**From:**

**Michael Wheeler**

# *Extended X: Recarving the Biological and Cognitive Joints of Nature*

## Draft Book Manuscript

## NB: Please do not quote or cite without permission

## Chapter 2 The Extended Cognition Hypothesis

I sometimes find, and I am sure you know the feeling, that I simply have too many thoughts and memories crammed into my mind.... At these times... I use the Pensieve. One simply siphons the excess thoughts from one's mind, pours them into the basin, and examines them at one's leisure.
Albus Dumbledore, in J.K. Rowling's *Harry Potter and the Goblet of Fire*

There is nothing specially oomphy about causal relations inside the skin, or inside the head, nothing specially capable of pushing and shoving. So there is nothing causally mysterious or inhospitable to materialism or naturalism or realism about relational states of persons. And there is no magical causal boundary around persons. Viewed subpersonally, they are in principle transparent to causality.
Susan Hurley, *Consciousness in Action*[1]

### 2.1 Magic and Oomph

The first extension view that will concern us trades under a variety of different names, including the extended mind (Clark and Chalmers 1998), active externalism (Clark and Chalmers 1998), vehicle externalism (Hurley 1998; Rowlands 2003), environmentalism (Rowlands 1999), and locational externalism (Wilson 2004). Beneath all this terminological diversity, however, lies what I take to be a shared and unifying thesis, which I shall label the *extended cognition hypothesis*, henceforth *ExC*. As I see it, ExC comes in three basic flavours, two modal (concerning the *possibility* of extended cognition) and one non-modal (concerning the *fact* of extended cognition in the here and now). Take the term *cognitive traits* to range over cognitive states, cognitive processes, cognitive mechanisms, cognitive routines and cognitive systems, as

well as the cognitive architecture as a whole. The three basic versions of ExC can then be stated as follows:

*General Modal ExC: it is possible for there to be a kind of thinker such that at least some of the cognitive traits of an individual thinker of that kind may sometimes be spatially located partly outside that thinker's skin (or whatever the corresponding boundary might be).*

*Human Modal ExC: it is possible for at least some of the cognitive traits of an ordinary individual human thinker sometimes to be spatially located partly outside her skin.*

*Human Non-Modal ExC: at least some of the cognitive traits of an ordinary individual human thinker are sometimes spatially located partly outside her skin.*

The dependencies here should be clear. If the human non-modal version is true, then so is the human modal version. If the human modal version is true, then so is the general modal version. However, the general modal version could be true without the human modal version being true. And the human modal version could be true without the human non-modal version being true. Having in mind the three different versions of ExC and the pattern of dependencies between them will occasionally help us to see what is at stake in a particular argument for or against cognitive extension. That said, I shall mostly write of ExC without further qualification. When I do that I am signaling a view that holds all three versions of the hypothesis to be true.

Certain other writers in this area (e.g. Rupert 2004; Adams and Aizawa 2008) have analyzed ExC in terms of a single modal claim (roughly, that extended cognition is possible) and a non-modal claim (roughly, that extended cognition is a fact of human life), with the fans of ExC depicted as accepting both claims and its critics standardly depicted as accepting the modal claim but denying the non-modal one.[2] As already intimated, however, there is some utility in separating out the two different modal versions of ExC, in order to distinguish claims about what might be true of some possible thinker from claims about what might be possible for an ordinary human thinker. This remains the case, I think, even if the resulting distinction is sometimes fuzzy and hard to police (see chapter 3). When Andy Clark (forthcoming a, p.3) claims that "in imaginable circumstances, ones that [involve] *no giant leaps of technology or technique*, we would be justified in holding that certain mental and cognitive states [extend]… into the non-biological world" and proposes this as an alternative to "a mere claim of 'logical possibility'" about cognitive extension, I take it that he is in the ballpark of the human modal version of ExC.

Precisely how do factors located beyond the thinker's skin get to have cognitive – and not merely causal – oomph, without there being any whiff of Dumbledore-like magic? Answering this question is far from straightforward, because unfortunately ExC is home to a tangled web of implicit assumptions, tricky complications and interpretative confusions. So the present, short chapter is dedicated to getting clear about precisely what ExC actually says. There will also be a smattering of signposts and promissory notes indicating where our future reflections will be taking us. In chapters 3, 4 and 5 we shall explore how one might plausibly argue for the view.

**2.2 The Heart of the Matter**

It is important to remind ourselves that ExC is a view about the whereabouts of mind that is distinct not only from the position adopted by orthodox cognitive science, but also from the position adopted by any merely embodied-embedded view. We can adapt an example originally due to Rumelhart et al. (1986), in order to retrieve the relevant conceptual space. 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 ExC theorist considers the coupled combination of pen-and-paper resource, appropriate bodily manipulations, and in-the-head processing to be a cognitive system in its own right. In a terminology introduced in chapter 1, for orthodox and embodied-embedded theorists alike, the pen and paper system is a non-cognitive part of a cognition-enabling system that also involves certain bodily movements and the brain. The cognitive enabling system, on the other hand, is restricted to just the brain (the orthodox view) or the brain and body (the embodied-

embedded view). For ExC theorists, however, the pen and paper system is itself a proper part of a cognitive enabling system that also involves the brain and the non-neural body. To put this yet another way, where cognition is embodied and embedded we have a case of explanatory spread with respect to orthodox cognitive theory; where cognition is extended we have a case not only of such explanatory spread, but also of ontological spread involving the category of the cognitive.

It should be clear from the foregoing remarks that what is on offer from ExC is not some insane vision of cognitive traits as being realized in everyday artefacts independently of the organic brain, or alternatively of some other system whose contribution to cognitive activity is similar to that made by the organic brain. The pen and paper system that enables long multiplication counts as part of a cognitive process because, and just when, it plays the right kind of role within a system of relations and interactions that includes various neural processes and bodily movements. (Of course, one needs to say precisely what the various contributions and roles that matter here are. More on such things as we go along.) To give another example, recall Dumbledore's pensieve. This device for storing thoughts and experiences (as described in one of the quotations at the beginning of this chapter) is not the whole of Dumbeldore's memory system, although if ExC is right it may be part of it. Assuming for the sake of argument a surely-too-crude 'information warehouse' conception of memory, one might explain this point by saying that being able to remember does not consist solely of some storage facility. It consists additionally of processes that place information in, and recover information from, that facility. (Here one might note that someone whom we describe as having lost their memory might have lost the data in question, but alternatively they might have lost the capacity to retrieve that data and/or the capacity to lay down new data.) In Dumbledore's case the pensieve may be conceived as a genuine memory storage facility. But the processes of placement and retrieval centrally involve his brain, as well as his wand and (therefore) certain bodily manipulations. His memory is thus *extended not (wholly) external*. Dumbledore's pensieve doesn't, on its own, remember anything. Similarly the pen and paper don't, on their own, think mathematical thoughts.[3]

This way of heading off a woefully inaccurate interpretation of ExC helps to make the point that although ExC theorists reject the seductive idea that cognition is organism-*bound* or, even more narrowly, brain-bound, that fact, in and of itself, does not prevent those same theorists from taking cognition to be organism-*centred*, and perhaps even brain-centred (cf. Clark 2007). Susan Hurley's account of cognitive extension in terms of dynamic singularities has this flavour. Thus she writes of "a tangle of multiple feedback loops of

varying orbits, some internal to the brain or body, others involving partly external motor feedback. Such a dynamic singularity is centred on the organism and moves around with it, but it does not have sharp boundaries" (Hurley 2003, p.9 manuscript). So ExC is compatible with organism-centredness. It is worth noting, however, that some ExC theorists – predominantly those who identify themselves as complementarity theorists or integrationists – endorse a view that has the effect of holding this organism-centredness in check. Such thinkers argue that the way in which external factors depend on brains to (as one might say) complete a cognitive trait is essentially equivalent to the way in which brains often depend on external factors to do the same. For example, John Sutton (2006, p.3 manuscript) notes, with a nod to Clark (2003), that external cultural tools, artefacts and symbols systems "(at present, in general) do no cognitive work on their own… but then neither do brains tend to do their cognitive work in isolation, because essentially incomplete creatures like us naturally use our neural resources in part to parasitize, lean on, and incorporate those external cultural-technological resources which have become apt for incorporating". The complementarity-integrationist view will concern us directly later in our investigation. For the moment it is worth noting that wherever one falls on the issue of organism-centredness, the shift from an orthodox cognitive framework to an extended one is unlikely to leave our account of what the brain is doing in psychological processing untouched. That is an issue to which we shall return in chapters 8 and 9.

### 2.3 Boundary Disputes

A further important clarification of ExC concerns its relationship with the internal-external distinction. To be clear, the core message of ExC is *not* that there are conditions under which the very idea of any internal-external boundary becomes problematic or misleading. Rather, what ExC suggests is that one perfectly reasonable way of drawing that boundary may become disassociated from various others. To bring this idea into view, let's begin with the following characterization of ExC: there are conditions under which something that counts as a single cognitive trait contains some elements that are internal and some that are external. The cognitive trait in question is thus *extended over* a still-in-place internal-external boundary. This demonstrates that although, for the ExC theorist, there is a clear sense in which the internal-external boundary is irrelevant when it comes to determining the spatial limits of a cognitive trait, that boundary has not thereby disappeared from the account altogether. But now notice something interesting. The analysis we have just given makes it a presupposition of ExC that the location of the internal-external boundary is not to be fixed by the limits of the mind. After

all, on this way of setting things up, the mind (as identified by the extended cognitive trait – more on this in a moment) is to be found *on both sides* of that boundary.

It seems to me that ExC theorists are not always as clear or consistent as they might be when it come to saying where, exactly, the internal-external boundary is to be located. Somewhat imprecise talk of, for example, 'the skull and skin boundary' is common. So what are the options? One might fix the internal-external boundary at the limit of the brain or the skull, so that the spinal cord and the pathways between the sensory receptors and the brain already count as external. Another option would be the limit of the nervous system as a whole, so that the muscles and the skin already count as external. And yet another would be the skin itself, so that it's only the beyond-the-skin physical, biological and social environments that count as external. No doubt there are other possibilities, but these are presumably the main candidates. In the present treatment I am going to reserve the idea of cognitive extension for cases in which cognitive traits are partly located beyond the skin, and not merely beyond the brain, skull or nervous system. Here's why.

Say it turned out that, according to our best science and philosophy, cognitive traits are sometimes extended into the beyond-the-skin environment. Such extension would no doubt typically involve bodily movements and/or manipulations. It seems to me that, in this situation, we would be strongly inclined to conceptualize the relevant contributions from the non-neural body within an ExC framework. But now say it turned out that the properly cognitive part of the behaviour-generating system stopped at the skin. In this second situation it seems to me that we would be far more likely to see the cognition in question as merely embodied and embedded, rather than as extended in any interesting sense. So although the overall truth of ExC does not of course require that *every* cognitive trait that involves, as a proper part, the non-neural body also involves, as a proper part, the beyond-the-skin environment, the fact remains that unless there are clearly established cases of cognitive extension in the environment-involving sense, the conclusion that the non-neural body makes a genuinely cognitive contribution to behaviour would support only a merely embodied-embedded picture. What this tells us is that ExC is not justified directly by the conclusion that, for example, bodily gestures are not mere side-effects of, or aids to, thought, but in fact are proper parts of thinking, as Clark (2007) seems to think it is. Of course, one might see this as just another strategy for marking out the distinction between merely embodied-embedded cognition and extended cognition, but it is a strategy that has the following beneficial pay-off. It suggests that the key cases of ExC – those that, given the controversial character of the view, its supporters need to establish – are those in which cognitive traits are partly located beyond the

skin. And for the moment at least, that justifies the methodological restriction of the position to such cases.[4]

As it happens, the idea of an internal-external boundary fixed by the skin is also conceptually necessary, if ExC is to be given its standard formulation as a form of *externalism* in the philosophy of mind. To see why, it will be useful to walk the well-worn path of distinguishing externalism ExC-style from what is probably the most commonly encountered species of philosophical externalism, namely content externalism (externalism about mental content). In philosophical circles the idea of a mental content is standardly cashed out in terms of the semantic content carried by the proposition that follows the 'that' in an intentional state (propositional attitude) attribution such as 'John believes that water is wet'. Then, roughly speaking (and without all the fun of the thought experiments that standardly motivate it, e.g. Putnam 1975), context externalism may be stated as the view that the internal (within-the-skin) properties of a thinker are, at least sometimes, insufficient to individuate the mental contents that figure in that thinker's intentional states. Given the common assumption that the contents of propositional attitudes are essential to them, this yields the striking conclusion that two thinkers who are internally identical in every way may nevertheless possess different intentional states merely through being located in different environments.

Content externalism, then, is a claim about the spatial location of the factors to which one needs to appeal in order to individuate certain agent-level mental states. ExC, by contrast, is a claim about the spatial location of the *vehicles of cognition*. As Rowlands (2006, p.32) explains, the vehicles of cognition are "the subpersonal mechanisms that allow cognitive processes to be run", although I'm inclined to put the idea more generally, by saying that the vehicles of cognition are *the subagential elements that allow cognitive traits to be realized*. (For an explanation of the agential-subagential distinction, as contrasted with the personal-subpersonal distinction, see chapter 1. For essentially equivalent, or at least complementary, ways of distinguishing ExC from content externalism, see Hurley 1998, forthcoming; Rowlands, 2003; Rupert 2004; Clark 2005.) With the idea of a vehicle of cognition on board, the radical shift that ExC invites us to endorse may be characterized as a shift from thinking of the subagential vehicles of agent-level cognition as being exclusively neural in character, or perhaps as consisting of neural factors and elements of the non-neural body, to thinking of the subagential vehicles of agent-level cognition as being *ecological* in character - that is, as spanning the brain, the non-neural body *and* the environment. And that's why ExC sometimes trades under the title of *vehicle externalism* (Hurley 1998; Rowlands 2003).[5]

By concluding that the internal-external boundary should be fixed by the skin, we have made important headway in our attempt to understand ExC. But now notice that it is equally common in philosophy and science to find the internal-external boundary positioned at the interface between the mind (the internal side) and the world in which the mind is embedded (the external side). In most thinking about thinking this boundary coincides in practice with either the limits of the nervous system or the boundary of the skin. In cases of cognitive extension, however, the mind-world boundary is shifted to the spatial limits of a system that includes certain beyond-the-skin elements. Of course, even where the vehicles of cognition are extended it remains the case that there are environmental elements external to the now-expanded mind, so from this perspective too the internal-external boundary remains in force. It has been shifted, not dissolved.

At this point one might wonder just what permits me to glide effortlessly from talk of ecological vehicles of cognition to talk of a repositioned interface between *mind* and world. Just why should the truth of vehicle externalism establish the existence of extended minds and (looking forward to later concerns) extended selves? Essentially the same question might be put as follows, in order to interrogate a particular aspect of the analysis being developed here: given the vehicular basis of ExC, why is it that, in our opening tripartite analysis of ExC into modal and non-modal hypotheses, the term 'cognitive traits' should be given both a subagential and an agential interpretation? It seems to me that if one is going to locate phenomena such as minds and selves in space at all, then it is plausible, *in the absence of good arguments to the contrary*, that one's agent-level cognitive traits, one's mind, and one's self are ultimately located where one's subagential cognitive machinery is.[6] Of course, such arguments to the contrary might be developed. For example, if one thinks (i) that there can be unconscious cognitive states and processes (I take it that this is the modern orthodoxy), (ii) that any state or process properly called mental must be conscious or potentially so (perhaps on the basis of something like Searle's (1992) connection principle), and (iii) that consciousness is always brain-bound, then one will have the conceptual space for cognitive extension without an extended mind. The burden of proof, however, remains with those who want to endorse ExC while imprisoning the mind and/or the self within the skin. We shall return to this issue briefly in chapter 8.

Our earlier formulations of ExC, plus everything that we have since learned about ExC in this clarificationary chapter, confirm that it is a hypothesis about the spatial location of the cognitive traits of an *individual* thinker. So although nothing in the basic idea of ExC rules out the possibility that my cognitive traits may be partly realized inside the skin of another thinker (e.g. regions of

your brain may realize part of my memory or decision-making systems), what is excluded is the idea of a group mind, where that mind is not owned by any individual thinker. As mentioned in chapter 1, the category of distributed cognition is broad enough to cover both ExC and the group mind hypothesis, and in some treatments developed explicitly under the distributed cognition banner (e.g. Hutchins 1995), the two narrower views appear alongside each other without always being clearly demarcated. I shall return to social extension ExC style and to the group mind hypothesis in later chapters.

That completes our opening round of clarifications regarding the basic shape of ExC. The next chapter turns to the central issue. How might one argue for the view?

**Notes**

1. Hurley, 1998, p.336.

2. In my view this is the correct interpretation of Adams and Aizawa's position. However, I should note two wrinkles in how they play it out. First, part of their response to the modal claim is to challenge the account that ExC theorists give of the conditions under which extended cognition would be actualized, which means that their acceptance of the modal claim still leaves room for debate with ExC theorists at an in-principle level (see Adams and Aizawa 2007, p.27). We shall engage with this in-principle debate in chapters 3 and 4. Secondly, Adams and Aizawa sometimes suggest that extended cognition might be an actual fact of human experience, but not one that is "a common or pervasive feature of our everyday world" (ibid. p.26). If we read 'some' in my formulation of the human non-modal version of ExC as 'at least one', and we read 'sometimes' as 'at least once', then it would in fact be true to say that Adams and Aizawa endorse the non-modal claim. However, since they also suggest that the conditions that establish the non-modal claim as being true are "the exotic conditions that philosophers entertain in thought experiments" (ibid. p.26), one must wonder just when and where Adams and Aizawa expect to find those conditions actualized. Because of this I think it is more illuminating to read them simply as rejecting the non-modal claim.

3. This 'extended not external' point is hardly news to ExC theorists (see e.g. Clark, forthcoming a, pp.3-4), but it bears emphasis because some of those same theorists have occasionally settled on descriptions of cognitive extension cases that invite the problematic reading. Thus when describing the imaginary characters Inga (whose on-board memory systems are normal) and Otto (an mild Alzheimer's sufferer whose notebook acts rather like

Dumbledore's pensieve), their creators write that "the notebook plays for Otto the same role that memory plays for Inga" (Clark and Chalmers 1998, p.13) and, later, that the "notebook is a constant for Otto, in the same way that memory is a constant for Inga" (ibid. pp.13-14). We shall meet Inga and Otto again.

4. A further consequence of the restriction of ExC to cases of beyond-the-skin cognitive extension concerns the relationship between ExC and the currently much discussed *enactive approach to cognition*. If one understands the enactive approach as being based on the claim that "cognition is the exercise of skillful know-how in situated and embodied action" (Thompson, 2007, p.13) or the idea that "perception… is not a process in the brain, but a kind of skillful activity on the part of the animal as a whole" (Noë, 2004, p.2), then, on the face of things, enactivism and ExC are not necessarily theoretical bedfellows. Indeed, it looks as if enactivism might support a kind of principled body-centrism (Clark 2008). The relationship between enactivism and ExC will be explored in chapter 8.

5. Does ExC entails content-externalism? Does content externalism entail ExC? These are questions which, for today at least, I shall leave unanswered, aside from the following vague thoughts. Even though most content externalists seem to want to remain vehicle internalists, it may be possible to construct an argument from content externalism to vehicle externalism by exploiting the context-externalist claim that two internally identical thinkers may nonetheless realize different mental states due to variations in the environment alongside a claim about the supervenience of the mental on the physical. In any case, it seems that it is a tougher call to be a vehicle externalist and a content internalist. (For further discussion, see Wilson, 2004, p.179; Hurley, forthcoming).

6. My inclusion of the qualifier 'ultimately' gestures at the way in which my position diverges from that adopted by Clark (2008b, chapter 6, p.13, manuscript). Clark suggests that "we don't find or individuate human agents by first finding their cognitive mechanisms", but rather by locating a "reliable, easily identifiable physical nexus of perception and action, apparently driven by a persisting and modestly integrated body of goals and knowledge". It seems to me, however, that if by 'find' one means the result of a rough and ready, pretheoretical method for locating agents in everyday life, and by 'individuate' one means the result of a scientifically and philosophically informed method for saying where the boundaries around agents finally lie, then the 'find' and 'individuate' aspects of our practices may come apart. Thus I might agree with Clark about finding agents but disagree with him about individuating them.