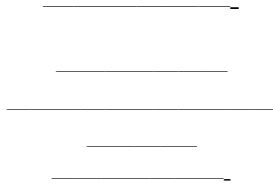# Why Some Version of the Extended Mind Hypothesis Must Be on the Table!

1   Introduction

Andy Clark and David Chalmers did not, of course, invent the extended mind hypothesis, versions of which can be found in enactive philosophy (Varela et al., 1991; Varela and Maturana, 1992), dynamic systems theory (Thelen and Smith, 1994), evolutionary psychology (Donald, 1993) and elsewhere. The idea has been around in one form or another for a long time and has its roots, among other places, in the semantic externalism of Hilary Putnam (Putnam, 1975)[1]. Nevertheless they have done more than anyone else to frame the contemporary debate and to show how various related lines of thought can usefully converge.

Clark and Chalmers open their now famous 1998 paper – a paper that has gone on to spawn something of a philosophical cottage industry – with a question: "Where does the mind stop and the rest of the world begin?" (Clark and Chalmers, 1998, p. 7) Although it is nowhere so well explicated as one might like, the concept of boundary is key. If *absolute fixed* boundaries of any kind are conceptually problematic – and I will argue that there is reason to think they are – then this one is in particular. Like many boundaries, its precise location may seem straightforward, until we look too closely and start finding borderline cases. This paper will offer a distinctive argument that the boundary between mind and world (or, more or less equivalently, between self and other or between self and non-self) is a pragmatic one that, as I believe Clark and Chalmers correctly assess, shifts over time with respect to needs and perspective.

I will take "boundary" in this context – perhaps controversially – to mean a categorical dividing line such that specific instances of entities for whom the boundary is relevant *must* fall to one side of the boundary or the other.[2] Any that fall directly on the boundary indicate a problem at the least with where the boundary has been drawn, if not with the boundary itself. I will argue that issues of boundary deserve a great deal more attention than they have mostly been given to date.

---

[1]Clark and Chalmers are, however, careful to distinguish their *active externalism* from Putnam's own passive externalism.

[2]Of course, many useful boundaries in other domains are not so sharply drawn, as one of the reviewers pointed out. However, I take the debate over the extended mind hypothesis to revolve, in large part, on whether the mind/world boundary *can* be so sharply drawn. If that boundary is sufficiently flexible, then Clark and Chalmers have made their point.

In particular, I will argue that considerations of the nature of concepts and conceptual abilities give us reason to think that these inevitably color our view on the world and, therefore, whatever lines we draw between ourselves and the world. Concepts I will take, roughly, to be those distinctive units of our structured thought that give those thoughts, *per* Gareth Evans' Generality Constraint (Evans, 1982, p. 100 ff.), *systematicity* (they can be deployed by a conceptual agent across uncountably many contexts without substantially changing their meaning) and *productivity* (a finite set of them can be used to produce a potentially unbounded number of complex concepts and, in linguistically capable agents at least, propositions). Concepts are compositional on any account: complex concepts can be built up out of primitive ones. Concepts are always *about* something – there are no empty concepts; and they always belong to an agent, under whose "endogenous control" (Prinz, 2004) they fall. For all of the debate over the precise nature of concepts and over such questions as whether concepts can change[3], these points I take to be non-tendentious.

The argument will proceed as follows:

1. An examination of the current state of the extended mind debate shows that such key figures as Rob Rupert, Frederick Adams, and Kenneth Aizawa are relying on strongly held intuitions; and more importantly, they are making assumptions based on metaphysical starting points that they are less than straightforward about (Section 3). To some extent Andy Clark and Dave Chalmers are as well, though Clark in particular is careful about hedging his bets.

2. If one does not accept these metaphysical starting positions, then the "just obvious" nature of many of the claims is much less obvious. In particular, their metaphysics predispose them to take concepts of inner and outer (internal/external) from the physical domain and apply them to cognition, where it might appear that such application is loosely metaphorical at best.

3. There are reasons independent of metaphysics for questioning the application of concepts like "internal" and "external" to the cognitive domain. However, if one allows anti-realism as at least a plausible position for sake of argument, then those concerns will be heightened.

4. Concepts provide a particularly fruitful area to focus on here, because they give us reason to believe that our concepts and conceptual prejudices affect all we experience (indeed, help to structure it), so that it really is meaningful to say in a McDowell(1996)-inspired way that concepts bleed "all the way out" into the experienced world (Section 4). In particular, there are sound logical reasons to believe that concepts and their referents cannot pull fully apart, relative to the perspective of any conceptual agent (who cannot step outside his own conceptual frame of reference).

5. If concepts bleed "all the way out" into the world, and if concepts are intrinsically cognitive entities, then cognition plausibly extends into the world as well (Section 5).

In *Supersizing the Mind* (2008), Clark talks about the ways in which the boundary we draw between the self and the non-self (self and other, self and environment) changes over time (in particular for any given individual, but perhaps collectively as well). This is to be expected if 3-4 are true and is, I argue, the best way of advancing a version of the Extended Mind Hypothesis. Yes, there are things that are clearly and untendentiously "self" and things that are clearly and untendentiously "non-self" (at least at any given point in time, but very often in a stable way over time as well). At the same time there are other, borderline cases that seem to fall on either one side of the boundary or the other depending on one's perspective at the time. The "inner"/"outer" "internal"/"external" distinction with respect to cognition is deeply problematic precisely because of this difficulty with boundaries.

---

[3]On Jerry Fodor's informational atomism account, for example (1998), concepts do not and, indeed, cannot change.

However, before proceeding to this positive account for advancing the extended mind hypothesis, I wish to address how I think the extended mind hypothesis should *not* be advanced. The rest of the paper is structured as follows. Section 2 addresses difficulties with the familiar Otto thought experiment and raises a familiar worry about "cognitive bloat". Section 3 presents my best understanding of the current state of the extended mind debate and raises questions about the driving intuitions and the metaphorical starting points of the various players. Section 4 sets metaphysical concerns aside to the side to suggest that, from a conceptual point of view, strict boundaries just are problematic entities, and certain boundaries (as between mind and world) even more so. It explores in some detail the argument that concepts and their referents do not and cannot logically pull *fully* apart. Section 5 summarizes the overall argument once more and offers some proposals for moving the extended mind discussion forward.

2   What's Wrong With Otto?

Clark and Chalmers introduce their readers to Otto and Inga, two New York residents who, for sake of the thought experiment, share a common belief (the location of the Museum of Modern Art) and a common desire (the intention to go there for a certain exhibition). For both of them the belief is meant to be *non-occurent*: so for example, Inga is not consciously thinking about the location of the museum before she hears about the exhibition, but we would still conventionally ascribe her the belief that the museum is on 53rd Street, based on her prior acquisition of that knowledge. Otto, also, is not consciously thinking about the location of the museum before he hears about the exhibition, but that is because, as an Alzheimer's patient with significant memory impairment, he cannot. Instead he has written the museum's location in the notebook that he carries everywhere with him, and where he records all information that might at some point prove important.

The only relevant difference between Otto and Inga is meant to be the location of the information, prior to conscious recall: conventional long-term memory for Inga, notebook for Otto. If Inga has a non-occurent belief, then so must Otto – or so the reasoning goes.

I find Otto and Inga a distraction for two reasons. The first is almost trivial: to the best of our current knowledge, Alzheimer's disease simply does not work that way. Alzheimer's is a degenerative disease, and although the rate of change may vary considerably from individual to individual and over the course of the disease, it is not known for reaching stable plateaus of the kind that would be needed. Routines like Otto's notebook take time to develop in all of us, never mind someone who is struggling with a disease like Alzheimer's.

The second concern is more substantial but relates to the first. In (Clark, 2008, p. 78), Clark attempts to clarify matters by giving Otto a "mild form" of Alzheimer's. But it is unclear whether this works; in initial stages of the disease (which one might suppose Clark is referring to), one simply does not have the *reliable* memory impairment that is critical to the hypothetical Otto's case. One might reasonably worry that Clark and Chalmers fail to pick out what is distinctive about hypothetical Otto: namely, the very *intimate* way he is bound to his notebook, the consulting of which seems to be the one thing he reliably can remember. He *cannot* consult his non-notebook-based memory of the museum's location on some semi-reliable basis, because then the notebook would be optional, and that would violate Clark's condition that "the resource be reliably available and typically invoked"(Clark, 2008, p. 79)[4].

---

[4]In case of doubt, Clark adds, tellingly, that "Otto always carries the notebook and won't answer that he 'doesn't know' until after he has consulted it."

On face value, most people simply are not dependent on their notebooks, laptops, mobile phones, and other such props, in anything like the same way. If Otto lost his notebook, he would, in a very real way, lose a part of himself: that seems plausible. In contrast, when the hard drive on someone's laptop crashes, they are generally and genuinely put out, but they are hardly transformed into a different person.

Clark would do well to remember where the term "supersizing" comes from: the American fast food industry; and what connotations the term had and has there. Over time the term became associated specifically with obesity and excess and more generally with saddling customers with what they did not need and did not really want. The moral of the story may be that a little extended mind goes a long way. Indeed, that is precisely the moral I will draw toward the end of this paper.

To be clear: I am not challenging the extended mind hypothesis – far from it! I am rather questioning the choice of supporting examples and, to some extent, the method of argument. Yes, Otto and his notebook would, if they existed, constitute a *prima facie* plausible candidate for an instance of extended mind; however, the example does not generalize in the way Clark and Chalmers want or, indeed, need it to. The extended mind hypothesis is a lot less interesting if it is limited to a few rare if not, indeed, merely hypothetical, cases. Worse, if it did generalize that way, issues of cognitive bloat raise their head. More on that in the next section.

3 Intuitions and Metaphysical Premises

Supersizing the Mind is in many ways an extended re-statement of the main ideas from the 1998 paper and response to the critics. It summarizes the extended mind hypothesis like this:

> Proponents of the extended mind story hold that even quite familiar human mental states (e.g., states of believing that so and so) can be realized, in part, by structures and processes located outside the human head. Such claims go far beyond the important but far less challenging assertion that human cognizing leans heavily on various forms of external scaffolding and support. Instead, they paint mind itself (or better, the physical machinery that realizes some of our cognitive processes and mental states) as, under humanly attainable conditions, extending beyond the bounds of skin and skull (Clark, 2008, p. 76).

The main points I take to be these:

- Any version of mind/brain identity (e.g., (Churchland, 1989)) is rejected. Mind is neither the same as nor reducible to brain; and in particular, they need not share the same boundary with respect to the world.

- It is important, but insufficient, to stress the rich interactions between any cognitive agent and its environment. At least some of those interactions are sufficiently rich as to blur the lines between the two.

It is telling that Clark refers to "scaffolding". It is the nature of scaffolding that it can be removed once the structure it is supporting is complete. The structures Clark believes to contribute to instances of extended cognition are *not* scaffolding precisely because they cannot be removed (at least, not without the cognitive agent becoming something substantially different from what it was). They become, in some non-trivial sense, part of the cognitive agent: or so goes Clark's argument.

3.1 Intuitions and Counter-Intuitions

Although considerable empirical evidence is cited, much of the argument for the extended mind hypothesis comes from an intuition pump commonly referred to as The Parity Principle (though it is not called that in the original paper): "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, p. 8).

Criticism of the Clark-Chalmers position has arisen most prominently from, on the one hand, Fred Adams and Ken Aizawa, who published a seminal paper in 2001 (Adams and Aizawa, 2001); and on the other, Robert Rupert, who published a similarly influential article in 2004 (Rupert, 2004). Borrowing a page from Clark, both have recently published books along the same lines as their earlier papers: (Adams and Aizawa, 2008) and (Rupert, 2009). In what follows I will make the most reference to Rupert's book, which is the more recent and makes many of the same arguments.

Adams, Aizawa and Rupert all claim sympathy for the Parity Principle. "To us, [the Parity Principle] means that the skull does not constitute a theoretically significant boundary for cognitive science. More specifically, it means that being inside the brain cannot be the mark of the cognitive. This seems to us true and obvious" (Adams and Aizawa, 2001, p. 46). "I sympathize with the motivation behind the Parity Principle. After all, why should it matter where a process takes place? If that process instantiates cognitive or mental properties when it is over *here*, why should things change simply because it is now over *there*?" (Rupert, 2009, p. 30)

All claim, however, a powerful *counter*-intuition. "...To many in the philosophical community, the extended view seems incredible on its face" (Rupert, 2010, p. 2). Aizawa is more colorful when he describes, in the introduction to *The Bounds of Cognition*, his introduction to the extended mind. "Some time in the early summer of 1998 or so, Fred came across a paper by Andy Clark and David Chalmers, advancing what seemed to us to be the outrageous hypothesis that, at least at times, cognitive processes extend into the tools people use" (Adams and Aizawa, 2008, p. vii).

Regardless – whatever value intuitions may have initially, they are – on Rupert's (or Adam's and Aizawa's) account – things ultimately to be set aside, like idiosyncrasies of perspective. "The average person's intuition-based applications of 'cognition', even the well-informed theorist's reactions, should not be trusted to reflect the *actual* structure of cognition – unless, of course, the subject has in hand the correct theory of cognition" (Rupert, 2009, p. 32). "After all, it's not up to our intuitive judgments to decide what cognition is; the property of being cognitive is a scientific construct, validated only by the causal-explanatory work it does" (Rupert, 2010, p. 20).

All three insist that the limitation of the mind to the boundaries of the body, if not the brain, are contingent and not *a priori*: i.e., no "fetish for the bodily boundary" (Rupert, 2009, p. 45). To wit, and *contra* Clark and Chalmers' reading, they claim that the extended hypothesis, in any meaningful form, flies in the face of nearly all empirical evidence, and that all of the explanatory work that the extended hypothesis might claim to do can be done as well or better simply by emphasizing the agent's *embodiment* (the agent takes a particular physical form, by which it interacts with its environment) and *embeddedness* (the agent is situated in a particular environment, with which it richly interacts).

Perhaps the biggest concern, which Adams, Aizawa, and Rupert again all share, is that – ironically, given the title of Clark's book – the extended hypothesis will lead to inevitable cognitive bloat.

Adams and Aizawa also call it "cognitive ooze". "These names do justice to the ugliness of the view, but not to its radical nature. The threat is of pancognitivism, where everything is cognitive. This is surely false" (Adams and Aizawa, 2001, p. 57). "Does the explanatory principle [on which the extended mind hypothesis might be grounded] entail that my house's state of being structurally unsound is partly located where the builders' corpses are? Or that the past mental states of these now deceased people are part of the physical substrate my house's property of being structurally unsound?" (Rupert, 2009, p. 20) "A human might design a computer, then drop dead. No one, I hope, would be tempted to say that, because of her explanatorily causal contribution to the computer's later processing, the designer's corpse partly constitutes that machine's computational processes" (Rupert, 2009, p. 139). Adams and Aizawa offer similar *reductios ad absurdum.*

Arguments back and forth have led to a succession of papers and books. But it does not seem to me that any of the principal parties to this debate have significantly moved on or changed their position. Instead they have often resorted to quite legalistic language reminiscent of lawyers arguing an obscure point of corporate law. Once one clears the legalese away, the differences – somewhat of emphasis, somewhat of substance – are fairly straightforward; and, unlike obscure matters of corporate law, they really do (and should!) matter to the rest of us.

3.2   Why Metaphysics Matter

It will be useful to begin a discussion of metaphysical premises by saying what the metaphysical disagreement is not. It is not, though it might sometimes claim to be, a debate about materialism or naturalism, or Cartesian dualism versus physical monism. All of the parties to this debate, with the notable exception of Chalmers[5], are keen to stress their materialist credentials: e.g., "In questioning BRAINBOUND[6], I shall not in any way be questioning the basic materialist vision of mind as emerging fully and without residue from physical goings-on" (Clark, 2008, p. xxviii).

Such posturing might seem tiresome since, in the absence of a clear and untendentious definition of the material or the physical, applicable in all contexts (i.e., not only at the so-called "macro" level) – and there seems to be none in sight – it is unclear what an immaterialist or anti-physicalist perspective is meant to look like. Solipsists and Berkleyans are notoriously hard to find. The real concern seems to be about Cartesian (substance) dualism versus substance monism, and here it is unclear whether Descartes himself (as wiser minds than I have been keen to note (Baker and Morris, 2002; Oakeshott, 1991, p. 22)) endorsed the view that bears his name. *Ceteris parabis*, one substance is preferable to two: well, yes. Offhand I do not believe in ghosts, either. But surely people are keen to be saying something more.

So if the metaphysical disagreement is not over materialism, or monism versus various forms of dualism, what is it about?

Rupert puts a lot of weight on the word "literal": the word is literally peppered throughout his book. In a typical passage he describes the extended view as "the view that human cognition – to some substantial degree – literally includes elements beyond the boundary of the human organism" (Rupert, 2009, p. 3). The implication, I assume, is that Clark and Chalmers are not "merely" speaking metaphorically: they "really" mean that. Such a crisp literal/metaphorical distinction – whether with one's language or one's concepts – assumes some form of realist metaphysics, as is

---

[5]Chalmers is an interesting case. Although subject to frequent claims of being a Cartesian dualist, Chalmers is avowedly *not* one (see e.g. (Chalmers, 1996)), preferring what he calls there (and elsewhere) a more "innocent" dualism that is meant to be fully compatible with an orthodox scientific world view and physical theory.

[6]BRAINBOUND is Clark's name for the position he contrasts with the extended view.

clear from many places in Rupert's writings (though nowhere have I seen it stated so baldly).[7] The literal meaning is the fact of the matter that realism aims to deliver. But there is nothing about physicalism (or materialism or naturalism), so far as I am aware, that entails realism. On this point Clark is holding his cards in his hands, while Chalmers is, so far as I can tell, best understood as both a physicalist (he rejects substance dualism) and an anti-realist (he takes experience as foundational: i.e., something that must be assumed from the beginning) (see e.g. (Chalmers, 1996)). Contrast Rupert with Peter Gärdenfors, whose anti-realist leanings inform his theories about concepts: "The upshot is that [in conceptual spaces theory] there is no sharp distinction between literal meaning and metaphor" (Gärdenfors, 2004, p. 187).

Realism I take to be the metaphysical assumption either (*per* direct realism) that the apparent transparency of the world should, in most instances at least, be taken at face value; or (*per* indirect realism) that if the apparent transparency cannot be taken at face value, it can, at least, be logically reconstructed. In either case, science talks about the world in a perspective-free (or essentially perspective-free) way. Meanwhile anti-realism is the position that, while the fully mind-independent world is conceded logically to exist, one cannot, as a matter of principle, say anything about it; or that the world we experience is always in some way touched by mind.

Let me be clear: anti-realism is not the perspective that world *is* mind; that would be idealism. Neither does anti-realism allow one to believe whatever one likes about the world: if the world constantly outruns our conceptual understanding of it, at the same time it constantly and forcibly constrains that understanding, sometimes on pain of injury or death.

Anti-realism, pragmatism, and pluralism go hand in hand, where pragmatism is taken as the position that there need be, in most instances at least (and perhaps all of any import), no single fact of the matter. Pragmatism can even tolerate apparent contradictions, so long as they are qualified by perspective: e.g., p from one perspective, $\tilde{p}$ from another. So long as one does not try to hold both perspectives at once – i.e., make them part of a single perspective – there is no contradiction in possessing both of them.

In this light, Rupert's statement that "even if one is inclined toward pluralism, an extended and an embedded model cannot both be true of a single cognitive process – else there is a single cognitive system that both extends beyond the boundary of the organism and does not" (Rupert, 2009, p. 9) is, on the face of it, simply wrong. Pluralism and pragmatism in no way guarantee that our conflicting perspectives can necessarily be reconciled[8].

I am arguing here neither for the correctness of anti-realism nor the falsehood of realism. It is enough to allow a modest anti-realism as a plausible position for sake of argument. If some form of anti-realism *should happen* to be true, then intuitions, like perspectives, cannot simply be set aside; they will play an unavoidable and substantive role in the theory.

---

[7]As one of the reviewers commented, in objecting to my portrayal of Rupert as a (direct) realist, Rupert is a representationalist. So far as I can see, there need not be no conflict between being a representationalist and being a direct realist, if the representations are suitably transparent: convenient mental shorthands, as it were. However, for my purposes, it is enough to note that there is *no* conflict between representationalism (in its many forms) and indirect realism, and all I need to show is that Rupert is assuming *some* form of realist metaphysics.

[8]Indeed, Rupert elsewhere might seem to allow this, given he acknowledges the likelihood that "different representations are used depending on context and thus that the subject represents what might normally be characterized as the same thing in different ways depending on context" (2009, p. 196). I see nothing in that passage to indicate either that one of the representations will be the "right" one or that the representations may not be in seeming conflict with each other.

3.3 Metaphysical Premises and Boundaries

Metaphysical premises become clearest when one looks at another word used ubiquitously by all parties to this debate, and the one with which I began this paper: namely, "boundary". After all, the extended mind debate at heart is about where one should draw the boundary between mind and world, and whether that boundary is fixed at the physical boundary of skin and skull. For all of the importance (rightly) placed on this boundary, one might expect there to be more attention paid not just to locating it correctly but determining its nature. Is the boundary "really" real, or is it something we construct (and can move)?

Overly rigid boundaries of any kind can prove problematic, if one probes them too closely. Consider cell boundaries, as clear of a boundary as one is likely to find. Any effective cell membrane must be permeable: a continuity to match the discontinuity. The problem is: at what precise point does a molecule pass from "outside" to "inside"? The closer one examines the cell boundary, the harder the answer becomes. The answer is only clear if one observes from a sufficiently detached perspective.

Boundaries at the level of multicellular organisms only become more difficult. Is my bodily boundary at my epidermis (layer of dead skin cells) or my dermis (live cells)? It depends upon the context in which you're asking. Likewise my body is, topologically speaking, torus shaped. Normally I think of my digestive tract as "inside", but from some perspectives it is "outside".

What of the bacteria living in my gut, who depend on me for their existence, and whom I likewise depend on for mine? Are they inside or out? Are they part of me or not? I am reliably informed that several kilograms of my body weight consist of single-celled organisms: some symbiotic, some neutral, some parasitic. When I weigh myself in the morning, I do not mentally subtract them.

The realist, of whatever persuasion, need not, of course, be bothered about any of this. *Prima facie,* it is enough for her to say e.g. that the boundary between mind and world is *roughly* at the physical boundary of skin and skull – or is it?

The difficulty (or, for the extended mind enthusiast, the opportunity) lies with how rough is "roughly", and in particular with the way Rupert (along with Adams and Aizawa) moves seamlessly from the boundary of the organism *as a biological agent* to the boundary of the organism *as a cognitive agent*. "Internal" and "external" are attributes of physical objects (or collections of such objects), and it is not immediately clear whether a cognitive agent is that sort of thing. Yet Adams and Aizawa write, "To ask about the bounds of cognition is to ask what portions of spacetime contain cognitive processing.... It is to ask about the physical substrate of cognition" (2008, p. 16).

This is, perhaps, not a problem, provided one sees mind reducing to brain *per* an eliminativist account (such as the Churchlands offer), or mind emerging from brain in a way that is either immediately transparent or reconstructably so. (The latter route is, I think, the one that Rupert wants to take: he seems ready to allow that mind *could* just be a functional description with no immediate physical translation.) Either mental boundaries "just are" physical boundaries, or they map straightforwardly to them.

Reconstructable in principle, however, need not mean reconstructable in practice; and herein lies the fruitful middle ground between anti-realist and realist perspectives: without that reconstructability, mental boundaries look woefully unclear. At the same time, a clear and at least relatively fixed mental boundary is essential to Rupert's arguments.[9]

---

[9]Contrast this with Clark's flexible notion of the same boundary, when applied to what he terms "profoundly embodied agents": "Such agents are able constantly to negotiate and renegotiate the agent-world boundary itself.

In keeping with much if not most of the literature in this field, Rupert talks about representations without defining what they are: the assumption is that the definition is already understood and agreed upon. Unfortunately, proceeding to label some representations as "internal representations", as Rupert does, does not help unless the application of "internal" in the mental domain is *also* already understood and agreed upon. Rupert's offer of a "systems-based criterion" (2009, ch. 3) is, on its own, no help recapturing a clear sense of boundary unless that criterion is assuming the very physical translation that is meant to be derived – no help, certainly, if one gives any weight to concerns like this one from Clark:

> Nontrivial causal spread... occurs whenever something we might have expected to be achieved by a certain well-demarcated system turns out to involve the exploitation of more far-flung factors and forces. (2008, p. 7)

One school of thought actively trying to occupy the middle ground is enactivism, as defined by Francesco Varela: "I have proposed using the term *enactive* to... evoke the idea that what is known is brought forth, in contraposition to the more classical views of either cognitivism or connectionism" (Maturana and Varela, 1992, p. 255). As does extended mind, enactivism – as used by Varela, Evan Thompson, John Stewart and others (Varela et al., 1991; Thompson, 2007; Stewart, 1995) – views cognition as spanning brain, body and environment (to paraphrase a phrase used by Clark and by many, many others). Like extended mind, enactivism includes but goes beyond notions of embeddedness and embodiment by:

- Understanding cognition, at least in the first instance, as a *skillful activity*;

- Stressing a continuity between agent and environment as underlying any conceptual distinctions between the two[10];

- Consequently presenting agent and environment as co-emergent;

- Emphasizing the ineliminable role of an observer, which is to say, a first-person perspective.

Although not explicitly enactive, Gärdenfors' conceptual spaces theory fits comfortably into such a perspective.[11] Clark, pointedly, has not endorsed enactivism. But perhaps he should.

4   The Argument from Concepts

One of the stock phrases in (Adams and Aizawa, 2001) and (Adams and Aizawa, 2008) is the "mark of the cognitive", the hallmark of which is "non-derived content". Regardless of what else does or does not bear the "mark of the cognitive", Adams and Aizawa must surely agree that concepts do. If concepts are not mental entities, what are? And if they are not mental entities, what are they?

Although our own capacity for such renegotiation is, I believe, vastly underappreciated, it really should come as no great surprise, given the facts of biological bodily growth and change." (2008, p. 34)

[10]Such an emphasis on continuity might on first blush seem at odds with the central importance and seeming inflexibility of boundaries – notably the cell boundary – to Maturana and Varela's notion of autopoiesis (see for example (Maturana and Varela, 1992, 1980)). This, I think, would be a misreading of Maturana and Varela. The membrane is essential, yes, but only relative to one's perspective as an observer; it is absolutely critical to their notion of autopoiesis that, from some other perspective, what matters is the *continuity* between organism and environment, each actively defining the other. That is to say, the boundary is only meaningful with respect to that continuity.

[11]... As Gärdenfors himself acknowledges (personal communication).

It is a trivial observation, almost a definitional truism, that our structured understandings of the world are conceptually mediated. The interesting question is whether or not concepts are transparent to the pre- or non-conceptual world (either immediately or reconstructably so). What is the relationship between clearly conceptual mind and seemingly non-conceptual world? To wit, is there a pre- or non-conceptual experiential Given that then rationally justifies our conceptually structured beliefs and experience? John McDowell, borrowing a page from Wilfrid Sellars (Sellars, 1956), calls this "the Myth of the Given": the idea that "experience, conceived in such a way that it could not be a tribunal, is nevertheless supposed to stand in judgment over our empirical thinking" (1996, p. xvii). In a famous phrase, McDowell writes that "...the idea of the Given offers exculpations where we wanted justifications" (McDowell, 1996, p. 8).

McDowell is likewise famous for saying that concepts go "all the way out". Though he is generally careful to qualify where they go "all the way out to" – e.g., "all the way out to the world's impacts on our receptive capacities" (McDowell, 2007, p. 338) – nonetheless, the practical consequence is generally taken to be that the world, itself, is somehow conceptual. So Jeremy Koons reads McDowell's attack on the Myth of the Given as meaning that "world is assimilated to mind: reality is in the space of reasons. Mind is able to represent world because world, like mind, turns out to be conceptual: reality is itself conceptual" (Koons, 2004, p. 130).

One need not embrace McDowell's conceptualism – the position that the content of experience is *fully* conceptual, fully within what he calls the "space of reasons" – in order to accept his attack on the Myth of the Given. That is to say, so long as concepts do not fully or reliably pull apart from their referents in the world, so long as some conceptual residual remains (or even may remain), then the Given will be a myth. More importantly for present purposes, one need not embrace any degree of anti-realism to allow the possibility (in practice!) of such residue.

4.1 Separating the Concepts from (the Rest of) the Content

> The prime problem is that the information received by the receptors is too rich and too unstructured. What is needed is some way of transforming and organizing the input into a mode that can be handled on the conceptual or symbolic level. This basically involves finding a more *economic* form of representation: going from the subconceptual to the conceptual level usually involves a *reduction of the number of dimensions* that are represented... (Gärdenfors, 2004, p. 221).

There is an important distinction to be made between concepts as we reflect on them, and concepts as we possess and employ them non-reflectively (which is, presumably, most of the time). What concepts seemingly "just are", when we reflect upon them, is representations (regardless of whether or not it is useful to qualify those representations as being "mental representations"). What they do, when we reflect upon them, is to represent (stand in place of) one aspect or another of the conceptual or the physical world we encounter.

No one would confuse the painting of a particular waterfall (itself a kind of representation) with the waterfall itself. On first blush, one would never confuse one's concept of a dog (as a general category) or of a particular dog (as a specific instance) with the dog itself, either, nor conflate the two together. And yet, I want to suggest, this is precisely what conceptual agents do all the time when they are employing concepts non-reflectively: when they are just getting on with using them. It is only when we reflect upon things that the concept and its referent pull apart – indeed, it takes a reflective mind to pull them apart; and even there, if we examine the distinction too closely, we invite contradiction or paradox.

Consider: there are concepts, and there are things that are not concepts, with a *prima facie* clear boundary between. But to reflect upon either is to bring them into the space of concepts: so now one has concepts of concepts (higher-order) on the one hand, concepts of non-concepts (first-order) on the other. But *that* reflection invites another layer of regress, and so on.

Concepts are the expectations that drive experience. Consider your pet dog Rover. Upon any specific Rover encounter, you bring a great deal of conceptual expectations to bear, whether you are reflectively aware of doing so or not (and, in the usual circumstances, you probably will not be). What you experience is an object, with all the expectations of e.g. object permanence that you have had from a very early age [12]; but not just any object: a dog, with all your expectations about dogs (that *ceteris parabis* they *do* bark, that they *never* purr or meow); but not just any dog: your dog Rover, whom you've raised since he was a puppy, who crawls into your bed every night, who you took to the vet's last Tuesday for de-worming. Whatever it is, the risk if not the certainty remains that the referent of "my dog Rover" is not, or is never just, the thing-in-itself, stripped of all conceptual shading. Even the concept "my dog Rover yesterday" is a generalization over, or an abstraction away from, many specific Rover moments over the course of the day.

To recognize Rover *as* Rover – or, more minimally, as a dog, or more minimally yet, as an object – is already to have passed beyond the possibility of any strictly in-the-moment, strictly non-conceptual experience. For it is in the nature of concepts that they step back from the present moment to take a wider view, seeing the present moment and context in the light of previous ones and others yet to come; abstracting away from any particular moment at the same time they are being applied back to a particular moment: the present moment of experience.

And yet our understanding of the distinction between x and our conceptual understanding of x is so basic as perhaps to appear trivially obvious. Developmentally it is not far behind the self/other distinction that, like object permanence, seems to have its origins in the first few months of life and is almost certainly a common feature of all agents we might want to call conceptual.[13] It is arguably flawed: not so much because of any problem with the distinction itself but because of the assumption – seemingly unavoidable – that we can, at least most of the time, make it reliably, without interference from conceptual prejudices.

The heart of the problem is the very transparency of concepts: the way they appear to present the world uninterpreted. As the Gärdenfors' quote suggests, conceptualization is a *lossy* compression algorithm, simplifying the "too rich and too unstructured" "information received by the receptors." The price of this simplification is generally, if not universally, a "reduction of the number of dimensions that are represented": that is, the reduction in detail is not just quantitative but qualitative. The original version of something, in all its detail, cannot be recovered or reconstituted from the simplified version – however the simplification is done – without additional (external) input. Yet, consistently, we take the simplification for the (lost) original *and not merely as an approximation*. In this way, concepts are what one might reasonably call "necessary fictions".

4.2 The Nature of the Enterprise

The conceptual difficulty if not impossibility of disentangling mind from world is to be expected if one looks at the nature of the enterprise, within the philosophy of mind, that is concept studies. As with consciousness studies, as with any of the sciences of the mind, theories of concepts take

---

[12] Jean Piaget, who coined the term, famously located this ability at nine months (1954); more recent research (e.g. (Ballargeon, 1987)) has shown reliable evidence for an expectation of object permanence at less than half that age.

[13] For a discussion of so-called animal concepts and why one might want to believe in them, see (Allen, 1999) or (Newen and Bartels, 2007).

empirical study of the external world – the more common domain of science – and turn it around, to focus attention on ourselves; and not just any aspect of ourselves, but that aspect that seems most essential to making us who we are: our minds, and our structured thoughts.

We can lose any other aspect of ourselves – an arm, or both legs, or even portions of our brain as happens in a stroke – and still feel, with apparent justification, that we are the same person. But if we lose our mind, then we really have lost ourselves. Rupert, given his concern for the "asymmetric relations between the persisting organismic portion of the purportedly extended cognitive system and the system's external portions" (Rupert, 2009, p. 106), would do well to remember this asymmetry, between our cognitive and biological identities.

The sciences of the mind are unique in this way: the lines between observer and observed become significantly *and unavoidably* blurred. I cannot speak of mind without, implicitly if not explicitly, intimately including *my* mind (for my mind is the mind I am presumably best acquainted with, and in any case am using to advance the theory). This line of thought – with its consequent limitations on what a science of mind can achieve – was explored by F.A. Hayek more than half a century ago (1952). Although quantum physics theory has told us for some time that the observer cannot be removed from the equation, nonetheless much of science has proceeded on the basis that, for all practical purposes, the observer and any bias she introduces can be safely disregarded, and a "pure" objectivity achieved.[14] At the least, observer and observed stand (or appear to stand) clearly apart. Metaphysics aside, for good practical *and* logical reasons, such a distancing seems not to be possible in studying concepts: we cannot set our conceptual nature aside. When we talk about concepts or when we offer a theory of what concepts are, we do so inevitably as (conceptual) observers of the very concepts we are postulating about, speculating about a conceptual system from within that conceptual system.

5   Conclusions

I can now return to the question with which I (and Clark and Chalmers) began: "Where does the mind stop, and the rest of the world begin?" After all, one has to draw the line somewhere: otherwise, if everything is (equally) cognitive, the term loses any usefulness. If the keyboard I type with might, plausibly, bear the "mark of the cognitive", Alpha Centauri should not.

As hopefully should be clear, I share many of Rupert's, and Adam's and Aizawa's, concerns about cognitive bloat. When one "supersizes" the mind, one should be careful not to go too far. Even if the keyboard might, in some circumstances, become part of my cognitive system, it should not do so too easily.

Besides the Parity Principle, the most well-known intuition pump for the extended mind hypothesis is the Otto-Inga thought experiment. If it works, I have argued, it is only because Otto is a very rare individual, in a way that Clark and Chalmers don't address. As Rupert writes, "Even if there are some cases like Otto's, they seem too rare to drive a paradigm shift in cognitive science" (2009, p. 90). If, on the other hand, Otto's case applies more generally, so that notebooks and pens and keyboards routinely become part of "literally" cognitive systems, then this seems to threaten cognitive bloat.

---

[14]"In order to be able to give a satisfactory account of the regularities existing in the physical world the physical sciences have been forced to define the objects of which this world exists increasingly in terms of the observed relations between these objects, and at the same time more and more to disregard the way in which these objects appear to us." (Hayek, 1999, p. 2)

The moral of the story is that a little extended mind may go a long way; and all that this requires, I have argued, is a sufficiently flexible boundary between mind and world: one that shifts not too much but *just enough*. An over-concern with the "literal" truth of the matter reflects metaphysical agendas. Metaphysical premises may lead one to see that boundary as more rigid (or, for that matter, more fluid) than experience suggests it is, and to downplay experienced reality in favor of the "literal" facts of the matter. In this light, it is revealing how Rupert dismisses the relevance of such experience:

> Given a lack of agreement between different sources of information, the brain's drive to construct a single consistent model allows one source of information to dominate, even if the resulting model of the situation diverges from reality. This kind of process, not the literal extension of the body, self, or the cognitive system, explains the experience of an extended nose (Rupert, 2009, p. 166).

If the boundary between mind and world is, ultimately, a pragmatic and flexible one, then where, as a practical matter, should it be located? It seems reasonable to suggest that mind stops where the sense of self does: i.e., with our sense of agency and control. Rupert himself shows some favorability toward such a view, though notably minus the dimension of *personal* control (2009, p. 169).

If concepts bleed "all the way out" into the world, and if concepts are intrinsically cognitive entities, then cognition plausibly extends (in some meaningful sense) into the world as well. One need not maintain – as McDowell is often read – that the world we encounter is *fully* conceptual[15], only that there is no part of that world that is fully or reliably free of the conceptual touch. Such cognitive tentacles into the world are all that the extended mind hypothesis requires.

References

Adams, F. and Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14(1):43–64.

Adams, F. and Aizawa, K. (2008). *The Bounds of Cognition*. John Wiley and Sons.

Allen, C. (1999). Animal concepts revisited: The use of self-monitoring as an empirical approach. *Erkenntnis*, 51(1):33–40.

Baker, G. and Morris, K. (2002). *Descartes' Dualism*. Routledge.

Ballargeon, R. (1987). Object permanence in 3 1/2- and 4 1/2-month-old infants. *Developmental Psychology*, 23(5):655–664.

Chalmers, D. J. (1996). Facing up to the hard problem of consciousness. In Hameroff, S. R., Kaszniak, A. W., and Scott, A., editors, *Toward a science of consciousness: the first Tucson discussions and debates*, pages 5–28. MIT Press.

Churchland, P. S. (1989). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. MIT Press.

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

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

Donald, M. (1993). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press.

---

[15]That might, indeed, lead to cognitive bloat.

Evans, G. (1982). *Varieties of Reference*. Clarendon Press.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Clarendon Press, Oxford.

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. Bradford Books.

Hayek, F. (1952). *The Sensory Order*. Unive.

Hayek, F. A. (1999). *The Sensory Order: An Inquiry Into the Foundations of Theoretical Psychology*. University of Chicago Press.

Koons, J. R. (2004). Disenchanting the world: Mcdowell, sellars, and rational constraint by perception. *Journal of Philosophical Research*, 29:125–152.

Maturana, H. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living (Boston Studies in the Philosophy of Science)*. Kluwer Academic Publishers.

Maturana, H. R. and Varela, F. J. (1992). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala, London.

McDowell, J. (1996). *Mind and World*. Harvard University Press, Cambridge, Massachusetts.

McDowell, J. (2007). What myth? *Inquiry*, 50(4):338–351.

Newen, A. and Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical Psychology*, 20(3):283–308.

Oakeshott, M. (1991). *Rationalism in Politics and Other Essays*. Liberty Press, Indianapolis, Indiana, USA.

Piaget, J. (1954). *The Construction of Reality in the Child*. Basic Books, New York.

Prinz, J. (2004). *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT Press.

Putnam, H. (1975). The meaning of 'meaning'. In Gunderson, K., editor, *Language, Mind, and Knowledge*. University of Minnesota 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 University Press, New York, NY.

Rupert, R. (2010). Supersizing the mind: Embodiment, action, and cognitive extension, reviewed by robert d. rupert. upcoming.

Sellars, W. (1956). Empiricism and the philosophy of mind. In Feigl, H. and Scriven, M., editors, *Minnesota Studies in the Philosophy of Science*, volume I, pages 253–329. University of Minn.

Stewart, J. (1995). Cognition = life: Implications for higher-level cognition. *Behavioural Processes*, 35(1-3):311–326.

Thelen, E. and Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, MA.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. Harvard University Press.

Varela, F. J. and Maturana, H. (1992). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.