# COMMON-SENSE FUNCTIONALISM AND THE EXTENDED MIND

*BY JACK WADHAM*

**ABSTRACT:**

The main claim of this paper is that Andy Clark’s most influential argument for ‘the extended mind thesis’ (EM henceforth) fails. Clark’s argument for EM assumes that a certain form of common-sense functionalism is true. I argue, *contra* Clark, that the assumed brand of common-sense functionalism does not imply EM. Clark’s argument also relies on an unspoken, undefended and optional assumption about the nature of mental kinds – an assumption denied by the very common-sense functionalists on whom Clark’s argument draws. I also critique Mark Sprevak’s *reductio* of Clark’s argument. Sprevak contends that Clark’s argument does not merely entail EM; it entails an extended mind thesis so strong as to be absurd. He goes on to claim that Clark’s argument should properly be viewed as a *reductio* of the very common-sense functionalism on which it depends. Sprevak’s argument shares the flaw that afflicts Clark’s argument, or so I claim.

Keywords: extended mind, common-sense functionalism, mental kinds, belief, memory

# 1. INTRODUCTION

Clark’s claim that the mind might extend beyond the brain has generated huge interest and a burgeoning literature. The portion of this literature that centres on Clark’s main argument for EM has a peculiar dialectical structure. Having presented his main argument for EM, Clark dictates terms to his opponents: he sets out a challenge that they must meet if they are to prove him wrong. And, after his opponents have obligingly followed his lead, he sets out to show that they have failed properly to meet his challenge. Rather than adding to this pattern of objection and reply, I argue that that the manner in which the debate is framed comes to seem wrongheaded once we drop an optional assumption.

# 2. EM AS EXTENDED FUNCTIONALISM

Many arguments have been put forward in support of EM, but I will be concerned only with the most well-known and influential of these – an argument that has been developed and refined by Andy Clark. Before examining his argument, however, we need to get clear on what the EM thesis, as formulated by Clark, amounts to.

Clark has recently started referring to EM as a form of ‘extended functionalism’.[[1]](#footnote-1) Clark is not alone in thinking about EM in this way; a growing number of theorists endorse the idea that EM is a functionalist thesis.[[2]](#footnote-2) I suggest that we capture this idea by defining EM as follows:

*EM:* *At least sometimes, in the actual world, the realisers of functional roles definitive of certain mental items include extra-neural components or processes.[[3]](#footnote-3)*

More needs to be said about this definition before we can proceed. The definition is broad enough to capture most conceivable functionalist versions of EM, but is also tailorable to particular versions of EM. So I now need to take the measure of Clark’s view and then cut the cloth accordingly.

First, we need to deal with issues of scope. To what class of ‘mental items’ is EM supposed to apply? EM can apply to conscious states, intentional states, proto-intentional cognitive states, or some combination of these options. Clark’s version of EM applies to intentional states like beliefs and proto-intentional representational states (those which are inaccessible to consciousness), but it does not apply to conscious states (or to the conscious aspects of intentional states, if such aspects exist).[[4]](#footnote-4) I will focus on his argument for the claim that EM applies to intentional states (specifically, non-occurrent beliefs).

I will also be understanding the ‘items’ part of ‘mental items’ such that it refers not only to mental states like beliefs but also to mental processes, like believing, remembering and so on. Clark takes EM to be true both of mental states and mental processes.[[5]](#footnote-5) Finally the reference to ‘extra-neural components or processes’ is ambiguous between that which is bodily but extra-neural and that which is extra-neural *and* extra-bodily. In other words, EM, as defined above, could be understood as the ‘embodied’ claim that the mind extends past the brain into the body or as the ‘extended’ claim that the mind extends past the body and into the world. Clark claims that the mind is both embodied and extended.[[6]](#footnote-6)

The next major task is to clarify the notion of ‘functional role’ in play in the above definition. Clark states explicitly that his argument for EM relies on common-sense functionalism, as formulated by Jackson and Braddon-Mitchell (see Clark 2008a: 88-9). So I will cash out his version of EM in terms of this view. On this brand of functionalism (as on most brands), to be a mental item is just to be that which plays (or is disposed to play) a certain functional role. So, for example, to be a belief is just to be a state disposed to be caused (typically) by perceptions of the appropriate kind, to interact with other beliefs in the appropriate way, to interact with desires in a way that typically generates actions of the appropriate kind, and so on. So the mental kind ‘belief’ is identified with a certain functional role. With this identification in place, Jackson and Braddon-Mitchell claim, we can say that the realiser of the kind ‘belief’ is just that which stands in the causal relations definitive of that kind. [[7]](#footnote-7) Normally, it is assumed that the realiser in question is just some physical process occurring in some region of the brain, but this is precisely the assumption that EM denies.

The claim distinctive to *common-sense* functionalism, of the sort advocated by Jackson and Braddon-Mitchell (2007), is that we should look to common-sense folk psychology in order to determine which functional roles are definitive of which mental states. This view stands in contrast with *empirical* functionalism (sometimes called psychofunctionalism), according to which we should look to our best scientific theories in order to determine which functional roles are definitive of which mental states. [[8]](#footnote-8) The idea, roughly, is that ordinary folk rely on a tacit theory about the minds of others, in order to make sense of the others’ everyday behaviour. The job of the theoretician is to make explicit the most central theoretical commitments of this folk psychological theory – the folk platitudes – and to build out of these an explication of the folk roles definitive of mental states.

But how are these platitudes to be identified? And, perhaps more importantly, why are they to be trusted? Jackson and Braddon-Mitchell claim that the folk platitudes are to be trusted because they are predictively and explanatorily *potent*, but at the same time theoretically *modest* (2007: 268-272).[[9]](#footnote-9) They are potent because they allow us to explain the behaviour of others. I can explain the fact that Jones is crouched in a river, passing mud through a sieve by attributing to Jones a) the belief that there is gold in the river and b) the desire to find gold. The platitudes are modest because they make no claims about realisers. My attribution of a belief to Jones does not commit me to any theory about what it is, in Jones’ head or elsewhere, that is realising the relevant belief-role. This modesty makes it the case that folk psychology is ‘peculiarly’ unlikely to be proved wrong by future scientific research, or so the claim goes (2007: 269-70).

With the ‘functional role’ part of Clark’s EM explained, my tailoring is complete. We now need to see how my definition handles the distinction between EM and content externalism. The traditional definition of the extended mind thesis as ‘vehicle externalism’ was supposed to mark out this distinction clearly.[[10]](#footnote-10) Content externalism, famously motivated by Hillary Putnam’s (1975/1985) twin earth thought experiment, is the view that the contents of at least some mental states are individuated (or, in some stronger versions of content externalism, constituted) by external, worldly features. By contrast, the standard line goes, EM is about the vehicles of mental states – the carriers of mental content.[[11]](#footnote-11)

My functionalist definition of EM gives a way of preserving what is important about this distinction while bringing it into the functionalist fold. To see how it does so, let’s consider the contrast between EM and the long-armed functionalism designed to accommodate content externalism.[[12]](#footnote-12) According to long-armed functionalism we must, when specifying the functional role definitive of some mental item, include functional role clauses that specify causal interactions with distal (environmental) objects. So, for example, we might specify perceptual states in terms of the worldly objects that are disposed to cause them, rather than specifying them in terms of the (proximal) retinal or neural inputs that are disposed to cause them. So according to long-armed functionalism, the specification of the *roles* definitive of mental items must contain essential reference to beyond-the-skin factors. By contrast, on my definition, EM is a thesis about the *realisers* of mental items; it is the claim that the functional roles definitive of mental items are *realised* in part by extra-neural goings on. One might at this point ask how Clark moves from common-sense functionalist claims about folk functional *roles* to the EM claim, which is a claim about the *realisers* of these roles. To answer this question, we have to look at Clark’s argument for EM.

# 3. EM: CLARK’S ARGUMENT, SPREVAK’S *REDUCTIO*

Clark’s best-known argument for EM is the Otto and Inga argument, originally developed in a paper co-authored with Chalmers (1998), but repeated and refined in Clark’s later work. It is on this argument that I will focus initially. The aim of Clark’s argument is to show that there are situations in which some mental items (non-occurrent beliefs in the case I consider) are realised by partly extra-neural resources. The argument starts with Inga, who is a benchmark case – a case in which non-occurrent beliefs are uncontroversially realised. Inga believes that the Museum of Modern Art (MoMA) is on 53rd street. When she wants to go to MoMA, she consults her memory, retrieves her belief about its location, and goes on her way. Compare Inga to Otto, who has Alzheimer’s disease, and so cannot form and retrieve non-occurrent beliefs in the normal way. Otto has a way around this problem. Instead of storing information in his brain, he stores it in his notebook. When he is told that ‘MoMA is on 53rd street’ he writes this sentence down in his book. When he wants to go to MoMA, he consults his notebook, discovers its location, and goes on his way. The aim of Clark’s argument is to show that Otto, like Inga, is capable of forming and retrieving non-occurrent beliefs (even though he does so in an unorthodox way).

The argument rests on the claim that, at the appropriate level of abstraction, the procedure performed by Otto is functionally analogous to the procedure performed by Inga. This is where common-sense functionalism comes in. The idea is to appeal to folk psychological intuitions in order to establish the claim that there is no intuitively significant functional difference between the procedure performed by Otto and that performed by Inga. And if there is no significant functional difference between Inga and Otto, then the common-sense functionalist has no justification for claiming that Inga has non-occurrent beliefs while Otto lacks such beliefs. For the common-sense functionalist, forming/retrieving beliefs is just a matter of realising certain functional roles. So if Otto and Inga are functionally analogous in the relevant respects, then they are both forming/retrieving beliefs.

And once it is established that Otto is capable of forming/retrieving beliefs when he uses his notebook in certain ways, it is a relatively short extra step to the claim that these beliefs are partly realised by the relevant parts of his notebook (or processes that constitutively involve these notebook-parts). Some object to this extra step, but it seems to me an easy one to make.[[13]](#footnote-13) If Otto really does believe that MoMA is on 53rd street once he has written this sentence in his book, it would be very odd to claim that the sentence written in his book forms no part of the realisation base for this belief, or so it seems to me. But my argument does not rest on this point, so I won’t argue for it. My focus is on the argument to the claim that Inga and Otto are functionally analogous in all relevant respects.

The argument for this claim essentially relies on the thought that we can get from Inga’s case to Otto’s without subtraction of any functional role properties that are essential to the mental item in question (i.e. non-occurrent belief). Clark claims that there are no relevant functional differences between Otto and Inga just in case the following conditions are satisfied by Otto’s notebook-involving procedure:

1) That the resource be reliably available and typically invoked (Otto always carries the notebook and won’t answer that he “doesn’t know” until after he has consulted it).

2) That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.

3) That information contained in the resource should be easily accessible as and when required. (Clark 2010b: 46)

Clark stipulates that Otto’s notebook-use satisfies these ‘extension-conditions’. And, he argues, when these extension-conditions are satisfied, Otto’s procedure is sufficiently similar to Inga’s to justify the claim that Otto is, like Inga, capable of belief storage/retrieval.

To establish this claim, he essentially lays down a challenge to his opponents. The challenge is to find some intuitively significant difference between the functional roles instantiated by Inga and those instantiated by Otto. Let’s call such a difference a *Deep Difference*. If, after a suitably long period of time and effort, nobody can point to a genuinely *Deep Difference* between Inga and Otto, we can (defeasibly) conclude that there are no such *Deep Differences* to point to. But how do we test such a difference in order to determine whether it is genuinely deep? Clark’s procedure for deciding on this matter is to create a thought experiment designed to test the putative *Deep Difference* against our folk intuitions. Clark justifies such appeal to folk intuitions by invoking Jackson and Braddon-Mitchell’s common-sense functionalist claim that ‘normal agents command a rich (albeit largely implicit) theory of the coarse functional roles distinctive of various familiar mental states’ (Clark 2008a: 88). He writes that it is this ‘coarse or common-sense functional role that (…) displays what is essential to a given mental state’ (2008a: 89).

Let’s look at an example of this procedure in action, putting to the test the recency/priming/generation difference, raised by Adams and Aizawa (2001) in their attempt to establish the falsity of EM.[[14]](#footnote-14) Let’s focus on recency: if I am asked to remember items on a list read out to me by someone, I am more likely to remember those listed items that were read out last (*ceteris paribus*). Not so with Otto.[[15]](#footnote-15) This is a difference between the functional profile of Otto’s putative memory system and what we take to be paradigm cases of memory systems. Otto’s notebook system also fails to exhibit other quirky features characteristic of human memory (priming and generation, for example). Assuming that Inga exhibits these quirks of memory, we have here a functional difference between Otto and Inga.

To show that this difference is shallow, Clark imagines a Martian, or just an unusual human, who doesn’t exhibit the recency effect, or memory effects of this type, but who is just like normal humans in all other psychological respects (2008a: 93). Would we claim that such people/aliens don’t remember? Plausibly not. As Clark puts it ‘to insist that some alien mode of storage and retrieval was not cognitive *just because* it failed to exhibit features such as recency, priming and crosstalk would be simultaneously to scale new heights of anthropocentrism and neurocentrism’ (2008a: 93 emphasis added). According to Clark, if we met such a creature, the appropriate thing to think would be that its memory system is slightly different from our own, but is a memory system nonetheless. And if this is the right thing to say about the alien case, what motivation, other than neural chauvinism, do we have for applying a different standard to Otto?

Notice how this thought experiment works. Clark claims that we would not deny that his imagined creature is capable of belief storage/retrieval *just because* it lacks the functional features in question. This is just to say that exhibiting this collection of features is not individually necessary for belief storage/retrieval. The idea then seems to be that if we can get from Inga to Otto, without the subtraction of any functional feature (or collection of features) that is individually necessary for realising belief storage/retrieval, then we are justified in claiming that Otto, like Inga, is capable of belief storage/retrieval. So Clark’s challenge to his opponents can be put as follows: show me some *Deep Difference* between Otto and Inga, where a *Deep Difference* is defined as follows:

1. Inga exhibits functional feature *Fi* but Otto doesn’t.
2. Exhibiting *Fi* is individually necessary for realising some mental item *M*.

In our case, *M* is the belief that MoMA is on 53rd street. Each time an opponent raises a putatively significant functional difference between Otto and Inga, Clark constructs a bespoke Martian thought experiment to show that this difference is merely shallow. Some illustrative examples are as follows:

*Fi* = the property of not inspecting one’s memories by perceptual means. Counterexample to (individual) necessity: Terminator (Clark and Chalmers 1998:16)

*Fi* = the property of involving non-derived as opposed to derived content.[[16]](#footnote-16) Counterexample to (individual) necessity: Martian ‘endowed with an extra biological routine that allowed them to store *bit-mapped images* of important chunks of visually encountered text.’ (2010c: 88-9).[[17]](#footnote-17)

*Fi* = the property of being a persisting and integrated part of the relevant cognitive system.[[18]](#footnote-18) Counterexample to (individual) necessity: Metamorpho/Metamento, whose brain includes detachable/re-attachable components. (Clark 2010a: 457).

This list is far from exhaustive, but it gives a sense of Clark’s argumentative methodology. Until someone finds some *Deep Difference* that is immune to counterexamples of this kind, there is a presumption in favour of the claim that Otto and Inga are functionally alike in the relevant respects. I do not go into the details of Clark’s particular counterexamples because such details are not relevant to my argument. I will be taking issue with the overall argumentative strategy employed by Clark, not the particular thought experiments he deploys. But first let’s look at Sprevak’s (2009) *reductio* of Clark’s argument.

Sprevak’s *reductio* works by applying Clark’s own argumentative strategy to Clark’s three extension-conditions. Sprevak tacitly assumes, as I will for now, that Clark’s extension-conditions must be individually necessary for belief storage/retrieval, if they are to block the powerful *reductio* to be examined shortly.[[19]](#footnote-19) He then builds counterexamples to the necessity of each condition:

1. **Condition**: the notebook must be ‘reliably available and typically invoked’. **Counterexample**: Imagine a Martian who’s putative ‘non-occurrent beliefs’ are neither reliably available nor typically invoked. Does this Martian lack the capacity to remember? Is the Martian entirely lacking in non-occurrent beliefs?[[20]](#footnote-20)

This procedure is then repeated on the other two conditions. To undermine the second condition (the automatic endorsement condition), Sprevak invites us to imagine a creature who ‘redundantly’ performs a quick plausibility-check on any putative belief they retrieve, and thus falls short of the automatic endorsement condition (2009: 514-5). Such a creature, he argues, would still count as a believer. And to undermine condition three, he claims we would not deny that someone has beliefs just because s/he finds those beliefs ‘difficult to access’, developing another Martian case to support this claim (2009: 515).

Let’s suppose (as seems plausible) that Sprevak’s counterexamples are successful in undermining the individual necessity of each of Clark’s three conditions. Let’s also suppose, for the sake of argument, that we agree with Clark that Otto realises belief storage/retrieval. One could justifiably ask at this point how far we should generalise from the Otto case. Does anyone who reads anything in a notebook (or any artefact bearing written symbols for that matter) also thereby count as retrieving non-occurrent beliefs? Clark’s three conditions were supposed to provide principled grounds for giving non-absurd answers to questions of this kind. The idea was supposed to be that most symbol-bearing artefacts we use are not reliably available, typically invoked and easy to access. So most ordinary cases of reading fall short of non-occurrent belief-retrieval, by the lights of Clark’s conditions.

But if Sprevak’s counterexamples are successful (and they seem every bit as compelling as Clark’s own Martian cases), then Clark is not entitled to rely on his three conditions when attempting to exclude obviously absurd cases of putative cognitive extension. This is because the conditions fail to pick out genuinely *Deep Differences* between Otto and a given absurd case of cognitive extension. So Sprevak can then confront Clark with the following challenge: show me some *Deep Difference* between Otto and Blotto (where ‘Blotto’ is just some obviously absurd case of mental extension). If Sprevak’s counterexamples are successful, none of Clark’s three conditions can succeed at this job. It therefore seems that until some more successful conditions are formulated, we can defeasibly conclude that Clark’s argument generalises to create absurd results.

# 4. DIAGNOSING THE PROBLEM

Something has clearly gone badly wrong here. Sprevak’s suggestion is that the problem lies with common-sense functionalism, and its attempt to systematise our intuitions:

The correct lesson might be that our intuitions about mental systems cannot be systematised without doing serious damage to our concept of mentality. Functionalism aims to provide an answer to what makes certain systems mental. Perhaps such an answer cannot be given. (2009: 522)

I argue that an alternative diagnosis is available. To see this, let’s review Clark’s basic strategy. There are many functional differences between the procedure performed by Inga and that performed by Otto. Let’s use *F1,…,n* to denote the full set of these differences. Each time Clark’s opponent claims that some difference on this list is a *Deep Difference*, Clark creates a Martian-style thought experiment to undermine that claim: the functional property lacked by Otto, *F1 ­­*say, is not individually necessary for realising the mental item in question, so Clark’s opponent fails. Sprevak repeats this procedure, but instead of using it to move from Inga to Otto, he uses it to move from Otto to cases like Blotto. This argumentative strategy relies on a critical (and *optional*) assumption.

To see what the assumption is, suppose Clark successfully undermines the individual necessity of *F1 ­*for realising mental item *M*. He does the same for *F2* and then for *F3*. Even if he does this successfully, he has not ruled out the possibility that satisfying some combination of these three conditions *is* necessary for realising *M*. To see why this possibility might undermine Clark’s argument for EM, let’s look again at the list of functional features (*F1,…,n*)which are realised by Inga but not by Otto. Clark might show that no feature on this list is individually necessary for realising *M*, while failing to show that there is no combination of features on the list whose satisfaction is necessary for realising *M*.

Suppose I put it to you that the drum stick I am holding in my hand is a pen, and defy you to convince me otherwise. I challenge you to point out a single feature, lacked by my drumstick, which is essential to penhood. You point out that I can’t write with my drumstick. But then I reply that something can be a pen even if it’s not possible to write with it. A pen that’s broken, out of ink, or locked in a secure safe to which nobody has the key is still a pen, despite the fact that we can’t write with it. You then object that it doesn’t even look like a pen, and I respond again that this feature is also inessential, concocting a range of clever thought experiments to prove my point. This goes on for some time and eventually, whether out of exhaustion or boredom, you give up trying to convince me. Suppose I lay down this challenge to lots of clever people, and each time wear them down in a similar fashion. Does the fact that I can do so give me a strong (though defeasible) justification for the claim that my drumstick is a pen? Clearly not. And it is possible that, in the relevant respects, Otto is to Inga as my drumstick is to some paradigm example of a pen.

And Jackson and Braddon-Mitchell, on whose common-sense functionalism Clark explicitly relies, use exactly this analogy to describe mental states:

There is a list of features that we regard as paradigmatic of pens. Something that satisfies every single one is by definition a pen. If it uses ink, is small enough to fit in the hand, is used to write with, is called a ‘pen’, is barrel shaped, and has a nib, then it is a pen. That follows from our concept of a pen. But nothing in that list is sacrosanct. Any single feature may be absent and yet the object still be a pen (…) what matters is that enough of the list or near enough is satisfied, and what counts as enough may itself be a vague matter (2007: 54)

The key point is that a pen can lack any of the properties enumerated by Jackson and Braddon-Mitchell, while still counting as a genuine pen. But, the idea goes, if enough of the relevant properties are absent, we reach a point where the right thing to say is that the thing in question is not a pen.

When applied to the mental, Jackson and Braddon-Mitchell’s idea is that each functional role property by which a given mental item is defined should be qualified with a *ceteris paribus* clause. Take beliefs, for example. Beliefs are typically action-guiding, but we could imagine lots of beliefs that are not. This point is illustrated by David Lewis’ resolute deceiver, who is ‘disposed come what may to behave as if his mental states were other than they really are’ (1994: 418). Such a person may be a slightly atypical believer, but is a believer nonetheless. Lewis claims this is the case while denying that ‘anything goes’ when it comes to mental state attribution (1994: 418).[[21]](#footnote-21) Jackson and Braddon-Mitchell, whose brand of common-sense functionalism is heavily indebted to Lewis, take a similar line.

More, however, needs to be said about how this proposal is supposed to work and how it might be deployed to block Clark’s and Sprevak’s arguments. Jackson and Braddon-Mitchell give less guidance than would be desirable about how their proposal is supposed to work, so I suggest we turn to Lewis, whose remarks on this matter are instructive. In discussing cases similar in kind to those of Otto and Inga, Lewis suggests that we posit a kind of ambiguity about exactly which folk roles are definitive of which mental items. He claims that this ambiguity is not concocted *ad hoc* by the common-sense functionalist, but is rather ‘a commonplace kind of ambiguity – a kind that may arise whenever we have tacit relativity and criteria of selection that fail to choose a definite *relatum*’ (1980: 221). I propose that ‘typicality’ is the relevant criterion of selection we should use when determining which folk roles are definitive of which mental item. There are many different respects in which a given mental item might be typical of its kind, so there are many different typicality criteria against which a given mental item might be judged.

To see how this proposal works, let’s take belief again. One axis of typicality along which a belief might be assessed is its action-guiding role. As the deceiver case shows, a state can count as a belief despite being atypical on this axis. But there are many other axes of typicality against which a given putative belief might be assessed. Below is an illustrative but inexhaustive list:

1. Evidence-sensitivity: typically, the credence we attach to a given belief is quite sensitive to the level of evidence we have in its support.[[22]](#footnote-22)
2. Involuntariness: for those unskilled in the art of self-deception, beliefs are typically involuntary in a meaningful sense.[[23]](#footnote-23)
3. Causal role: typically, beliefs cause and are caused by, events of a distinctive kind. Otto’s ‘beliefs’ are atypical with respect to causal role in that we can, for example, render him incapable of accessing them by poking out his eyes.

Beliefs can be atypical in certain respects, but typical in others. If a putative belief is atypical in one respect, but typical in all others, it might make sense to say that it is still a belief. But if it is atypical in just about every respect, we might justifiably wonder if it is a belief in any meaningful sense. Ambiguity arises when we try to determine what should be said about the cases in the middle. Take some putative belief that is belief-typical on certain axes of typicality, but atypical along others. It might be difficult to say one way or another whether it is suitably close to paradigm cases of belief to count as a mental state of the same kind. The typicality axes referred to here are derived directly from the folk platitudes – the folk platitudes dictate that mental item *M* typically exhibits a certain set of functional features. But the folk platitudes do not tell us what we should make of putative mental items that are typical in some respects, but atypical in others.

If something like this story is right, then there is something seriously wrong with Clark and Sprevak’s methodology. They challenge their opponents to identify some respect in which Otto or Blotto are atypical. Once the opponent identifies some relevant respect, Clark and Sprevak proceed to show that a person (or an alien) could be atypical *in the respect identified* while still realising the relevant mental item. They do so by imagining some creature whose mental states are typical in all respects except the respect *Fi* whose significance is under consideration. They then show, by use of a Martian case, that one could be atypical with regards to *Fi* and yet still count as realising the relevant mental item.[[24]](#footnote-24) They then repeat this procedure for each respect identified by some opponent.

But this isn’t good enough. One has to show that the sum total of respects (*F1,…, n*) in which Otto or Blotto are functionally atypical (relative to benchmarks like Inga) do not, *when taken together, or in any combination*, undermine the claim under discussion: that Otto or Blotto also bears some mental item borne by Inga. Until this is done, the question of whether Otto or Blotto bears the relevant mental item must remain open. And from this it follows that common-sense functionalism of the kind outlined here has not been shown to imply either EM or the absurd conclusions that Sprevak attempts to derive.

So Clark and Sprevak must either change their argumentative methodology or they must give principled reasons for rejecting the common-sense functionalist’s claims about *ceteris paribus* typicality conditions. Until they do so, they cannot claim to have established the claims for which they argue; Clark has not established the truth of EM and Sprevak has not provided a successful *reductio* of common-sense functionalism.

# 5. OBJECTION AND REPLY[[25]](#footnote-25)

Is there room for Clark to accommodate this point about *ceteris paribus* conditions while leaving his argument in tact? One option open to him is to claim that, while his opponents assume that *their* anti-Otto conditions are individually necessary for realising the relevant mental items, *his* anti-Blotto conditions need not be so. He could argue as follows: ‘my opponent thinks that involving non-derived content is necessary for counting as a non-occurrent belief, but I have shown this to be false with a Martian counterexample. Another opponent thinks that such-and-such is individually necessary for counting as a non-occurrent belief, and I have shown *that* to be false using another bespoke Martian counterexample. It’s not me making such individual necessity assumptions, it’s my opponents. I merely show their assumptions to be false.’

If Clark takes this line, it looks like he can use his Martian counterexamples to silence the anti-Otto brigade without himself relying on any individual necessity assumption. If he does this, there is no reason why he should build the individual necessity assumption into his anti-Blotto conditions. This means he can treat his conditions in the manner suggested in the previous section. He can claim that while none of his three conditions are individually necessary for non-occurrent belief, some combination of them (or some combination *involving* them) is necessary.[[26]](#footnote-26) If he does this, he can neutralise Sprevak’s *reductio*, which worked by taking each of Clark’s three conditions in turn and undermining each’s claim to individual necessity. If the conditions were never meant to be individually necessary, then Sprevak’s counterexamples don’t automatically work. And there is some textual evidence suggesting that Clark does *not* treats his three anti-Blotto criteria as individually necessary in the above sense.[[27]](#footnote-27) So it looks like Clark might be safe after all.

This appearance is misleading. The problem is that the strategy just described is overly contingent on the manner in which Clark’s opponents happen to argue. Let’s grant, for the sake of argument, that as it happens, members of the anti-Otto brigade all argue against Clark by a) picking some functional feature that differentiates Otto from Inga, b) assuming that having this feature is individually necessary for having the relevant non-occurrent beliefs and c) concluding that Inga has the relevant beliefs while Otto does not. This being the case, Clark’s bespoke Martian strategy looks successful against the anti-Otto brigade.

But suppose his opponents gang up together and claim that while the features they previously pointed to may not be *individually* necessary for counting as a believer, having some combination of the features they point to *is* necessary for so counting. Suppose, in other words, that his opponents adopt the strategy suggested in the previous section. In fact, to make the supposition less hypothetical, I hereby declare myself an anti-Otto brigadier, and one who adopts the strategy just suggested. If the argument in the last section is right, there is nothing obviously incoherent about the idea that conditions may be ‘combinatorially necessary’ even if they’re not individually necessary. And recall that Clark’s challenge was: show me some functionally significant feature exhibited by Inga, but not Otto, such that only Inga has non-occurrent beliefs. I now have a set (*F1…, n*)of functional features which look *prima facie* to be significant, for all the standard reasons raised by the anti-Otto brigade. Clark has only argued that these are not individually necessary. He has not considered (let alone *undermined*) the possibility that they might be significant in the ‘combinatorially necessary’ way. To undermine this possibility, he cannot simply generate bespoke Martian counterexamples to pick off my conditions one by one, as he has done to date.

It seems that unless Clark comes up with some new argumentative strategy, he must show why his opponents are *forced* to adopt the individual necessity assumption, while he is *free* to drop it (when developing his anti-Blotto conditions). If he fails to do this on principled grounds, it looks like he is adopting a double standard: a tough rule for the anti-Otto brigade, and a lax rule for the anti-Blotto brigade. It’s not enough to show that the anti-Otto brigade happen to impose the tough rule on themselves, unless it can be shown that they are *obliged* to do so for some reason. Otherwise, I can simply pronounce myself a new-and-improved anti-Otto brigadier, impose the lax rule on my list of anti-Otto features (*F1,…,n)*, and use these to block Clark’s argument for EM.

We end, again, with Clark between a rock and a hard place. Either adopt the individual necessity assumption, in which case Blotto becomes a problem. Or drop the assumption, and thereby defang the Martian counterexamples on which Clark usually relies to motivate EM. To avoid this dilemma, Clark must either give principled grounds for adopting what would otherwise look like a double standard, or he must develop a whole new argumentative strategy. Either way, the ball is now in his court.[[28]](#footnote-28)

# REFERENCES:

Adams, F and K, Aizawa (2001) ‘The Bounds of Cognition’, *Philosophical Psychology*,14: 43-64.

Adams, F. and K. Aizawa (2010) ‘Defending the Bounds of Cognition’ in R. Menary (ed) *The Extended Mind*, 67-80. Cambridge, Mass.: MIT Press.

Block, N. (1990) ‘Inverted Earth’, *Philosophical Perspectives*, 4: 53-79

Bortolotti, L. (2009) *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press.

Clark, A*.* (2003) *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*,New York: Oxford University Press.

Clark, A. (2005) ‘Intrinsic content, active memory, and the extended mind’, *Analysis*, 65/285: 1-11.

Clark, A. (2008a) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*,Oxford: Oxford University Press.

Clark, A. (2008b) ‘Pressing the flesh: A tension in the study of embodied, embedded mind?’ *Philosophy and Phenomenological Research*, 76/1: 37-59

Clark, A. (2009) ‘Spreading the joy? Why the machinery of consciousness is (probably) still in the head’, *Mind*, 118/472: 963-993

Clark (2010a) ‘Finding the Mind’, *Philosophical Studies,* 152/3: 447-61

Clark (2010b) ‘*Memento’s* Revenge: The Extended Mind, Extended’, in R. Menary (ed.) *The Extended Mind*, 43-66. Cambridge, MA.: MIT Press

Clark, A. (2010c) ‘Coupling, Constitution and the Cognitive Kind’, in R. Menary (ed.) *The Extended Mind*, 81-100. Cambridge, MA.: MIT Press

Clark, A., and Chalmers, D (1998) ‘The Extended Mind’, *Analysis*, 58: 7-19

Dennett, D.C. & Kinsbourne, M. (1992) ‘Time and the observer: The where and when of consciousness in the brain’, *Behavioural Brain Sciences*, 19: 183-247

Di Paolo, E. (2009), ‘Extended Life’, *Topoi*, 28: 9-21.

Hohwy, J. and Rajan, V. (2012) ‘Delusions are forensically disturbing perceptual inferences’, *Neuroethics*, 5/1: 5-11

Hurley, S. (1998) *Consciousness in Action*,Cambridge, MA.: Harvard.

Hurley, S. and Noë, A. (2003) ‘Neural Plasticity and Consciousness’, *Biology and Philosophy*, 18: 131-68

Jackson, F., Braddon-Mitchell, D. (2007) *Philosophy of Mind and Cognition: An Introduction*, 2nd edn. Oxford: Blackwell Publishing

Lakoff, G. and Johnson, M. (1999) *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, New York: Basic Books.

Lewis, D. (1980) ‘Mad Pain and Martian Pain’ in N. Block (ed.) *Readings in Philosophy of Psychology*, *Vol. 1.* 216-222.Cambridge, MA: Harvard University Press

Lewis, D. (1994) ‘Reduction of Mind’ in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*,412-31.Oxford: Blackwell.

Noë, A. (2004) *Action in Perception*. Cambridge MA: MIT Press.

Noë, A. and Thompson, E. (2004a) ‘Are There Neural Correlates of Consciousness?’, *Journal of Consciousness Studies,* 11: 3-28.

Noë, A. and Thompson, E. (2004b) ‘Sorting Out the Neural Basis of Consciousness: Authors’ Reply to Commentators’, *Journal of Consciousness Studies*, 11: 87-98

Putnam, H. (1975/1985) ‘The meaning of “meaning”’ in *Philosophical Papers, Vol. 2: Mind, Language and Reality*, 215-272. Cambridge University Press.

Rowlands, M. (2006) *Body Language: Representations in Action.* Cambridge, Mass.: MIT Press.

Rowlands, M. (2010a) *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. London: MIT Press

Rowlands (2010b) ‘Consciousness, Broadly Construed’ in R. Menary (ed.) *The Extended Mind*, 271-294. Cambridge, MA.: MIT Press

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

Rupert, R. (2009) *Cognitive Systems and the Extended Mind*, Oxford: Oxford University Press.

Rupert, R. (2010) ‘Extended Cognition and the Priority of Cognitive Systems’, *Cognitive Systems Research*, 11: 343-356

Shapiro, L. (2004) *The Mind Incarnate*, Cambridge: Cambridge University Press.

Shapiro, L. (2008) ‘Functionalism and Mental Boundaries’, *Cognitive Systems Research*, 106: 503-527

Sprevak, M. (2009) ‘Extended Cognition and Functionalism’ *Journal of Philosophy*, 106/9: 503-527

Sterelny, K. (2004) ‘Externalism, epistemic artefacts, and the extended mind’ in Richard Schantz(ed.) *The externalist challenge: New Studies on Cognition and Intentionality*. New York: de Guryter.

Switzgebel, E. (2001) ‘In-between Believing’, *Philosophical Quarterly*, 51/202: 76-82

Wheeler, M. (2010a) ‘Minds, things, and materiality’ in C. Renfrew and L. Malafouris(eds) *The Cognitive Life of Things.* 29-37,Cambridge: Cambridge University Press

Wheeler, M. (2010b) In defense of extended functionalism’ in R. Menary (ed) *The Extended Mind*, 245-270. Cambridge, Mass.: MIT Press.

Wilson, R. (2010) ‘Meaning Making and the Mind of the Externalist’ in R. Menary (ed) *The Extended Mind*, 167-188. Cambridge, Mass.: MIT Press.

*University of Sheffield, UK*

1. See Clark (2008b: 37) and Clark (2010a: 449). [↑](#footnote-ref-1)
2. See, for example Michael Wheeler (2010a; 2010b). For a commentary on the link between EM and functionalism, see Wheeler (2010b). See also Di Paolo (2009) for an attempt to develop an anti-functionalist, ‘autopoetic’ version of the extended cognition thesis. [↑](#footnote-ref-2)
3. There are good exegetical reasons for thinking that Clark has something like this definition in mind when he talks of EM. Due to limitations of space, I cannot go into these reasons here. But see (2008a: xviii) for what I take to be a more evocative (but less precise) functionalist definition of EM. [↑](#footnote-ref-3)
4. See Clark and Chalmers (1998) for original statement of the view and its scope. See Clark (2009) for an argument against the view that EM is true of conscious states. For arguments in favour of EM about conscious states/processes, see Hurley and Noë (2003), Noë and Thompson (2004a; 2004b), Rowlands (2010a; 2010b). [↑](#footnote-ref-4)
5. There are some, like Wilson (2010: 183) and Rowlands (2010a: 63-65), who dissent from this analysis. They claim that EM applies only to mental processes. I will not go into their reasons for doing so. The issue has no direct bearing on the dialectic presented here. [↑](#footnote-ref-5)
6. In the literature, there has been some tension between writers in the ‘embodied’ camp – e.g. Lakoff and Johnson (1999), Shapiro (2004), Noë (2004) – and those in the ‘extended camp’, like Clark (2008a). For a discussion of this tension, see Clark (2008b) and Rowlands (2010a: 102-6). The tension is normally traced back to difference between the ‘liberal’ functionalism presupposed by writers like Clark and the less liberal variety adopted by those in the ‘embodied’ tradition. I will be focusing on Clark’s position, and its relation to the brand of functionalism on which it rests. [↑](#footnote-ref-6)
7. Jackson and Braddon-Mitchell in fact go further than this. They claim that a given token belief that *p* is to be identified with a certain functional role. Put more generally, they think the content of a given mental state is determined by its functional role, and nothing besides. As far as I can see, Clark’s argument for EM does not rely on this additional claim. [↑](#footnote-ref-7)
8. In this respect, his argument for EM differs from that put forward, for example, by Wheeler, which seems to rest on a (nuanced) version of empirical functionalism (see Wheeler 2010b: §5). Robert Rupert, a critic of EM, makes direct appeal to empirical functionalism, or psychofunctionalism, in his critiques (2004: 423; 2009: 91-4; 2010: 345) [↑](#footnote-ref-8)
9. What follows is a brief sketch of the motivations for common-sense functionalism, and should not be taken as an attempt to give a full-scale defence of the position – for one of these, see Jackson and Braddon-Mitchell (2007). I will return briefly to the debate between empirical and common-sense functionalism in §4. [↑](#footnote-ref-9)
10. See Hurley (1998) and Rowlands (2006) for seminal characterisations of the EM thesis as vehicle externalism. [↑](#footnote-ref-10)
11. See Dennett and Kinsbourne (1992) for a seminal and entertaining treatment of the vehicle/content distinction. [↑](#footnote-ref-11)
12. See Jackson and Braddon-Mitchell (2007: 119) for a defence of this kind of functionalism. See Block (1990: 58) for a more detailed treatment of ‘long-armed functionalism’. [↑](#footnote-ref-12)
13. Shapiro (2008), for instance, raises the problem of distinguishing between a realiser of some functional role and something that merely contributes causally to the realisation of said role. [↑](#footnote-ref-13)
14. See Rupert (2004) for a similar argument against EM. [↑](#footnote-ref-14)
15. Assuming, that is, that Otto has a neat filing system in his notebook that prevents old entries from becoming swamped under by new ones. [↑](#footnote-ref-15)
16. See Adams and Aizawa (2001; 2010) for more on this distinction. The rough idea is that linguistic representations derive their meaning from convention, or from individual stipulations. By contrast, the representational status of mental states like beliefs and desires is not supposed to be derivative in this sense [↑](#footnote-ref-16)
17. See Clark (2003), Clark (2005) for more details. [↑](#footnote-ref-17)
18. See Rupert (2004; 2009). [↑](#footnote-ref-18)
19. See the final section for a discussion of this assumption. [↑](#footnote-ref-19)
20. I construct my own version of the counterexample, rather than quoting laboriously from Sprevak. For the original version, see (2009: 514). Nothing I say rests on the exact details of the counterexample. [↑](#footnote-ref-20)
21. See also Lewis (1980) for a seminal statement of this view. [↑](#footnote-ref-21)
22. Some people, like Hohwy and Rajan (2012), question the doxastic status of extremely irrational beliefs (delusions). But this fact should call us to question the claim here made. First, there are many who think it is wrong to deny the doxastic status of delusions (e.g. Bortolotti 2009). Second, one could deny the doxastic status of delusions while still maintaining that states which are irrational in a less extreme sense can still count as beliefs. Third, many of the arguments against the claim that delusions are beliefs rest on the fact that delusions are not only epistemically atypical, but atypical in other respects too (for discussion, see Bortolotti 2009). In the case I am imagining, the beliefs are epistemically atypical, but in all other respects typical. See also Schwitzgebel (2001) for the view that delusions might be cases of ‘in-between’ believing. [↑](#footnote-ref-22)
23. See Sterelny (2004) for discussion of a similar point in the relation to Otto’s notebook. [↑](#footnote-ref-23)
24. Note, importantly, that the relevant Martian cases are *not* total functional isomorphs of Otto/Blotto. For example Metamento, whose brain includes detachable/reattachable components, is not a perfect functional isomorph of Otto. Like Otto, Metamento falls foul of the ‘persistence and integration’ condition. But unlike Otto, Metamento (for all we are told) does not fall foul of other conditions, like the non-derived content condition. Because Metamento is not functionally identical to Otto, the intuition that Metamento counts as a believer does not automatically generalise to Otto. [↑](#footnote-ref-24)
25. My thanks to an anonymous reviewer for suggesting the line of response considered here. [↑](#footnote-ref-25)
26. It might seem as if he also has another option: he could treat his three conditions as jointly sufficient for non-occurrent belief, rather than as individually necessary. But, as Sprevak points out (2009, p.17), this strategy fails to achieve Clark’s aim, which is to prevent ‘Blotto’ cases. To actively block the iteration-to-absurdity of Clark’s argument with respect to some mental item *M*, you need to say: ‘such-and-such is *needed* for counting as *M* and bloat cases lack such-and-such’. For this you need necessary, rather than sufficient conditions. [↑](#footnote-ref-26)
27. For example, he has described his conditions as a ‘rough-and-ready set of additional criteria to be met by nonbiological candidates for inclusion into an individual’s cognitive system’ (2008a, p.79). See also Clark and Chalmers (1998, p.17). I do not analyse these quotes in detail because I am less interested in what Clark actually says on this score than in what he *needs* to say in order to keep his argument plausible. [↑](#footnote-ref-27)
28. I would like to thank Rob Hopkins, Dominic Gregory, Michael Wheeler, Stephen Laurence and Rosanna Keefe, who all gave extremely useful comments on earlier drafts of this paper (and invaluable advice on the publication process). Thanks also to the two anonymous reviewers, whose comments were very useful. And thanks to Steve Wright his proof-reading help, and to the audience at the Sheffield graduate seminar, whose comments were very helpful. [↑](#footnote-ref-28)