], but let's assume that any concerns can be met. What we are pursuing is a benchmarked, fine-grained approach to cognitive kinds, since we are concentrating on the sorts of fine-grained causal-explanatory properties (those patterns of combinatorial symbol structure and coupled network dynamics specific to linguistic behaviour, natural deduction or mathematical reasoning) that an area of established human cognitive psychology (classical computational psychology) takes to be theoretically important. And yet, contra Rupert, the approach delivers extended cognitive kinds, in that some of the kind-constituting elements (the symbol structures) are externally located. By NKA, then, the Bechtel architecture is not only an extended PSS; it is also an extended cognitive system. With this, the first horn of the Rupert dilemma, and thus the dilemma itself, is neutralized. Resisting the temptation to retire early, however, let's see what the prospects are for nullifying the second horn too.

4. Cognitive Kinds, Martian Kinds and the Sprevak Dilemma

The second horn of the Rupert Dilemma would be defused, if there are benchmarked, coarse-grained cognitive kinds with extended instances. Recall that Rupert endeavours to block this thought by arguing that the coarse-grained structures on offer fail to count as natural kinds, because they fail his family-resemblance test. As Rupert explains: 'If, for example, a Martian exhibits memory-related behavior, but that behavior is produced by a collection of *very* different mechanisms from the ones that produce memory-related behavior in humans (and the Martian process is not amenable to tweak-and-extend modeling), then the Martian behavior is not produced by memories, at least not if we want to use 'memory' as a natural-kind term, rather than, say, as a merely organizational term'.[14]

This sets Rupert against a view that is expressed (although not thereby endorsed) by Sprevak as the *Martian intuition*. The Martian intuition is a multiple-realization principle that Sprevak takes to be at work in standard functionalist theorizing about the mind.

> [I]t is possible for creatures with mental states to exist even if such creatures have a different physical and biological makeup to ourselves. An intelligent organism might have green slime instead of neurons, it might be made out of silicon rather than carbon, it might have different kinds of connections in its "nervous" system… The Martian intuition applies to fine-grained psychology as well as physiology: there is no reason why a Martian should have exactly the same fine-grained psychology as ours. A Martian's pain response may not decay in exactly the same way as ours, its learning profiles and reaction times may not exactly match ours, the typical causes and effects of its mental states may not be exactly the same as ours, even the large-scale functional relationships between the Martian's cognitive systems (e.g. between its memory and perception) may not exactly match ours.[15]

If we interpret the Martian intuition in terms of cognitive kinds, the moral for us is that the grain of our cognitive kinds needs to be set sufficiently coarsely so that Martian and human thinkers may share cognitive kinds, even if, at the fine-grained functional level, collections of very different mechanisms are operative. For reasons explained earlier, this sort of coarse-grained approach to cognitive kinds delivers ExC. More specifically, in the present context, the Martian intuition is supposed to strengthen NKA, because its independent plausibility adds weight against Rupert's apparently ExC-hostile account of what constitutes a genuine coarse-grained natural kind. Unfortunately, by using the Martian intuition to tip the scales in her favour, the proponent of NKA runs into trouble. For the Martian intuition helps to generate a second putative dilemma for ExC, a dilemma that has been lodged in the literature by Sprevak. Let's call this the *Sprevak dilemma*.[16]

To bring the Sprevak dilemma into view, we need one more idea that is arguably at the heart of ExC, the so-called *parity principle*.[17] The parity principle asks us to consider an actual system that generates some psychologically interesting outcome and whose operation involves an important functional contribution from certain externally located elements. It then encourages us to imagine a hypothetical scenario in which exactly the same functional contribution, to an equivalent outcome, is made by certain internally located elements. Having taken this imaginative step, if we then judge that the internal realizing elements in the hypothetical case count as bona fide parts of a genuinely cognitive system, we ought to conclude that the very same status (i.e., cognitive status) should be granted to the external realizing elements in

the actual, environment-involving case. To do otherwise would be to succumb to neural chauvinism.

For the sake of argument, let's agree that the Martian intuition and the parity principle are indeed keystones of ExC. Given this, Sprevak argues as follows. Take an example of an externally located element that intuitively looks to be a wildly unlikely candidate for cognitive status. Now imagine a functionally equivalent element located inside the head of a Martian. According to the Martian intuition, we should grant that inner Martian analogue cognitive status. But then, according to the parity principle, we must also grant cognitive status to the externally located element identified at step one, the element, that is, that was offered as a wildly unlikely candidate for any such status. Thus, Sprevak concludes, ExC as characterized is unavoidably and wildly over-permissive with respect to what counts as cognitive, a fact which gives us good reason to reject the view. Of course, the advocate of ExC could avoid this problem of over-permissiveness, *if* she could either give up the Martian intuition (and so not count the inner Martian analogue as cognitive, thereby blocking the application of the parity principle) or give up the parity principle (and so not count the external element as cognitive, even if the internal Martian analogue does count); but (we have agreed) the Martian intuition and the parity principle are keystones of ExC, so she can give up neither. That's the Sprevak dilemma.

To see what has gone wrong[18], we need to focus on how the above argument unfolds in the case of one of Sprevak's flagship examples of an external factor to which the granting of cognitive status would be excessively permissive, namely a so-far-unused, not-even-known-about desktop computer program that calculates the dates of the Mayan calendar five-thousand years into the future. Sprevak invites us to consider a functionally equivalent program inside a Martian head. At this point the Martian intuition is supposed to kick in and the latter program qualifies for cognitive status. Since, by hypothesis, the external program and the in-the-head Martian program are functionally equivalent, the application of the parity principle then drives us into the arms of the troublesome conclusion that the desktop program is part of the computer user's extended mind.

How should the proponent of ExC respond? One promising move would be to deny that the Martian intuition requires the in-the-Martian-head program to be awarded cognitive status. This would block the subsequent, parity-based step in the argument. This looks to be eminently achievable. After all, the desktop program is presumably supposed to be an isolated and removable application, one that the human user would have to somehow find and learn how to use. It's plausibly this lack of any functional integration into an organized economy of existing cognitive states and processes that makes us want to resist the claim that the desktop program has cognitive status. But surely if we put a program like *that* in a Martian head, then

we continue to have warrant to deny it cognitive status on precisely the same grounds, namely that it's insufficiently functionally integrated into the Martian's cognitive economy. Of course, the implication of this move is that we may well all carry around, in our heads, neurally realized structures that don't count as parts of our mental machinery, because those elements fail to meet the functional integration condition. But, unless one holds that merely being inside a head is sufficient for an element to enjoy cognitive status, whatever else may be true of that element, that idea seems perfectly innocuous. For example, the glial cells in our brains perform various incontestably non-cognitive tasks, such as holding neurons in place and supplying them with nutrients and oxygen. Furthermore, the idea of non-cognitive inner elements seems to be in harmony with the fair treatment ethos of the parity principle (which denies that spatial location is a relevant factor in determining cognitive status) and the Martian intuition (since where there is a lack of functional integration on the scale of the inner or outer Mayan calendar program, we are beyond talk of large-scale functional differences). With this, the Sprevak dilemma is neutralized.

It looks as if the only remaining challenge for the fan of NKA is to provide an explanation of how it is possible to secure genuine coarse-grained cognitive kinds with extended instances, while remaining mindful of Rupert's worry about merely organizational terms. Here is a proposal. Earlier in this chapter, I argued that, if we accept that being a PSS is sufficient for certain cognitive capacities, then the first horn of the Rupert Dilemma may be defused, because capacity-specific distributed arrangements of inner connectionist networks and external compositional symbol systems, coupled via sensorimotor control, provide us with examples of extended, benchmarked, fine-grained cognitive kinds. This way of salvaging NKA was purchased by highlighting the fine-grained properties and dynamics of such systems at a capacity-specific level. However, if we adjust our metaphysical spectacles, it is plausible that each of these capacity-specific arrangements counts as an instance of a bona fide *coarse-grained* cognitive kind, namely a PSS. Any temptation to think of 'PSS' as a 'merely organizational term' is surely misplaced, in virtue of the fact that the different capacity-specific arrangements in question will presumably exhibit the right sort of family resemblances to satisfy Rupert, namely overlaps between the causal-explanatory elements (structure-sensitive syntactic rules, compositional symbol structures) that figure in the different instances of PSSs which contribute to different high-end cognitive achievements. If this is right, then there are benchmarked, coarse-grained cognitive kinds with extended instances – namely extended PSSs – and the second horn of the Rupert dilemma is disarmed. Thus we arrive at a situation in which the proponent of ExC can comfortably occupy either horn of the Rupert dilemma. And that's another way of saying that there is no such dilemma after all.

5. Conclusion

Faced with the Rupert dilemma and (in a supporting role) the Sprevak dilemma, it might have seemed as if this is the worst of times to be arguing for extended cognition by way of the thought that cognitive kinds – the causal-explanatory natural kinds that figure in the practice of our best cognitive science – have extended instances. But, as I have argued, it is in truth the best of times for this argument, because paying close attention to the relevant scientific practice reveals it to be hardily resistant to both dilemmas. We are in the epoch of the extended mind.[19]

---

[1] The maiden voyage of ExC was in A. Clark and D. Chalmers , 'The Extended Mind', *Analysis*, 58: 1 (1998), pp .7-19; see also A. Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (New York: Oxford University Press, 2008). For a field-defining collection, see R. Menary (ed), *The Extended Mind* (Cambridge, Mass.: MIT Press, 2010). Throughout this chapter, I will use the terms 'mind' and 'cognition' interchangeably. This strikes me as standard practice in cognitive science. From this perspective, 'extended cognition' and 'the extended mind' are alternative names for the same view.

[2] R. Rupert, 'Memory, Natural Kinds, and Cognitive Extension; or, Martians don't Remember, and Cognitive Science is not about Cognition', *Review of Philosophy and Psychology*, 4: 1 (2013), pp. 25-47, p.28, note 2.

[3] Clark and Chalmers, 'The Extended Mind', p.14.

[4] R. Rupert, *Cognitive Systems and the Extended Mind* (Oxford: Oxford University Press, 2009); R. Rupert, 'Memory, Natural Kinds, and Cognitive Extension'.

[5] See, for example, M. Wheeler, 'In Search of Clarity about Parity', *Philosophical Studies*, 152: 3 (2011), pp.417-425, and the response in A. Clark, 'Finding the Mind', *Philosophical Studies* 152: 3 (2011), pp.447-461.

[6] Rupert, 'Memory, Natural Kinds, and Cognitive Extension', p.29.

[7] Rupert, 'Memory, Natural Kinds, and Cognitive Extension'. An earlier presentation of the dilemma appears in Rupert, *Cognitive Systems and the Extended Mind*. My treatment draws on both pieces, but mostly on the former.

[8] M. Wheeler, 'In Defense of Extended Functionalism', in R. Menary (ed), *The Extended Mind*, pp.245-270.

[9] W. Bechtel, 'Natural Deduction in Connectionist Systems', *Synthese*, 101 (1994), pp.433-463; W. Bechtel, 'What Knowledge must be in the Head in order to Acquire Language', in B. Velichkovsky and D. M. Rumbaugh (eds), *Communicating Meaning: the Evolution and Development of Language* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1996).

[10] Bechtel, 'Natural Deduction in Connectionist Systems', p.436.

[11] See M. Wheeler, 'Embodied Cognition and the Extended Mind', in J. Garvey (ed), *The Continuum Companion to Philosophy of Mind* (London: Continuum, 2011), pp.220-38; M. Wheeler, 'Revolution, Reform, or Business as Usual? The Future Prospects for

Embodied Cognition', in L. Shapiro (ed), *The Routledge Handbook of Embodied Cognition* (London and New York: Routledge, 2014), pp.373-383

[12] A. Newell and H. A. Simon, 'Computer Science as Empirical Inquiry: Symbols and Search', *Communications of the Association for Computing Machinery*, 19: 3 (1976), pp.113-126, p.116.

[13] For further discussion, see Wheeler, 'Embodied Cognition and the Extended Mind'.

[14] Rupert, 'Memory, Natural Kinds, and Cognitive Extension', p.40.

[15] M. Sprevak, 'Extended Cognition and Functionalism', *The Journal of Philosophy*, 106: 9 (2009), pp.503-527.

[16] Sprevak, 'Extended Cognition and Functionalism', pp.507-8.

[17] Clark and Chalmers, 'The Extended Mind', p.8.

[18] The argument I am about to present is a streamlined and more focussed version of one that appears in Wheeler, 'In Defense of Extended Functionalism'.

[19] For useful critical discussion, many thanks to Rob Rupert and to audiences at Boulder, Stirling and Sussex. Some passages in sections 2, 3 and 4 of this paper were adapted from passages in Wheeler, 'In Defense of Extended Functionalism' and Wheeler, 'Revolution, Reform, or Business as Usual?'.