**Strange but True:**

**On the Counter-Intuitiveness of the Extended Mind Hypothesis**

Zed Adams

Chauncey Maher

[Penultimate Draft]

Forthcoming in *Journal of Consciousness Studies*

"[C]ommon sense leaves us completely in the dark as to the true intrinsic nature of physical objects, and if there were good reason to regard them as mental, we could not legitimately reject this opinion merely because it strikes us as strange."

(Russell, 1997/1912, p. 38)

1. **Introduction**

The *Extended Mind Hypothesis* (hereafter, EM) strikes many as counter-intuitive.[[1]](#footnote-1) It is the claim that there are parts of the world, outside of human bodies, that are literally parts of human minds. As counter-intuitive as it might sound, EM rests on a plausible idea: that the world itself is minded when parts of it are functionally equivalent to parts of human minds. But even that idea can seem counter-intuitive.

In this paper, we assess two intuitive criticisms of EM recently expressed by Sam Coleman (Coleman, 2011). The first is that the examples of extended mind offered by advocates of EM are not parts of minds, because subjects are not “conscious” or “immediately aware” of those parts of the external world. The second is that the principle at the heart of the argument for EM is biased in favor of EM. We argue that both of these intuitive criticisms of EM fail. Our ultimate aim is to suggest that the counter-intuitiveness of EM is not a barrier to its acceptance.

Our paper has three parts. In §2, we sketch the *Functional Equivalence Argument* for EM (hereafter, FE). In §§3-4, we explain and undermine these two intuitive criticisms of EM. We conclude by indicating what we think is really at stake in debates about EM.

1. **The Functional Equivalence Argument for the Extended Mind**

The FE is the master argument for EM.[[2]](#footnote-2) It presupposes *Vapid Functionalism*, the view that cognitive processes and their components are best understood functionally—in terms of what they do—rather than in terms of what they are made of.[[3]](#footnote-3) This claim is *vapid* since it is neutral about both the function of such processes and the substances that instantiate these functions. According to vapid functionalism, the identity conditions (and not merely the identifying conditions) of cognitive processes are specified in terms of what these processes do, not in terms of what they are instantiated in.

Vapid Functionalism supports the *Principle of Functional Equivalence* (PFE) at the core of the FE:

(PFE) Ifthere are items that are functionally equivalent to items that are *already* granted to be parts of cognitive processes, thenwe should grant that these items are also parts of cognitive processes.

The PFE implies that cognition extends beyond the body when there are external items that are functionally equivalent to internal items are already granted to be parts of cognitive processes.

The second step in the FE gives examples of such external items. For example, Clark and Chalmers (hereafter, C&C) offer the imaginary case of Otto, an Alzheimer’s patient who remembers where the Museum of Modern Art is by writing down its address in a notebook. C&C propose that if the following four criteria are met, then the contents of Otto’s notebook are functionally equivalent to stored internal representations (i.e., memories): (i) Otto always carries the notebook with him, (ii) the notebook is easy to access, (iii) Otto automatically endorses whatever he reads in it, (iv) Otto only writes things in it that he has explicitly endorsed in the past. C&C argue that given these four features, the contents of the notebook are functionally equivalent to inner mental items.

Here is the FE, in outline:

(P1) Principle of Functional Equivalence

(P2) There are cases of functional equivalence.

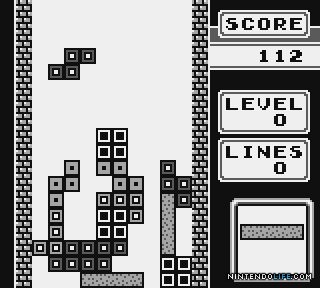
(C) The mind sometimes extends into external items.

If the FE is sound, then there are external items, outside of human bodies, that are literally parts of minds.

1. **Can we be “immediately aware” of worldly processes?**

In “There is No Argument that the Mind Extends,” Sam Coleman gives expression to two intuitive criticisms of EM. We discuss these criticisms in turn. The first alleges that that the examples offered by advocates of EM are not examples of parts of the world that are functionally equivalent to parts of minds.[[4]](#footnote-4) In other words, this first criticism rejects P2.

Coleman focuses on a second example offered by C&C involving the videogame Tetris. In Tetris, players arrange falling rectilinear blocks into contiguous horizontal layers. Since the blocks have different shapes and can fall from any part of the top of the screen, the task is to determine where they best fit in the existing layers at the bottom of the screen. This must be done before the block has landed on a layer, at which point it becomes immobile. Pieces can be moved left and right, and rotated in ninety degree increments. If a layer contains all and only contiguous pieces, it vanishes. If it does not, it remains, leaving less room to manipulate subsequent falling blocks. You lose when there is no room left. Here is a representative snapshot of a Tetris screen:



Many Tetris players play by rotating an inner mental representation of each falling block, as a way of determining where it best fits. Call this the *internalist* way of playing. C&C point out, however, that there is another way of playing, an *externalist* way. In the externalist way, one determines where a falling block best fits by rotating the block itself, on-screen. Although this externalist way might seem more time-consuming and therefore less effective in a time-sensitive game, it turns out that solving the problem of where a piece fits by rotating it on-screen can be more than three times faster than solving this same problem by rotating an inner representation of the piece in one’s head.[[5]](#footnote-5) Since the manipulation of this external, worldly item allows one to play Tetris just as well as, if not better than, the manipulation of an inner representation of this item, C&C contend that this external, worldly process is functionally equivalent to an internal mental process.

Coleman contends that the external process of rotating on-screen blocks is not functionally equivalent to the internal process of rotating mental representations. He thinks there is a functional difference between these two ways of playing Tetris: externalist players are *not* consciousof on-screen rotations in the same way that internalist players are conscious of their inner rotations (Coleman, 102). Thus, for Coleman, the crucial difference between internal and external rotations is that only internalist players are “immediately aware” or “directly conscious” of the rotations involved (102). [[6]](#footnote-6) Coleman thinks this shows that C&C’s Tetris example is not, in fact, an example of functional equivalence.

The problem with this criticism is that it does not, in fact, identify a relevant functional difference between externalist and internalist players. Advocates and critics of EM can agree that there are differences between internalist and externalist ways of playing Tetris. First, the externalist way is faster. Second, the steps that make up each process are different: each involves steps that are not in the other. On the one hand, in the internalist process, once a fit has been determined by rotating an inner representation, the player must still go on to rotate the on-screen block to the desired position. That final step is not part of the externalist process. On the other hand, in the externalist process, after each on-screen rotation, the player must look and see if the block is now in position to fit into the existing layers. That intermediate step is not part of the internalist process. These differences between the two processes are undeniable. What is deniable is that any of them constitutes a relevant functional difference.

Coleman thinks the relevant functional difference is that the internalist process is conscious in a way that the externalist process is not. Specifically, he claims that internalist players are *immediately aware* of their inner rotations, whereas externalist players are not immediately aware of their on-screen rotations.

But what is it for a player to be immediatelyaware of part of a cognitive process? If immediacy here is temporal, then the externalist process of rotating blocks on-screen is more immediate than the internalist process. After all, the externalist process is faster. If immediacy here is infallibility, then neither process is immediate, since both processes are liable to failure. If immediacy here is epistemic directness, then both processes are equally immediate, since one can directly know whether a piece fits by internal or external rotation.[[7]](#footnote-7)

So, on any of these plausible conceptions of immediate awareness, internalist players are no more immediately aware of their internal rotations than externalist players are of their on-screen rotations. Therefore, this first criticism does not identify a relevant functional difference between internalist players and externalist players.

One might still think that there must be some other conception of immediate awareness that identifies a difference between these cases. There is, but it will not help motivate the first criticism. An internalist player simply imagines whether a block fits, whereas an externalist player must look and seewhether a block fits. On this conception of immediacy, the internalist process is immediate in a way that the externalist process is not. But this sort of immediacy is not a relevant functional difference between the two ways of playing. Both processes *do* the same thing, after all. This difference simply re-states something that all parties to the debate already agree on: that the externalist way of playing Tetris has an external component that the internalist way lacks. (That is what makes it externalist!)

The phrase “immediate awareness” has tremendous intuitive appeal: it appears to mark a fundamental difference between internalist and externalist ways of playing Tetris. When one considers what this phrase might mean in the context of playing Tetris, however, it turns out that in all relevant functional respects, players are just as much immediately aware of external rotations as they are of internal rotations.[[8]](#footnote-8) As such, the intuitive appeal of this phrase is simply an illusion.[[9]](#footnote-9) Thus, the first intuitive criticism fails to undermine the third premise in the FE. It does not given us any reason to doubt that the examples offered by advocates of EM are examples of functional equivalence.

1. **Is the Argument for EM Biased?**

The second intuitive criticism of EM expressed by Coleman is that the second premise in the FE is *biased* in favor of EM. That is, this criticism rejects P1. Following C&C, Coleman calls this premise the “Parity Principle” (PP).[[10]](#footnote-10) Here is how Coleman puts this principle:

(PP) If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive/mental process, then that part of the world is (for that time) part of the cognitive/mental process. (p. 102)[[11]](#footnote-11)

Coleman argues that PP is biased because it is question-begging. He claims that no one would accept PP without having already accepted EM.

Coleman admits that PP does not seem biased (p. 106). It is merely a claim of equivalence, and it might seem that claims of equivalence are not biased in favor of one or the other side of that equivalence. For instance, the claim that rubber doorstops are functionally equivalent to wooden doorstops does not appear to be biased in favor in any particular view about what door stops can be made of—wood, rubber or anything else.

In order to show that PP is biased, Coleman introduces another equivalence principle, one that is analogous to PP but which Coleman claims is obviously biased *against* EM.

(PP\*) If, as we confront some task, a part of the head functions as a process which, were it to go on in the world, we would have no hesitation in rejecting as part of the cognitive/mental process, then that part of the head is (for that time) not part of the cognitive/mental process. (p. 105)

Coleman argues that PP\* is biased by applying it to the case of standing beliefs. Such dispositional or non-occurrent beliefs store content, but, as long as they remain non-occurrent, their content is not consciously entertained. In this regard, standing beliefs are functionally equivalent to other content-storing items in the outside world, such as notebooks. But, Coleman asserts, “we do not *prima facie* regard content-storing states of the environment [such as notebooks] as mental” (p. 105). Given this, if one accepts PP\*, one is committed to thinking that standing beliefs are not mental. Coleman takes this to imply a radically *narrow* view of the mind, according to which only a subset of internal content-storing states are genuinely mental: namely, occurrent ones. Taking himself to have shown that PP\* implies a narrow view of what counts as mental, Coleman asserts that “PP\* wears its internalism on its sleeve” (p. 107) and is therefore clearly biased againstEM.

Coleman concludes that since PP\* is analogous to PP, PP is equally well biased in favor of EM, against internalism.

This second criticism goes wrong in two places. First, neither PP\* nor PP is biased. Second, even if one insists that PP *appears* biased, it can be replaced by a different formulation that is manifestly not biased.

Coleman’s first mistake is to claim that PP\* is biased in favor of internalism. As we will show, this is simply false, for PP\* is perfectly compatible with EM. Clearly, it matters here what it is for an equivalence claim to be biased. In several places, Coleman states that bias involves begging the question, that a claim is biased in favor of a view if it presupposes the truth of that view (and thereby implies the falsity of an opposing view).

However, PP\* does not presuppose the truth of internalism. PP\*, remember, is the claim that if something inside the head is functionally equivalent to something outside the head that is not mental, then that inner item is equally well not mental. Coleman says that this implies that standing beliefs are not mental. For the purposes of argument, we will grant this implication.[[12]](#footnote-12) Even if it turns out that standing beliefs are not mental, that does not presuppose internalism. Quite the contrary: that some things in the head turn out to be non-mental does not at all presuppose that there is nothing outside the body that is mental. Simply put, even if PP\* shows that some things insidethe head are not mental, it tells us nothing whatsoever about the mentality (or lack thereof) of things outside the body. It is entirely neutral on this question.

To see this, consider the following principle:

(D) If something outside of one’s house walks and talks like something that, were it to be inside of one’s house, we would have no hesitation in accepting as a duck, then it is a duck.

To all appearances, D does not imply or presuppose that there are any ducks outside of one’s house. Yet Coleman would have it that D is biased against those who hold that there are no such ducks. And he would have it that this bias is revealed by the following analogous principle:

(D\*) If something inside of one’s house walks and talks like something that, were it outside of one’s house, we would have no hesitation in rejecting as a duck, then it is not a duck.

But D\* does not, in any way, imply that there are no ducks outside of one’s house! It simply holds that *some* of the things inside of one’s house *might* not be ducks. Similarly, PP\* simply holds that some of the things inside of one’s head might not be mental. It does not, in any way, imply that nothing outside of the body is mental.[[13]](#footnote-13) It is entirely neutral on that question. So, if bias is simply a matter of begging the question, PP\* is not biased in favor of internalism, and PP is not biased in favor of EM.

What has led Coleman astray? He seems to have mistakenly equated “internal” with “narrow”. One might think that PP\* encourages a narrowview of the mind, according to which only occurrent thoughts are mental.[[14]](#footnote-14) But such a narrow view of the mind is not equivalent to an internalist view of the mind, according to which mental activity only goes on inside the human body. PP\* is entirely neutral on wheremental activity goes on; it only holds that wherever it goes on, it functions the same way. That is all that PP\* or PP holds.

However, in a couple places, Coleman seems to think that bias is something other than begging the question. He writes:

depending upon the interests of the person drawing a particular equivalence, it is possible to build priority into that equation; this hangs on which side of the equation the person takes as basic. The equation thereby can be imbued with directionality—bias. (p. 106).

The idea here is that a claim is biased when priority is built into it, when the person making the claim takes one side of the equivalence to be more basic than the other side.[[15]](#footnote-15) For instance, the claim that natural and artificial hearts are functionally equivalent would be biased in favor of natural hearts if someone making that claim took natural hearts to be more basic for understanding what hearts are. Likewise, PP\* would be biased in favor of internalism if the person offering it took internal items to be more basic for understanding what minds are.

Coleman might be right that this kind of building-in of priority makes a person biased, but he is certainly wrong that it makes a claim biased. After all, one and the same claim can be used by different people who treat opposite sides of the same equation as basic. One could take eithernatural or artificial hearts to be more basic.[[16]](#footnote-16) Given this, it is absurd to think that PP\* or PP in itself is biased. Furthermore, this sort of bias that persons can exhibit does not support Coleman’s ultimate conclusion that the argument for EM is biased. The fact that a person arguing for EM can be biased in favor of EM—in the sense just described—does not imply that her argument for EM is biased. So, neither PP\* nor PP nor the FE as a whole is biased.

Even if one continues to insist that PP appears biased, one could replace it with a manifestly unbiased formulation of the equivalence claim at the heart of the FE. In §2, we already provided one such formulation, the PFE. Here it is again:

(PFE) Ifthere are items that are functionally equivalent to items that are *already* granted to be parts of cognitive processes, thenwe should grant that these items are also parts of cognitive processes.

Granted, PP and PFE are similar. But PP mentions locations (“the world,” “the head”) and PFE does not. In their original essay, C&C had good reason to mention location.[[17]](#footnote-17) They mentioned it as a way of emphasizing that being “in the head” was not what mattered for something to count as part of a mind. However, the very mentioning of location potentially invites worries about bias. Moreover, mention of location is utterly incidental to an adequate formulation of the claim needed for the argument for EM. PFE, for instance, does not mention location.[[18]](#footnote-18)

Most importantly, PFE does not presuppose that EM is true; it does not presuppose that anything outside of the head is part of a cognitive process. It is neutral with respect to the location of parts of cognitive processes; PFE could be true and it could turn out that there simply are not any items outside of the head that are functionally equivalent to parts of admitted cognitive processes. The second intuitive criticism fails to identify any bias in the argument for EM.

1. **Conclusion**

We began this paper by noting that the Extended Mind Hypothesis (EM) conflicts with many intuitions about minds, especially intuitions about the spatial locations of minds.[[19]](#footnote-19) For example, it is a widely-shared intuition to think that minds do not – and perhaps cannot – extend beyond human bodies. As such, however, such intuitions simply beg the question against EM. Sam Coleman’s paper (Coleman 2011) is useful because it attempts to offer non-question-begging support for these intuitions. As we have seen, however, Coleman putative support for these intuitions fails to convince. Although EM may seem counter-intuitive to many, it remains a viable hypothesis about the mind.

That said, EM is not neutral with respect to every thesis about the mind. As noted in §2, the master argument for EM presupposes Vapid Functionalism. EM assumes that cognitive processes are functionally individuated; it says that doing the same thing as part of a cognitive process is sufficient for being part of one. This indicates one way to resist EM: deny Vapid Functionalism![[20]](#footnote-20)

Of course, most contemporary philosophers of mind are not prepared to do that. So, a better way to challenge EM would be to argue that given the function of cognitive systems, they cannot extend beyond the body. That would require abandoning the vapidity of Vapid Functionalism and taking a stand on a tough question: what, exactly, is the function of cognitive systems? Thus, if nothing else, the EM is valuable because it raises that question in the starkest terms. What is it for something to function as part of a cognitive system? What is the function of such systems as wholes? We think that is really what is at state in debates about EM.

# Works Cited

Adams, F., & Aizawa, K. (2001). The Bounds of Cognition. *Philosophical Psychology, 14*(1), 43-64.

Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition.* New York: Blackwell.

Adams, F., & Aizawa, K. (2010). The Coupling-Constitution Fallacy Revisited. *Cognitive Systems Research, 11*(4), 332-342.

Adams, Z., & Maher, C. (2012). Cognitive Spread. *European Journal of Philosophy*.

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

Clark, A. (2009). Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head. *Mind, 118*(472), 963-993.

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analsysis, 58*, 10-23.

Coleman, S. (2011). There is No Argument that the Mind Extends. *The Journal of Philosophy*, 100-108.

Fryer, D., & Marshall, J. (1979). The Motives of Jacques de Vaucanson. *Technology and Culture, 20*(2), 257-269.

Grush, R. (2003). In Defense of Some ‘Cartesian’ Assumptions Concerning the Brain and Its Operation. *Biology and Philosophy, 18*(1), 53-93.

Haugeland, J. (1990). The Intentionality All-Stars. In J. Tomberlin (Ed.), *Philosophical Perspectives, IV: Philosophy of Mind and Action Theory* (pp. 383-427). Atascadero, CA: Ridgeview Publishing Company.

Haugeland, J. (1995). Mind Embodied and Embedded. (L. Haaparanta, & S. Heinämaa, Eds.) *Acta Philosophica Fennica, 58*, 233–267.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*, 513-549.

Menary, R. (2010). Introduction. In R. Menary (Ed.), *The Extended Mind.* Cambridge, MA: MIT Press.

Menary, R. (Ed.). (2010). *The Extended Mind.* Cambridge, MA: MIT Press.

Russell, B. (1912/1997). *The Problems of Philosophy*. New York: Oxford University Press.

Sellars, W. (1956). Empiricism and the Philosophy of Mind. In *Minnesota Studies in the Philosophy of Science* (Vol. I, pp. 253-329). Minneapolis: University of Minnesota Press.

Sprevak, M. (2009). Extended Cognition and functionalism. *The Journal of Philosophy, 106*, 503–527.

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind and the Civilizing Process. In R. Menary (Ed.), *The Extended Mind* (pp. 189-225). Cambridge, MA: MIT Press.

Wittgenstein, L. (1953/1958). *Philosophical Investigations* (3rd ed.). (G. Anscombe, Trans.) New York: Blackwell.

1. Of course, not everyone finds EM counter-intuitive. Furthermore, many think intuitions are an unreliable guide to the truth. We are inclined to agree. However, since some people think EM should be rejected because it appears strange, we confront this appearance head-on in this paper. [↑](#footnote-ref-1)
2. (Haugeland, 1995) and (Clark & Chalmers, 1998). [↑](#footnote-ref-2)
3. In calling this formulation of functionalism “vapid” we do not mean to denigrate it. Rather, we mean to echo Haugeland’s discussion of “vapid materialism” in (Haugeland, 1990), which was intended to capture a minimal form of materialism that finds widespread acceptance in philosophical discussions today. [↑](#footnote-ref-3)
4. Whether some part of or process in the world is functionally equivalent to some part of or process in a human body is a primary topic of dispute. See, for instance, (Grush, 2003) and (Sutton, 2010). [↑](#footnote-ref-4)
5. C&C cite (Kirsh & Maglio, 1994) at (1998, p. 8). [↑](#footnote-ref-5)
6. Coleman assumes that consciousness is *the* mark of the mental. Since he does not provide any support for this contentious claim (e.g., it excludes subconscious, unconscious, and non-occurrent states and processes from being mental), we will not discuss this assumption further. The putative functional difference he identifies between internalist and externalist ways of playing Tetris is independent of this assumption. [↑](#footnote-ref-6)
7. That is, unless one presupposes a “veil-of-ideas” conception of perception according to which people are never *directly* aware of states of affairs in the external world. That conception of perception is contentious enough that it cannot be what Coleman means by ‘immediate awareness’. If it is what he means, his criticism loses much, if not all, of its intuitive appeal. [↑](#footnote-ref-7)
8. (Clark, 2009) questions whether the *conscious* mind extends. [↑](#footnote-ref-8)
9. Indeed, (Haugeland, 1995) and (Clark & Chalmers, 1998) highlight this idea: in all relevant senses of the phrase, a person can be “immediately aware” of the external, worldly processes described in their examples. In this respect, Haugeland and C&C are part of a line of philosophers that includes Wittgenstein (1953/1958) and Sellars (1956) who are dissatisfied with traditional appeals to immediate awareness in discussions of the mind and want greater clarity about what such awareness might really involve. [↑](#footnote-ref-9)
10. The so-called parity principle is at the center of many debates concerning EM. See, for instance, (Menary, The Extended Mind, 2010). In several publications, Fred Adams and Ken Aizawa contend that this principle leads to fallacious reasoning, which they have dubbed “the coupling constitution fallacy” (2001, 2008, 2010). Questioning the requisite degree and character of the parity involved is arguably characteristic of so-called second-wave developments of EM, see, e.g., (Menary, Introduction, 2010) . [↑](#footnote-ref-10)
11. See (Clark & Chalmers, 1998, p. 8). [↑](#footnote-ref-11)
12. We doubt this follows. We think that whether “content-storing states of the environment” are mental depends on whether they exhibit a certain sort of semantic activity. See (Adams & Maher, 2012). [↑](#footnote-ref-12)
13. In essence, Coleman assumes that if some Is (internal items) are not Ms (mental items), then that implies that all Ms are Is. That is clearly wrong. The fact that some Is are not Ms does not even imply that some Ms are Is, let alone that all Ms are Is. [↑](#footnote-ref-13)
14. We ourselves do not think PP\* encourages such a view. But, for the purposes of argument, we are willing to grant that it might, if only to show that granting that it might is entirely independent of the question of whether it is biased in favor of internalism. [↑](#footnote-ref-14)
15. We wish Coleman said more about what “more basic” means. We offer a reading of what he might mean, but Coleman’s casual use of the phrase belies the number of significant philosophical issues it brings up. [↑](#footnote-ref-15)
16. In fact, it is arguable that certain automata designers took their artificial creations to be more basic in this sense. See (Fryer & Marshall, 1979). [↑](#footnote-ref-16)
17. As did Haugeland, whose own version of the PFE also mentions location. [↑](#footnote-ref-17)
18. As Menary says, “it is functionality and not location that matters when determining whether or not a process is cognitive” (Introduction, 2010, p. 6). [↑](#footnote-ref-18)
19. Some philosophers deny that EM is counter-intuitive. Clark writes, “I am also inclined… to dispute the claim that the Extended Mind Model runs so wildly contrary to common sense” (2008, pp. 105-6). [↑](#footnote-ref-19)
20. See, for instance, (Sprevak, 2009). [↑](#footnote-ref-20)