Drayson Z (2010) Extended cognition and the metaphysics of mind, *Cognitive Systems Research*, 11 (4), pp. 367-377.

**This is the peer reviewed version of this article**

Zoe Drayson
December 2009

# Extended cognition and the metaphysics of mind

## 0. Introduction

Advocates of the 'extended mind' – the claim that cognitive processes can and do extend outside the head – have generally had little to say on the metaphysics of mind, preferring to concentrate on the explanatory role which extended cognition can play in empirical cognitive science research. Recently, however, claims have been made about the relationship between extended cognition and traditional functionalism in the philosophy of mind. In this paper I explore these claims and suggest a way of clarifying the debate.

Clark and Chalmers (1998) put forward the original case for the extension of cognitive processes beyond the skin and skull. While they do not explicitly present their position as a development of the functionalist program in philosophy of mind, the links between the two are often remarked upon. Rupert (2004), Shapiro (2008), and Weiskopf (2008), for example, all claim that the argument for the extended mind relies to a certain extent upon functionalism.

Sprevak (forthcoming), however, makes a much stronger claim regarding the relationship between extended cognition and functionalism. Sprevak argues not only that the proponents of the extended mind rely on functionalist principles, but also that functionalism actually *entails* extended cognition – and a version of extended cognition which is sufficiently radical as to be obviously false. In this paper I introduce the standard argument for extended cognition, and the principles and examples which are

invoked to support it (section 1). I then review the basic tenets of functionalism and the intuitions behind it (section 2), before setting out Sprevak's claim that functionalism and extended cognition are more closely related than previously thought (section 3). I focus first on Sprevak's argument that the version of the extended mind entailed by functionalism is so radical as to be false (section 4), and I propose two ways one could defend a more moderate version of the extended mind (sections 5 and 6). I shall follow convention by using the terms 'extended cognition' and 'the extended mind' interchangeably throughout, until I explore the nature of the relationship between functionalism and the more moderate version of extended cognition/the extended mind (section 7). Here I will suggest that the original Clark and Chalmers argument considered by Sprevak is actually two separate arguments: one for extended cognition, and another for the extended mind. I will argue that the two positions bear different relationships to functionalism. Finally, I suggest that even if Sprevak's argument were correct, and he'd provided a *reductio* of functionalism, the proponent of extended cognition need not worry, as their argument does not rely on the truth of functionalism.

## 1. Extended cognition

The hypothesis of extended cognition (hereafter HEC) was introduced by Clark and Chalmers (1998) as the claim that cognitive processes can and do extend outside the head.

Clark and Chalmers ask the reader to consider the following three scenarios, in which a person watches two-dimensional 'Tetris-like' geometrical shapes on a computer screen.

> (1) The person is asked to answer questions about the potential fit of the shapes into depicted 'sockets'. The shapes do not move: the person must mentally rotate the shapes to assess their fit.
>
> (2) As above, but this time the person can choose to press a button to physically rotate the shape on screen (which is feasibly faster), or to mentally rotate the shape as before.
>
> (3) Described as 'sometime in the cyberpunk future', the person has a neural implant which can rotate the shape as quickly as the computer in (2). They can choose to assess the fit using either the neural implant or standard mental rotation.

Clark and Chalmers want the reader to consider how much *cognition* is present in each case, and suggest that all three are similar. In particular, if we're willing to think of the neural implant in (3) as cognitive, why not in (2), given that it displays the same sort of computational structure? As long as we're not already committed to the *a priori* claim that cognition can only take place internal to the skin/skull boundary, it seems difficult to argue that the external processes in (2) are somehow not cognitive in the way that the mental rotation in (1) and the neurally-implemented rotation in (3) are.

.

Clark and Chalmers suggest the argument works not just for cognitive processes but also for more traditional mental states:

> "In particular, we will argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes. If so, the mind extends into the world." (Clark and Chalmers 1998)

The well-known example introduced at this point involves the character of Otto and his notebook. First, Clark and Chalmers ask us to consider Inga, who retrieves her beliefs from her memory in a normal way; and Otto, an Alzheimer's patient who relies on environmental props to structure his life. Inga and Otto are then compared in the following thought experiment:

> (I) Inga hears from a friend that there is an exhibition at the Museum of Modern Art, and decides to go see it. She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. It seems clear that Inga believes that the museum is on 53rd Street, and that she believed this even before she consulted her memory. It was not previously an *occurrent* belief, but then neither are most of our beliefs. The belief was sitting somewhere in memory, waiting to be accessed.
>
> (O) Otto suffers from Alzheimer's disease, and like many Alzheimer's patients, he relies on information in the environment to help structure his life. Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down. When he needs some old information, he looks it up. For Otto, his notebook plays the role usually played by a biological memory. Today, Otto hears about the exhibition at the Museum of Modern Art, and decides to go see it. He consults the notebook, which says that the museum is on 53rd Street, so he walks to 53rd Street and goes into the museum.
>
> (Clark and Chalmers 1998)

Clark and Chalmers argue that Otto's case and Inga's case are analogous, because the explanatory role played by Otto's notebook matches that played by Inga's memory: the content of Otto's notebook plays the same function as Inga's dispositional belief. The mere fact that Inga's belief was inside her skull while Otto's was outside his skull does

not seem to make any relevant difference to the explanatory role (or 'functional poise') of his mental state.

Both the Tetris example and the example of Otto and his notebook rely on the assumption that we're not already biased in favour of the inner when making judgements about cognition. This assumption has become known as the parity principle (hereafter PP) and is stated by Clark and Chalmers as follows:

> "If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process." (Clark and Chalmers 1998)

PP is seen by many as being a broadly functionalist principle, in that it emphasises causal role over the details of material implementation.

## 2. Functionalism

Functionalism is the position in the philosophy of mind that what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the functional role it plays in the system of which it is a part (Levin 2009). A particular type of mental state could therefore be realized by multiple different types of physical state. One of the motivations for functionalism is that it can preserve the 'Martian intuition' (hereafter MI): the idea that a creature could have the same types of mental states as us despite major physical differences between them and us. (It should be

noted that not all varieties of functionalism aim to preserve MI: I shall return to this matter in the section 7.)

Although Clark and Chalmers (1998) do not mention functionalism in association with HEC, several commentators have claimed that there is a close link between functionalism and HEC. Rupert (2004) believes that the main argument for HEC "contains a clear functionalist strain", and Shapiro (2008) argues that "support for extended minds often rests on the functionalist theory of minds". Similarly, Weiskopf (2008) remarks:

> "The argument for this [the extended mind] thesis rests on a simple, orthodox functionalist principle [...] Unusual realizers are a staple of the functionalist literature. The hybrids described by advocates of extended minds differ only in lying outside of the normal brain-body system." (Weiskopf 2008)

While several commentators claim that functionalism seems to lend support to HEC, for the most part they do not go so far as to claim that functionalism *entails* HEC: they acknowledge that "functionalist theorizing alone does not resolve the issue of extended states" (Rupert 2004), and that "the case for extended cognition is not going to follow *a priori* from a theory of mind" (Shapiro 2008).

The claim made by Sprevak (forthcoming), however, is a much stronger one: Sprevak argues that functionalism doesn't just support HEC or raise its possibility, but rather that functionalism entails that cognition is actually extended:

> "the relationship between functionalism and HEC goes beyond support for the relatively uncontroversial claim that it is logically or nomologically possible for

> cognitive processes to extend [...] functionalism entails that cognitive processes do extend in the actual world." (Sprevak, forthcoming)

Sprevak's argument relies on functionalism's commitment to preserving the Martian intuition (MI), and on the proponent of extended cognition's commitment to the parity principle (PP).

## 3. The argument from functionalism to extended cognition

Sprevak (forthcoming) points out that in order to preserve MI, functionalism has to be pitched at a relatively coarse-grain: an overly fine-grained functionalism would deny mentality to creatures who exhibited slightly different reaction times or different pain decay responses, for example. Sprevak argues that if we set the grain parameter high enough to preserve MI, then HEC follows automatically.

Sprevak introduces the example of a Martian with an 'ink-mark' memory: his memory is stored internally in a series of ink-marks rather than in patterns of neural activity. To store new information, he activates a process to create new ink-marks. Accommodating MI into one's functionalist approach dictates that these ink-marks count as dispositional beliefs; and thus that the ink-mark contents of Otto's notebook must too:

> "One could imagine a Martian whose memory, instead of being stored in patterns of neural activity, was stored internally as a series of ink-marks. If the Martian wished to store new information, it would activate a process that would create new ink-marks in its storage system [...] But if the functional roles are set this coarse [as to allow for the Martian in question to have beliefs], then they are also

> satisfied by the Otto–notebook system. Therefore, Otto's notebook counts as an extended belief." (Sprevak, forthcoming)

Sprevak concludes that any brand of functionalism which preserves MI will entail HEC. This is an argument which doesn't just establish the modal version of HEC – that extended cognition is possible (conceded by most of the opponents of extended cognition) – but that HEC is a fact about the actual world.

> "the argument concerns the actual existence of extended cognitive processes, not their mere possibility [...] If functionalism is coarse-grained enough to admit possible intelligent Martians, then actual extended systems also qualify as mental. The claim is that one's attitude to Martian worlds commits one to the truth of HEC in the actual world." (Sprevak forthcoming)

So far, this looks like a bonus to proponents of the extended cognition argument: Sprevak has shown that HEC – normally seen to be a controversial claim – is entailed by a reasonably widespread position in philosophy of mind. But Sprevak takes his argument further, and claims that the sort of extended cognition entailed by functionalist is one which even the proponents of HEC would reject.

## 4. Radical extended cognition and a *reductio* of functionalism

Sprevak argues that the version of HEC entailed by functionalism is so radical as to be evidently false. He claims that once one is prepared to accept MI, it looks like we can conceive of all manner of internal Martian states as cognitive: a Martian could have a

memory that worked like a library, or the internet, or an address book, and we'd still be prepared to think of this as an instance of memory.

But now recall the parity principle: PP requires that we treat functionally similar cases equally with regard to their cognitive properties, whether they are internal or external. If we'd consider an *internal* process cognitive, we are committed to considering a functionally equivalent *external* process cognitive too. Sprevak's point is that for any imaginable (internal) Martian cognitive process, the functionally equivalent *extended* process is also cognitive: if a Martian whose memory worked like a library, the internet, or an address book would be considered cognitive, then *actual* cases of people using libraries, the internet, or an address book should also be considered cognitive. These seem to be more radical cases of extended cognition than Clark and Chalmers (1998) had in mind.

To push this point further, Sprevak introduces the example of a program for calculating the dates of the Mayan calendar, and asks us to think of it as an algorithm which could be run on a desktop computer or inside a Martian brain.

> "Imagine that my desktop computer contains a program that calculates the dates of the Mayan calendar 5,000 years into the future. As a matter of fact, I never run this program, entertain the question of what the Mayan calendar is for any year, or even know that my computer contains such a program. However, if I wanted to know the Mayan calendar and explored the resources of my computer, the program would allow me to find the answer quickly. According to the functionalist argument above, I possess a mental process that calculates the dates of the Mayan calendar. The justification: one could imagine a Martian with an internal cognitive process that calculates the dates of the Mayan calendar using

> the same algorithm. The Martian's ability could be innately present as an unintended by-product of the unfolding of its genetic program. The Martian may never happen to use this cognitive process; it may be unaware that it has this cognitive process." (Sprevak forthcoming)

The idea here is that we can imagine a Martian who has an internal process which can calculate the dates of the Mayan calendar, but who has never used this process. Once we're willing to call this process cognitive (by MI), then it follows that my never-used desktop computer program which runs the same algorithm counts as one of my cognitive processes (by PP). Even the most radical proponents of HEC would want to deny this conclusion. Sprevak assumes that this version of HEC is obviously false, and so that it constitutes a *reductio* of the functionalist position from which it was derived.

## 5. A defence of moderate extended cognition (I)

Proponents of HEC do not, in general, want to claim that a computer program one has never used could be part of one's cognitive processes. Clark and Chalmers (1998) distinguish acceptable cases of extended cognition, such as Otto's notebook, from unacceptably radical cases. They claim that the Otto-notebook system had features which weren't exhibited by (for example) the never-used Mayan calendar program on one's computer:

> "First, the notebook is a constant in Otto's life – in cases where the information in the notebook would be relevant, he will rarely take action without consulting it. Second, the information in the notebook is directly available without difficulty.

> Third, upon retrieving information from the notebook he automatically endorses it" (Clark and Chalmers 1998)

One might think that the same sort of distinction between acceptable and overly-radical applications of HEC be used to counter Sprevak, but not according to Sprevak: he thinks that Clark and Chalmers' attempt to distinguish between the Otto case and an overly-radical case must be rejected as it violates their own PP. According to PP we can't set standards for potential *extended* cognitive processes that aren't met by existing *internal* cognitive processes; it is a 'fair treatment' principle and should work both ways. Sprevak claims that there are *internal* human examples of cognition wouldn't be considered cognitive on Clark and Chalmers' above conditions of typicality, availability, and endorsing. His examples include resources used in acts of outstanding human creativity, which aren't typically invoked; the information in my visual system about the position of my eyes, which isn't directly available to me; and the fact that we don't automatically endorse the outputs of our imagining or desiring processes. Furthermore, Sprevak argues, Clark and Chalmers can't add or substitute different conditions: due to the liberalness of the Martian intuition, *any* conditions added to an argument for extended cognition to restrict it to a moderate version will be violated by cases of hypothetical, if not actual, internal cognition. And if extra conditions for extended cognition are added which do not have to be (and indeed are not) met by cases of internal cognition, Sprevak argues, goes against PP and the functionalist grounding of HEC.

But Sprevak misrepresents the position of Clark and Chalmers. He claims that Clark and Chalmers "add extra conditions to the functionalist credo… [which are] individually necessary conditions for cognition" (Sprevak forthcoming), but in fact Clark and Chalmers do no such thing. When they mention that the information in Otto's notebook

is directly available, typically invoked, and automatically endorsed, they are not using these features of the notebook's contents to characterize what makes something *cognitive* rather than *non-cognitive*: rather they are highlighting ways in which the contents of Otto's notebook have the same functional poise as our (and Inga's) normal *dispositional beliefs*. Clark and Chalmers are merely claiming that typicality, availability, and endorsing are features of dispositional beliefs; they are not, *contra* Sprevak's interpretation, claiming that these features are necessary and sufficient conditions for what it is for something to be cognitive.

Given that the criteria used by Clark and Chalmers to support the function of Otto's notebook in his mental life are only concerned with the specific case of dispositional belief, it is simply irrelevant whether there are other forms of cognition (e.g. those involved in human creativity, the visual system, or desire and imagination) which are not typically invoked, readily available, or automatically endorsed. This blocks Sprevak's claim that Clark and Chalmers have no way to restrict extended cognition to moderate cases. Once we start thinking about Clark and Chalmers' criteria in terms of individual mental states types rather than as necessary and sufficient conditions for cognition, it's not clear that attempts to limit HEC to a more moderate form will violate the parity principle.

But even so, this leaves the proponent of HEC to come up with ways of distinguishing between acceptable and non-acceptable cases of cognitive extension for all other mental states, and – so as not to violate the parity principle – ensure that these distinguishing features are met by all internal examples of the relevant mental state. This is a big ask, so it is worth looking at another way of defending moderate HEC from Sprevak's argument.

## 6. A defence of moderate extended cognition (II)

Another way to block Sprevak's *reductio* would be to challenge his argument about the Mayan calendar algorithm. Wheeler (forthcoming), does just this, by claiming that Sprevak's argument relies on a stronger version of MI than that generally assumed by functionalism. He reconstructs Sprevak's Mayan calendar argument (outlined in Section 4 above) roughly as follows:

i) Take an extended process X which isn't cognitive

ii) Imagine a process functionally identical to X inside the head of a Martian

iii) Declare that the Martian system is cognitive (by MI)

iv) Claim that X is thus also cognitive (by PP)

v) Derive a contradiction from (i) and (iv)

vi) Acknowledge that functionalism entails HEC (from a previous argument)

vii) Use the contradiction in (v) as a *reductio* of functionalism

In step (iii) of the reconstructed version of Sprevak's argument above, it seems clear that MI is doing a lot of work. MI is supposed to be the idea that a physically different creature could still have cognitive states and processes like us; Sprevak's use of MI in step (iii), on the other hand, requires that *any* process inside the head of a physically different creature would count as cognitive. The latter can easily be rejected by proponents of HEC (Wheeler forthcoming).

It seems clear that Wheeler is right to claim that not just any process going on inside the head of a Martian – or indeed a human – is cognitive: many brain processes are purely physiological and function merely to regulate low-level bodily systems such as respiration and digestion. But if there were an algorithm for calculating the Mayan

calendar in the head of a Martian, wouldn't we consider this process to be a cognitive one? It seems to depend on the nature of the process: we're more like to consider it cognitive if it is fully-integrated into the Martian's other cognitive resources. But if this were the case, while step (iii) would hold in this instance, it's not clear that step (iv) would go through. Step (iv) relies on PP, which requires that internal and external processes are given equal treatment only insofar as they are functionally equivalent; but the more fully-integrated the Martian's calendar algorithm is, the less functionally equivalent it looks to the isolated desktop computer component. And if we imagined the desktop computer component to be as integrated with the cognitive agent as its counterpart in the Martian case, then it's not clear that we'd deny cognitive status to the former; in which case the *reductio* would fail.

The argument of this section and the previous section propose two ways of blocking the move from HEC to radical HEC. However, there is also a way to block the initial argument: the supposed entailment from functionalism to HEC.

## 7. Mind and cognition: some distinctions and a diagnosis

In sections 5 and 6 I focused on the second part of Sprevak's argument: the attempt to show that the form of extended cognition entailed by functionalism is so radical as to be obviously false, thus providing a *reductio* of functionalism. I argued that it's possible to defend a moderate version of extended cognition, by understanding Clark and Chalmers' 'conditions' correctly and by adhering strictly to the right versions of MI and PP.

I now turn my attention to the first part of Sprevak's argument, the claim that functionalism entails HEC. Recall that HEC only follows from functionalism if the functionalist theory in question is coarse-grained enough to preserve the Martian intuition, MI. But not all varieties of functionalism aim to preserve MI. Psychofunctionalism, for example, is committed to the idea that mental states and processes should be introduced and individuated in terms of their roles in producing behaviour, according to the best scientific theories of behaviour (Levin 2009). The guiding force of psychofunctionalism is to capture generalisations relevant to theories of human psychology, rather than to allow attributions of mentality to non-human hypothetical creatures such as Martians. As such, psychofunctionalists are unlikely to embrace MI. Sprevak (forthcoming) is incorrect to assume that all (or even most) functionalists are committed to the metaphysical task of giving a solution to the mind-body problem.

Clark and Chalmers' argument for HEC is usually understood as one argument for a single position which is variously know as 'the hypothesis of extended cognition' or 'the extended mind argument'. The original 1998 paper, however, comprises two arguments for two distinct conclusions: one, featuring the 'Tetris-rotation' thought experiment, argues for extended cognitive processes; the other, featuring the Otto thought experiment, argues for extended mental states. In the first argument, no 'Martian intuition' is required: the question is one of how our best cognitive science would explain the behaviour in question. As such, the argument for extended cognitive processes needs nothing more than the generalisation-capturing psychofunctionalism outlined above. (And perhaps not even psychofunctionalism, as I shall discuss shortly.) The second argument, for extended mental states, is more complicated. It is concerned with folk-psychological mental states such as dispositional belief, and it is unclear that these are

going to be vindicated by cognitive science. The arguments for considering extended beliefs on a par with normal beliefs start to look a lot like an analysis of our ordinary concepts of the mental states in question. But to decide upon the functional specifications of mental state terms by adopting *a priori* conceptual analysis instead of looking to empirical science to is to turn to analytic functionalism – and analytic functionalism, unlike psychofunctionalism, *does* seem to be committed to the Martian intuition. So if Sprevak's argument works at all, it works against the extended mind argument, which is distinct from the argument for extended cognition.

(Ultimately, the argument rests on our intuitions about cognition and mentality. And it's natural that our intuitions should diverge: our concept of the mind and mental states is tightly bound up with ideas about consciousness, rationality, agency, and the self. While cognitive processes can be agent-level, they can also be sub-agential processes and states (e.g. edge detectors in low-level vision, postural control states of the vestibular system) which aren't integrated with the propositional attitude states of folk-psychology, and thus aren't laden with the same intuitions that we have about mentality. We're more likely, I maintain, to attribute cognition to animals or to robots, than we are to think of them as having minds. As a result, extended cognition is a less controversial – and less philosophically interesting – hypothesis than that of extended minds.)

But does either argument really require functionalism at all? PP emphasises causal role over material implementation, but this does not make it akin to functionalism. Functionalism is the claim that types of mental state are defined by their functional role within a system: the functional role in question is necessary and sufficient for being that type of mental state. PP, on the other hand, makes only a defeasible claim about sufficiency: PP states that if something plays a certain role, then – unless we've got a

good reason not to – we should consider it to be a certain type of mental state. PP is *not* committed to the claim that all instances of that type of mental state will take the same functional role. Furthermore, as Clark and Chalmers (1998) make no claims about consciousness being extended, PP is clearly not intended to apply to all mental phenomena in the way that functionalism is. One could reject functionalism and still argue for extended cognition using PP; there is ultimately no reason why a proponent of extended cognition should have any commitments in the metaphysics of the mind.

Why has there been a general tendency to think about the arguments for extended cognition in functionalist terms? I end with a diagnosis. The parity principle, PP, clearly requires that cognitive processes can be multiply realized; and for many people, the concept of multiple realizability is associated with functionalism. However, this merely points to what Fodor (2000) refers to as "the widespread failure to distinguish the computational program in psychology from the functionalist program in metaphysics" (Fodor 2000, p.105 n.4). Functionalism individuates mental states by their functional relations, but is neutral on how these should be characterized; computationalism claims that these functional relations are computational, but remains neutral on whether they constitute the nature of mental states (Piccinini 2004). The problems come from reading PP as a functionalist claim about the nature of mental states rather than as a computationalist claim about the nature of cognitive algorithms.

## References

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

Fodor, J.A. (2000), *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*, The MIT Press.

Levin, J. (2009) Functionalism. Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/entries/functionalism/>.

Piccinini, G. (2004). Functionalism, computationalism, and mental states. *Studies in the History and Philosophy of Science*. *35*, 811–833.

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

Shapiro, L. (2008). Functionalism and the Boundaries of the Mind. *Cognitive Systems Research*. *9*: 5-14.

Sprevak, M. (forthcoming). Extended cognition and functionalism. *Journal of Philosophy*.

Weiskopf, D. A. (2008). Patrolling the mind's boundaries. *Erkenntnis, 68*(2), 265-276.

Wheeler, M. (forthcoming), Extended functionalism. R. Menary (ed.) *The Extended Mind*, MIT Press.