

# Extended Cognition and Extended Consciousness\*

David J. Chalmers

It's a pleasure to contribute this paper to *Andy Clark and his Critics*. Andy has been an influential figure for me ever since I read his first book *Microcognition* while I was in graduate school. A few years later I took up my first job as a post-doctoral fellow working with him in the newly initiated Philosophy-Neuroscience-Psychology program at Washington University in St Louis. We had a lot of adventures, but the one that seems to have left the most traces is our joint paper on "The Extended Mind", which argued that cognitive processes can extend outside the brain and the body to include objects that we interact with in our environment.

I recently came across hard copies of two early drafts of "The Extended Mind" in my files. The first draft was by Andy on his own, and is covered in handwritten comments by me. The second draft is a co-authored version with revisions by me, and is covered in handwritten comments by Andy. (I have put these two online at [consc.net/papers/e-drafts.html](http://consc.net/papers/e-drafts.html).) There were many further iterations after that. The whole thing is a wonderful record of distributed and extended cognition, involving a complex cognitive process spread between Andy, me, various notes on paper, and computer files. I'm proud to have played a role in constituting this process. That said, I have to acknowledge that as with many cases of distributed cognition, there was an individual at the core who set the whole process into motion and who bears primary responsibility, and that individual was Andy.

I am still very much sympathetic with the extended mind thesis, though I also have some doubts about its truth that I articulated in my foreword to Andy's 2008 book *Supersizing the Mind*. I will return to those doubts shortly, but here I will focus especially on two other issues. First, what exactly is the extended mind thesis? Second, are there cases of extended consciousness, and if not, why not? In addressing the first issue, I will be criticizing and revising the statement of the thesis in our original article. In addressing the second issue, I will be criticizing and revising a thesis that both of us have endorsed separately in more recent work. So these remarks will to some

---

<sup>0</sup>To appear in (M. Colombo, E. Irvine, and M. Stapleton, eds.) *Andy Clark and His Critics*.

extent criticize Andy, as the title of this volume requires. But really I am in the situation of a split brain hemisphere criticizing a previous whole brain of which it was a part, as well as criticizing both itself and the other hemisphere on their own.

## 1 What is the Extended Mind Thesis?

What exactly is the extended mind thesis? In “The Extended Mind”, we never quite state a single official thesis. We say that we advocate an active externalism based on the active role of the environment in driving cognitive processes. We say that parts of the world can be parts of cognitive processes and that cognitive processes are not all in the head. We argue that some mental states, including especially beliefs, can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes.

If there is an official thesis in the article, it is something like this:

A subject’s cognitive processes and mental states can be partly constituted by entities that are external to the subject, when those entities play the right role in driving cognitive processes.

To unpack this a little: the thesis applies to both cognitive processes and to mental states, which we separate in the article, since it is not obvious that the extension of one implies the extension of the other. “Partly constituted” is naturally understood as what is often now called partial grounding, though perhaps one might want to allow weaker relations in order to accommodate property dualist versions of the extended mind thesis in which mental states depend nomologically on features outside the head. “Entities” can be (perhaps inter alia) objects, instantiated properties, or states of affairs whereby objects instantiate properties. As for “external”, at various points we appeal to the boundaries of brain, body, head, skin, and skull. Each boundary leads to an interesting thesis, but perhaps to distinguish the extended mind thesis from various embodied mind theses, it makes sense to use the body as the key boundary.<sup>1</sup> Finally, “the right role in driving cognitive processes” is required to distinguish our externalism from that of Putnam, Burge, and others. The article gives only limited guidance about just what roles count as the right ones, but minimally what is needed is an active and ongoing role, as opposed to a passive and distal role.

---

<sup>1</sup>I am on record ([www.youtube.com/watch?v=FhybIsN5IZc](http://www.youtube.com/watch?v=FhybIsN5IZc)) as once endorsing the thesis that consciousness is in the hair, which is presumably outside the boundaries of brain, skin, and skull, but still part of the body, and perhaps a borderline case of being in the head?

Something like this is a fairly standard formulation of the extended mind thesis. But there is good reason to doubt that it is the right formulation. This reason has been touched on by various theorists over the years in formulating versions of the thesis, but it has been pressed most forcefully as an objection by Katalin Farkas (2012).

The objection is roughly this: as formulated, the thesis is too weak to be interesting. In particular, it is near-obviously true in *extended circuit* cases: cases in which a brain circuit is replaced by an external silicon circuit that is connected directly (e.g. by wiring or radio transmitters) to the rest of the brain so that it plays the same role as the original brain circuit.

In his reply to Jerry Fodor's review of *Supersizing the Mind*, Andy himself spells out such a case, the case of Diva:

But now imagine a case in which a person (call her Diva) suffers minor brain damage and loses the ability to perform a simple task of arithmetic division using only her neural resources. An external silicon circuit is added that restores the previous functionality. Diva can now divide just as before, only some small part of the work is distributed across the brain and the silicon circuit: a genuinely mental process (division) is supported by a hybrid bio-technological system. That alone, if you accept it, establishes the key principle of *Supersizing the Mind*. It is that non-biological resources, if hooked appropriately into processes running in the human brain, can form parts of larger circuits that count as genuinely cognitive in their own right.

Andy says here, in effect, that if one accepts that an extended circuit like Diva's can partially constitute a mental process (call this the extended circuit thesis), one has thereby accepted the key thesis of his book. Furthermore, if one accepts the extended circuit thesis, one has more or less accepted the official thesis of "The Extended Mind" as described above: mental processes will be partly constituted by something external because of the active role it plays in cognition. Now, some later remarks (which I will discuss shortly) suggest that Andy may think of his book's thesis as something distinct from the original extended mind thesis. Nevertheless, all this tends toward a view on which the extended circuit thesis suffices for extended cognition.

Farkas says, and I agree, that this cannot be quite right. The extended circuit thesis is too weak for it to support the interesting and controversial version of the extended mind thesis. One way to bring this out is to observe that many who take themselves to be opponents of the extended mind thesis explicitly accept the extended circuit thesis. Adams and Aizawa (2008) and Rupert (2007), who have mounted perhaps the most sustained opposition to the extended mind thesis, are happily

to accept the extended circuit thesis: they agree that in science-fiction cases such as Diva's, there can be extended cognition. This suggests that the official formulation of the extended mind thesis does not really capture what is centrally at issue in the debate.

The point can also be brought out by observing that the Diva case is quite different from the key cases in "The Extended Mind", such as the Otto case and the Tetris case. Most obviously, the Otto case and the Tetris case use only existing technology, where the Diva case uses hypothetical future technology. The first two cases involve ordinary perception and action, whereas the Diva case requires science-fiction-style extended circuitry. Furthermore, compared to the other two cases, the Diva case is much more clearly functionally isomorphic to non-extended cases of cognition. Partly as a result, the claim of extended cognition in the Otto case and the Tetris case is far more controversial than the corresponding claim about the Diva case.

Now, Andy and I could stand our ground and stick with our stipulated definition of the extended mind thesis, so that Adams, Aizawa, Farkas, and Rupert all count as supporters of the thesis. That would be a little akin to the US declaring victory in Vietnam and going home. I think it makes more sense to find a stronger formulation of the extended mind thesis that captures what is really at issue in the debate. This stronger thesis should be one that is not supported by a verdict of extended cognition in the Diva case, but that is supported by such a verdict in the Otto and Tetris cases.

Various stronger theses have been suggested in the literature, but no thesis that I have seen is ideal. Adams and Aizawa (2007) distinguish a modal version of the thesis (weak), holding that there *could* be cases of extended cognition, from a nonmodal version (strong) saying that there *are* cases of extended cognition. But this makes the strong thesis uncomfortably contingent (if no one used notebooks and the like, the thesis would be false?). One will also need to index the thesis to times as well—so that if certain opponents are right, the 2017 version of the thesis is false but the 2117 version of the thesis will be true. (My colleague Ned Block likes to say that the thesis was false in 1995 when we wrote the article, but it has since become true with the advent of smartphones and the like.) But I do not think that a thesis about the years 1995 or 2017 really captures what is of deepest philosophical interest in the article.

A related move distinguishes science-fiction from non-science-fiction cases, so that the strong thesis says there can be externally constituted cognition without science-fiction resources, while the weak thesis allows science-fiction cases such as extended circuits to count. Both the strong and weak theses here are modal theses, but the modality in the strong thesis invokes only worlds fairly close to the actual world. I think something like this was what Andy and I had in mind in

writing the paper. In recent work (2010), Andy appeals to a version of this criterion: “in fairly easily imaginable circumstances – ones that involved no giant leaps of technology or technique—we would be justified in holding that certain mental and cognitive states extended”. But the science fiction of today is the non-science-fiction of tomorrow, and a giant leap from today is standard tomorrow, so as before this formulation ties the extended cognition thesis uncomfortably to specific times. Perhaps there is something philosophically important about the difference between present technology and future technology, but if so a good formulation of the thesis ought to capture the philosophically important distinction directly.

Farkas (2012) suggests articulating a stronger thesis specifically in terms of standing states such as beliefs. Her suggested thesis says that the typical role of standing states can be extended to include states that produce conscious manifestations in a somewhat different way than normal beliefs and desires do. I think that while this thesis follows from our analysis of the Otto case and is interesting in its own right, it is not a good candidate to be the core extended mind thesis that is at issue. For a start, the thesis does not say anything directly about extension. Furthermore, the core thesis should certainly cover mental rotation in the Tetris case, but mental rotation is not a standing state so is not covered by Farkas’s thesis. Many wish to apply the extended mind thesis to other non-standing states such as perception, emotion, and the like— perhaps they are wrong to do so, but I do not think their view should be ruled out by definition. So we need to look elsewhere for a core extended mind thesis.

I suggest the following. What is distinctive about the extended mind thesis is that cognition can be extended in a certain way: via perception and action. In the Otto case, his belief is extended in virtue of his interactions with the notebook by perception and action (he sees the notebook, and he writes in it). In the Tetris case, mental rotation is extended by the subject’s seeing the tiles and pressing the button. In the Diva case, by contrast, Diva’s belief is extended by wires or radio transmitters that do not involve perception or action—instead the extension is a subpersonal process.

This leads to the following formulation of the extended mind thesis:

A subject’s cognitive processes and mental states can be partly constituted by entities that are external to the subject, in virtue of the subject’s interacting with these entities via perception and action.

The “interacting via perception and action” clause plays the key role of counting the Otto and Tetris cases as relevant while counting the Diva case as irrelevant. It captures the plausible

core of the nonmodal and non-science-fiction construals of the thesis: what is important about existing and non-science-fiction modes of extension is that they all involve extension through the mundane means of perception and action, rather than extension through novel means such as radio transmitters and brain wiring. It also captures what is distinctive to the role of extended standing beliefs: unlike ordinary beliefs, these produce their manifestations via perception and action. As a bonus, this clause can also replace the “right role in actively driving cognitive processes” clause in distinguishing the thesis from Putnam/Burge externalism, which do not turn on interaction of this sort.

I think this definition is at least extensionally correct in capturing paradigmatic cases of the extended mind. All such cases seem to involve interaction via perception and action. An opponent could try to construct cases of cognitive extension involving neither a perception-action link or an extended-circuit link: perhaps brain-computer interfaces, or intravenous stimulation. But insofar as these are not perceptual and involve subpersonal processes they seem closer in kind to extended circuit cases than to paradigmatic cases. One could likewise try to find cases of perception-action extension that do not seem like paradigmatic extended cognition: cases involving unconscious or unattended perception and action are a possibility here, partly because the link here is at least moving in the direction of the subpersonal character of the extended-circuit case. I think these cases are nevertheless interesting and controversial enough to count as vindicating the extended mind thesis, but if one wanted to exclude them, one could always restrict the notions of perception and action at play in the thesis accordingly.

Furthermore, I think the definition is intensionally correct in capturing what we were after. In “The Extended Mind” we talked of “epistemic actions” as playing a key role in constituting extended cognition. As for perception, a little later in Andy’s response to Fodor, he says:

If you next gently alter the details so that the device communicates with Diva’s brain through Diva’s sense organs (piggybacking on existing sensory mechanisms as cheap way stations to the brain) you end up with what David Chalmers and I dubbed ‘extended minds’.

This suggests that Andy too sees the use of sensory mechanisms, as opposed to the mere extended circuit found in the Diva case, as crucial to cases of the extended mind. Putting these pieces together, I think we are both at least tacitly committed to extended-mind cases as involving something like a perception-action link.

A complication here is that at a couple of points in the original article, we suggest that in cases of extended cognition, what looks like perception and action may not really be perception and action. For example:

In a very real sense, the rearrangement of tiles on a rack is not part of action; it is part of thought.”

and

“From the standpoint of [Otto’s extended] system, the flow of information is not perceptual at all... It is more akin to information flow within the brain.”

Now, as I said in my foreword to *Supersizing the Mind*, our points here were something of a stretch. Extended cognition or no extended cognition, it seems clear that Otto is seeing the notebook (and thereby perceiving), and that someone is moving the tiles (and thereby acting). Still, to finesse this point we might restate the strong extended mind thesis in terms of sensorimotor interaction, rather than action-perception interaction. Presumably no one will deny that that there is sensory and motor interaction in these cases, whether or not these constitute genuine perception and action. This leads to an alternative statement of the revised extended mind thesis:

A subject’s cognitive processes and mental states can be partly constituted by entities that are external to the subject, in virtue of the subject’s sensorimotor interaction with these entities.

This seems to me to be the strong and interesting thesis that is at the heart of the extended mind idea. I invite Andy to join me in accepting it as the retrospectively official statement of our thesis.

## **2 Interlude: The Perception-Action Objection**

Another advantage of this formulation is that it captures what is genuinely controversial about the extended mind thesis, and it leads directly to what I see as the most important objection to the thesis.

There have been many prominent objections to the thesis over the years, but few of them seem strong to me. They include: (1) extended “cognition” has a different psychological profile from ordinary human cognition and so isn’t genuine cognition (reply: this leads straight to

an implausible human chauvinism about cognition); (2) the content of extended items is not “intrinsic” but is derived from relations to a human cognizer (reply: likewise, the content of items in biological memory is derived from relations to conscious states); (3) the extended mind thesis commits a coupling-constitution fallacy (reply: this confuses the thesis with the argument for the thesis, which turns not on a coupling-constitution inference but on the better-justified parity principle); (4) extended states are not truly mental as they are not conscious (reply: if they are as mental as ordinary standing beliefs, that is good enough for us); (5) extended cognition leads to “cognitive bloat” where we believe everything on Google (reply: one can reduce the bloat with an endorsement-on-retrieval criterion and perhaps even further with a past-endorsement criterion, and happily accept any residual bloat as a correct-if-surprising consequence of the thesis).<sup>2</sup>

In my foreword to *Supersizing the Mind*, I set out what I see as by far the strongest objection to the thesis.<sup>3</sup> The objection is that perception and action constitute a principled boundary to the mind. On this view, the mental is constituted by processes between perception and action, so that what falls outside these boundaries cannot constitute the mental. When one combines this with the observation above that the extended mind thesis requires constitution by external entities that the subject interacts with via perception and action, and that therefore falls outside these boundaries, the extended mind thesis is ruled out immediately. Even if one works with the sensorimotor version of the thesis, there remains some plausibility in the claim that sensory and motor processes constitute a principled boundary for the mental. If so, this will suffice to rule out extended cognition.

I think this objection captures the force of much intuitive resistance to the thesis. Otto’s notebook seems problematically external not because it is outside the skin or the skull, but it is connected to the organism only through perception and action. External circuits such as Diva’s do not seem problematically external precisely because they are connected to the organism in a different way that allows them to fall between the boundaries of perception and action.

Now, I do not think this is a knockdown objection, and in the 2008 piece I outlined what I see as the best way of replying: reject the proposed boundary even at cost of some revisionism, since the important explanatory roles of mentality need not respect the perception-action distinction.

---

<sup>2</sup>(1) Rupert 2009; (2) and (3) Adams and Aizawa 2008; (4) Gertler 2007; (5) Clark and Chalmers 1998 and many others.

<sup>3</sup>I recall Philip Pettit making a version of this objection (concerning intentional action) in the discussion period after my first-ever oral presentation of “The Extended Mind”, at the International Conference on Logic, Methodology, and the Philosophy of Science in Florence in August 1995. I cannot recall seeing anyone make it in print in the intervening years, though I would not be surprised if it appears somewhere in the literature and I would be interested in pointers.



Perhaps surprisingly, extended belief states can function in explanation very much as non-extended states belief states do, even though their connection to rest of the organism involves perception and action.

But in any case, seeing things in terms of perception and action helps to capture what is genuinely interesting and controversial about the extended mind thesis. For opponents, it helps to clarify a principled line of opposition to the thesis: states that are available to the organism only via perception and action are not genuinely mental. For proponents, it also helps to clarify what is radical and compelling about the thesis. The thesis does not just overthrow the hegemony of skin and skull as boundaries for cognition (as we claimed at the end of the original article). It also overthrows the hegemony of perception and action.

### **3 Extended Consciousness**

Finally, what about extended consciousness? In “The Extended Mind”, Andy and I allowed that conscious experience might depend only on the internal state of the organism, and in later work both of us are doubtful about extended consciousness. In part, this is simply because it is hard to find potential cases of extended consciousness that are nearly as plausible as the cases of extended belief while meeting the required criteria. Many have tried to find such cases, but no cases have won much support.

Of course if we allow extended circuit cases such as Diva, it is easy to find cases of extended consciousness. We need only ensure that the neural correlates of consciousness are transposed into an extended circuit outside the head. Farkas (2012) notes that this gives rise to an apparent tension in Andy’s view: Diva cases suffice to establish extended cognition, but not to establish extended consciousness? The tension can easily be avoided, though, if we require the stronger perception-action link for both extended cognition and extended consciousness. Then the Diva case will not be a case of extended consciousness, since the extended circuit does not involve a sensorimotor interaction.

If extended consciousness is impossible, it is natural to ask for an explanation of why it is impossible. In the 2008 piece I briefly suggested an explanation, which Andy subsequently developed in much more depth. That explanation appealed to the thesis that consciousness requires high-bandwidth access to information, and that extended processes (at least realistic processes involving perception and action) all involve only relatively low-bandwidth access to information.

Karina Vold (2015) has argued against this explanation, and I am inclined to think she is

correct. A key point is that vision is actually an extremely high-bandwidth process. An enormous amount of information is transmitted via electromagnetic radiation from objects to the eye. Once the process hits the eye, there is a drastic reduction in information flow, which is reduced further once the process makes it to the brain. So it is not really true that perception has a lower bandwidth than the internal core of consciousness. Perception often transmits a large amount of information to the brain very fast. It is true that attentional bottlenecks pose some limits on how much of this information can be processed in a way that makes it available to cognition, but the same is true for internal information too. And if there is a relevant pathway for internal information that does not require attention in order to give rise to consciousness, it is hard to see why that pathway should not be available for perceptual information too.

To explain why there is no extended consciousness, something more is needed. I think the right explanation is not that consciousness requires high-bandwidth access to information, but that it requires relatively *direct* access. This is a fairly familiar point. In *The Conscious Mind* and in “On the Search for a Neural Correlate of Consciousness”, I argued that consciousness correlates with *direct availability for global control*, and that the physical processes that correlate with consciousness are those that support availability of this sort. Various others have argued for something similar: for example, Tye (1995) suggests that consciousness requires representational states that are “poised” for control of reasoning and action.

When this thesis is combined with the perception-action criterion for extended consciousness, it yields a suggestive explanation of why extending consciousness is so difficult. Internal brain processes support information that is directly available for global control. Processes that are extended via perception and action do not. These processes support information that is only indirectly available for global control: in order to be used in control, it must travel causal pathways from object to eye, from eye to visual cortex, and from visual cortex to the loci of control. By contrast, the internal neural correlates of consciousness need only travel some portion of the third pathway, from certain intermediate areas of the brain to the loci of control. Since consciousness requires direct availability, extended consciousness is impossible.

Again, there is no problem in an extended circuit case (which Vold uses to argue that extended consciousness is possible). The direct availability criterion provides no obstacle in these cases, since the extended circuit may support relatively direct access to information. But we have seen that the interesting sort of extended cognition requires that the interaction (and so the availability) is mediated by perception, and perceptually mediated availability is indirect availability.

Vold (2015, pp. 21-22) acknowledges the possibility of an explanation in terms of direct avail-

ability, and responds that “the access that a biological perceptual system has to information from the environment is no more direct than an extended perceptual system, including a pair of glasses or corrective lenses”. This may be true, but the key contrast is not between two modes of access to information present in the environment. Rather, it is between control systems’ access to information present in the brain compared to their access to information present in the environment. Access to the latter is mediated by perception where access to the former is not. So the latter is more directly available than the former.

It is useful to compare the case of conscious experience with that of standing beliefs. Both plausibly require some sort of availability of information for reasoning and for control. But even prior to the extended mind, we had good reason to think that conscious experience requires a stronger sort of availability. One reason is that standing beliefs plausibly require some sort of availability to consciousness itself: for example, availability to bring about a conscious thought or an episode of conscious reasoning. If consciousness itself requires availability for control, standing beliefs will then involve a two-step sort of availability: first to consciousness, then to control.

It is in effect this two-step availability that is exploited in the Otto case. Perception makes information in Otto’s notebook available to consciousness, from where it is available to control action. If the extended mind thesis is correct, the two-step availability required for belief can be perceptually mediated. But the one-step availability required for consciousness is more strongly constrained, and it is plausible that it cannot be perceptually mediated. If so, this explains why there are no cases of extended consciousness involving a perception-action link.

I do not think this is any sort of knockdown argument against extended consciousness. Proponents of extended consciousness can simply deny that direct availability is needed for consciousness, or they can argue that the right sort of direct availability can be present in extended processes. But if one has independent reason to doubt that there is extended consciousness, for example by noting that there seem to be no plausible cases of it, the account in terms of direct availability can provide a reasonable explanation of why this is.

I conclude that even if perception and action do not provide a principled boundary for cognition, they do provide a principled boundary for consciousness. Given that the sort of extension at issue is understood in terms of perception-action interaction, this explains why even if there is extended cognition, there is no extended consciousness.

## References

Adams, F. and Aizawa, K. 2008. *The Bounds of Cognition*. Wiley-Blackwell.

- Chalmers, D.J. 2008. Foreword to *Supersizing the Mind*. Oxford University Press.
- Clark, A. 2008. *Supersizing the Mind*. Oxford University Press.
- Clark, A. 2009a. Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head. *Mind* 118:963-93.
- Clark, A. 2009b. Letter to the editor. *London Review of Books* 31:6 (26 March 2009).
- Clark, A. 2010. Coupling, constitution, and the cognitive kind: A reply to Adams and Aizawa. In (R. Menary, ed.) *The Extended Mind*. Ashgate.
- Clark, A. & Chalmers, D.J. 1998. The Extended Mind. *Analysis* 58:7-19.
- Farkas, K. 2012. Two Versions of the Extended Mind Thesis. *Philosophia* 40:435-447.
- Gertler, B. 2007. Overextending the Mind? In (B. Gertler & L. Shapiro, eds.) *Arguing About the Mind*. Routledge.
- Rupert, R. 2009. *Cognitive Systems and the Extended Mind*. Oxford University Press.
- Tye, M. 1995. Ten problems of consciousness. MIT Press.
- Vold, K. 2015. The parity argument for extended consciousness. *Journal of Consciousness Studies* 22:16-33.