

THE *EXTENDED MIND*: A CASE OF MENTAL VEHICULAR NEGLIGENCE

by

Owen Alexander

Thesis submitted in partial fulfillment of the
requirements for the Degree of
Bachelor of Arts with
Honours in Philosophy

Acadia University

March, 2018

© Copyright by Owen Alexander, 2018

This thesis by Owen Alexander
is accepted in its present form by the
Department of Philosophy
as satisfying the thesis requirements for the degree of
Bachelor of Arts with Honours

Approved by the Thesis Supervisor

Dr. Anna Frammartino Wilks

Date

Approved by the Head of the Department

Dr. Stephen Maitzen

Date

Approved by the Honours Committee

Dr. Matthew Lukeman

Date

I, OWEN ALEXANDER, grant permission to the University Librarian at Acadia University to reproduce, loan or distribute copies of my thesis in microform, paper or electronic formats on a non-profit basis. I, however, retain the copyright in my thesis.

Signature of Author

Date

Acknowledgements

A single page is no adequate measure of the gratitude I feel towards all the people who made this paper possible. Even so, I would like to express my deepest thanks, first and foremost, to Dr. Anna Frammartino Wilks, whose instruction furnished me with the knowledge and know-how required to do philosophy and who was absolutely indispensable in encouraging and guiding me throughout the writing of this thesis. I would like to thank Dr. Ian Wilks for his meticulous second reading of my thesis, and for his guidance in things both philosophical and personal. I would like to express my gratitude to Dr. Stephen Maitzen for convincing me that I had it in me to undertake and successfully complete an Honours Degree in Philosophy. I would also like to thank Dr. Paul Abela and Dr. Marc Ramsay for the excellent instruction they provided me during my time at Acadia. Lastly, I would like to thank Husna Farooqui for enriching my life with her encouragement, friendship, and understanding.

Table of Contents

Acknowledgements	vii
Table of Contents	ix
Abstract	xi
Introduction: <i>The Extended Mind</i> and <i>Vehicle Externalism</i>	1
Chapter 1: Functionalism and the Extended Mind.....	5
1.1: Two Dogmas of Internalism.....	5
1.2: Why Minds are Like Mousetraps: Functionalism as a Theory of Mind	9
1.3: The Path is Wide, but the Gate is Narrow: Limitations on Multiple Realizability .	12
1.4: Functionalism and Mental State Inclusivity	15
1.5: Brains, Notebooks, Ectoplasmic Slime – What’s the Difference?.....	19
Chapter 2: The Problem of Cognitive Bloat	23
2.1: Reliable, Accessible, and Trusted: Nice but not Necessary Conditions for Mental Constitution	23
2.2: It’s not What’s on the Inside that Counts: Mental State Constitution is a Fair Game	27
2.3: Raindrops are Falling on my Mind: The Proper Parts of Mental States	29
2.4: The Extended Mind as a Coupled System.....	35
2.5: One of These Things is not Like the Other: Where is the Mind in a Cognitive System?.....	40
Chapter 3: The <i>Extended Self</i>	47
3.1: Where am I?	47
3.2: Locke, Stock, and Personal Identity as Psychological Continuity	49
3.3: Mental State Internalism	52
3.4: The Scattered Self	57
Concluding Thoughts	60
Bibliography.....	70

Abstract

In 1998 a thought experiment made a splash in the *philosophy of mind*. Its objective was to demonstrate that the mind is not all in the head. The mind can be *extended* when certain conditions obtain between thinker and environment. Typically, the sufficient condition is functional similarity between what goes on in the head, and what goes on beyond the head. Take notebooks, for example. They can be used to record information that represents beliefs, thoughts, or desires that are normally contained by the brain. The intuition is that if some external object plays a role in some process that, if done in the head, we would consider a *mental* process, then that object is a part of the mind, location notwithstanding.

But without the *de facto* boundaries of the skin and skull, what is the new limit to the extension of the mind? Functional similarity is sufficient to get the mind out of the head, but it is silent as to where exactly to redraw the line of the mind. Limiting conditions are needed because if, for example, a notebook is sufficiently similar so as to count as a part of your mind, then would the *Yellowpages* count too, just because you used them to order pizza? In the quest for new limits on the mind I look to such places as *proper parts theory* and *dynamical systems theory*, but ultimately find no new mental boundaries.

I conclude by considering a few of the many absurd consequences admitting the possibility of extended minds has for the notions of *selfhood* and *personal identity*.

Introduction: *The Extended Mind and Vehicle Externalism*

What would life be like if our minds were not in our heads? The view that this is, in fact, the case is a position in the philosophy of mind referred to as the *extended mind thesis*¹. It is also known as *vehicle externalism*. A mental vehicle is the material stuff which gives rise, in some way, to a mental state. In constructing arguments for the *extended mind thesis*, *functionalism* is the theory of mind one is expected to bring to the table. According to functionalism, mental states are essentially functional states. Functional states are identified by the causal roles that are performed by stimuli, behaviour, and other mental states. For example, to possess the mental state *pain*, one must have experienced an appropriate stimulus (stubbing one's toe), produced the appropriate behaviour (copious cursing), and possess appropriate relations between other mental states (*desiring* that you never bought that gosh darn dresser). By "appropriate" I mean performing the causal role necessary to identify a given functional state as such. The relevance and well-suited-ness of functionalism in explaining and supporting external mental vehicles is readily discernable: a mental state can be made of any conceivable material so long as the functional state with which it is identical has parts that play the appropriate causal roles. A movie reference might serve well here. In the movie *Memento*, Guy Pearce plays a character with retrograde amnesia named Leonard. In the movie, Leonard's memory resets at regular intervals, often at very dramatic and suspenseful points in the movie. To combat his amnesia he takes Polaroid photographs, inscribed with vague hints, to stand in for memories. While, under a more conventional interpretation, the photographs would be

¹ The phrase "extended mind" is italicized only when referring to the theory itself and not when referring to an actual or possible instance of an extended mind.

considered “stand-ins” for his memories, on an extended mind interpretation they literally *are* his memories; Leonard has an extended mind. Some of his mental vehicles lie outside of his head, and the reason why is that they play the appropriate causal roles that hold between stimuli, behaviour, and other mental states. Location of mental vehicle is a difference, but is it a significant one? Well, some philosophers, if they were to make some popcorn and sit down to watch *Memento*, would see Leonard’s usage of the Polaroids as a functional equivalent of their own neurological memory. These philosophers would argue that there is no reason why the vehicles that perform appropriate causal roles in mental state identification *must* be in the head or *must* be made out of brainstuff – Polaroids could serve just as well. Having a belief or remembering something is a functional state, not a material state; it shouldn’t matter whether the vehicle of the content is a Polaroid or a brain; as long as they perform similar causal roles in the functional state, it *is* a part of the mind.

But if Polaroids could be part of a mind, what else could be? Conceivably, if the mind is out of the head, then the sky is the limit. Our intuitive belief that the mind is skull-bound is chauvinistic; there is *stuff* out there just as capable of mental state realization as the brain. The organism/environment boundary picks out a real physical boundary in nature, but the metaphysical screen between the mind and the not-mind is *not* laid on top of it, if it is indeed laid anywhere. Although transhumanists and science fiction writers might rejoice at some philosophical underpinnings for their pet beliefs that perhaps one day they will be able to upload their consciousness to the internet or become the biomechanical cyborg they always wanted to be, a mind that includes external mental vehicles is about as intelligible as an air-filled balloon without the balloon. If the mind is like the volume of air in the balloon, and you pop the balloon, where *exactly* are the

boundaries of the volume of air that was formerly encapsulated in a rubber membrane? You really cannot say – either because there *are* no boundaries to the volume of air anymore, or because there are boundaries, but they are just too *vague* to be identified. I argue that is what the *extended mind thesis* has done to the mind. It has popped the balloon of cranial hegemony. The popping has been lauded as a good thing – I argue however, it is not. On the *extended mind* view, the mind is a scattered object comprised of neural and environmental mental vehicles. But which ones are actual *constituents* of a mind? I mark this as *the* central question within the debate surrounding the *extended mind* for two reasons. One, it is a difficult question to answer, meaning that it is likely a good philosophical question; two, any answer given to the question will have many interesting and somewhat troubling implications for our notions of personal identity and self-hood.

If I can successfully show that an extended mind is as scattered an object as dust in the wind or a particular cup of water poured in the Pacific Ocean, then I am entitled to the claim that *I* am more like a dispersed quantity of water than a spatially discrete organism. The reason is this: My mind is a part of *me*, and I am located partially wherever a part of me is located. So if my mind contains the Polaroids I took of the Grand Canyon that I lost on the side of Highway 66, then *there* I am (partially) while I sit here in Eastern Canada typing this sentence. In short, the *extended mind* implies the *extended self*, and the *extended self* means I am not *only* where my head or my feet are. In fact, if you argue that the “self” is identical with a series of mental states, which is a not uncommon position in the philosophy of personal identity, then since many parts of my body are not vehicles of any of my mental states, such as “my” ears, toes, gall-bladder, etc., and so are not parts of me. The Polaroid on Highway 66 thousands of kilometers away is a part of me, but the toes in my shoes are not. As intuitive as the arguments for extraneural mental vehicles

might be, which are, additionally, especially attractive to anyone with functionalist sympathies, *extended mind theory's* tension with our common-sense notions of ourselves as spatially determinate individuals renders it a difficult position to adopt. Nonetheless, I think some of the arguments offered in its defense merit serious consideration. In my examination of these arguments I attempt to show how the conceptual tension between extended minds and individual selves can be relieved.

Chapter 1: Functionalism and the Extended Mind

1.1: Two Dogmas of Internalism

The *extended mind thesis* is the view that not all the parts of the mind are located in the head. To clarify this position, it is necessary to explicate several key concepts in philosophy of mind. The *extended mind thesis* can be located in a specific quadrant within a matrix consisting of two predicates and two subjects, respectively *internal/external* and *content/vehicle*. By “internal” I simply mean depending *only* on “in the brain” factors and by “external” I mean depending on *some* “outside the brain” factors.² As to the objects to which these predicates apply, I understand “content” as that which a mental state is about, i.e. what it represents concerning the way the world is. A “vehicle” is that material object or process that gives rise to the content of a mental state. To see this classifying scheme in action, take the thought that the earth is round. The thought has content, it is about the earth and its roundness, and it is instantiated by a vehicle, a particular activation of neurons.

Traditionally, this matrix has been used exclusively in the domain of *content*, with philosophers of language and philosophers of mind trying to figure out whether or not what a given mental state is about is in some way determined by the way the world happens to be. Usually, thought experiments involving physically duplicate persons embedded in different possible worlds are brandished to argue that the persons can have different content pertaining to their mental states despite physical sameness. They point to

² The sense of “depending” depends on your theory of mind. For example, it could refer to a relation of supervenience, identity, or emergence.

the only variable in the thought experiment, the difference in worlds in which the original and the twin respectively reside, and thereby conclude that these external environmental conditions are determining what is capable of being thought by each. That is to say, mental state content is determined in part by how the world is. This view is known as *content externalism*, which I will be using as an explication device for the topic of my thesis, i.e., the *extended mind thesis*. Hilary Putnam provided the most influential argument for content externalism: the famous *Twin Earth Thought Experiment*.³ There is a man named Oscar who lives on Earth and speaks English. He possesses the commonplace belief that *water is wet*. However, there is also a man named Twoscar, who lives on Twin Earth, but also speaks English and is a perfect physical duplicate of Oscar. Twoscar, too, believes that *water is wet*. But here's where things get strange: on Twin Earth, what is called "water" is actually the chemical substance XYZ. So when Twoscar says that "water is wet" he is actually referring to XYZ and not water as Oscar would on Earth. So Oscar and Twoscar actually *mean* something different when they utter "water is wet" because Oscar can only refer to water and Twoscar to XYZ. In other words, the content of their respective beliefs is different owing to the difference in how the world is. Oscar's belief that *water is wet* is about water, yet Twoscar's identically phrased belief is about XYZ. Since Oscar and Twoscar are physical duplicates, the difference in belief cannot be accounted for by a difference in brain-states. The only variable in the thought experiment is the world in which Oscar and Twoscar, respectively, find themselves; Earth has water, Twin Earth has XYZ, and this fact makes all the difference to the content of

³ Hilary Putnam, "The Meaning of 'Meaning'," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers (Oxford: Oxford University Press, 2002), 584–585.

each thinker. As Putnam colloquially put it “Cut the pie anyway you like, ‘meanings’ just ain’t in the *head*!”⁴

Putnam effectively called into question the assumption that all *mental content* is safely insulated in the head from all outside influence. This is the first dogma of internalism. The *extended mind thesis* is not about meaning and content, however. I understand the *extended mind thesis* to be about *vehicle externalism*, thereby serving as an alternative to the second dogma of externalism: *vehicle internalism* – the view that all *mental vehicles* are in the head. I draw this conclusion from how it was originally introduced by Andy Clark and David Chalmers: “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.”⁵ Further corroboration for this interpretation can be taken from Baker, Hurley, and Wheeler.⁶ This is the claim that it is possible that the material vehicles of a mental state can be outside the head. Therefore, the sense in which “externalism” is being used is spatial *simpliciter*. When can vehicles of mental states be outside the head? Clark and Chalmers say when there is sufficient *functional parity* between outside-the-head vehicles and inside-the-head vehicles. Luckily, just as content

⁴ Putnam, “Meaning,” in Chalmers, *Philosophy of Mind*, 587.

⁵ Though Clark and Chalmers speak about a “cognitive process,” I take them to be speaking about the mind as philosophers would, i.e., a collection of folk psychological states such as beliefs, desires, and fears, and not as some mechanistic system geared for completing cognitive tasks as cognitive scientists would. The reason why I hold them to be talking about the mind and not cognitive systems in general is that they make extensive reference later in their paper to dispositional belief as a paradigm case of a mental state with possible external vehicles. See: Clark and Chalmers, “The Extended Mind,” in Chalmers, *Philosophy of Mind*, 644.

⁶ See: Lynne Rudder Baker, “Persons and the Extended-Mind Thesis,” *Zygon* 44 (3) (2009): 645; Susan L. Hurley, “Varieties of Externalism,” in *The Extended Mind*, ed. Richard Menary, (Cambridge: MIT Press, 2010), 102; Michael Wheeler, “In Defence of Extended Functionalism,” in Menary, *The Extended Mind*, 245.

externalists had their pet thought experiment, vehicle externalists have theirs too. What the thought experiment is designed to show is this essential notion of *functional parity*, or the intuition that neither location nor material constitution are essential properties of parts of the mind. Rather, it is what the putative parts *do*, or the causal role that they play that is essential to mental state parthood. The thought experiment introduces the characters Otto and Inga, who happen to possess an identical belief: that *the Museum is on 53rd street*. They both want to go see the Museum of Modern Art Exhibit on 53rd street and consequently must form the belief that *the Museum is on 53rd street* in order to realize their desire. This is easy for Inga, who has a working biological memory. She simply stores the information using her marvellous internal vehicle, her brain, or more specifically, some coalition of neurons and larger brain areas such as the hippocampus, for future recall as a dispositional belief to guide her future behaviour.⁷ Otto is not so fortunate: he suffers from memory loss and is unable to reliably use his brain as a vehicle for the belief that *the Museum is on 53rd Street*. He must resort to an alternative vehicle, one that lies outside his head. It just so happens there is an object perfectly suited to being the repository of dispositional beliefs: a notebook. Here he writes the location of the Museum and uses the information to later guide his behaviour in finding the location of the art exhibit and fulfilling his desire. The question Clark and Chalmers ask is: are we justified in our conviction that Inga's brain is part of her mind, but that Otto's notebook is not? After all, the notebook seems to be doing the exact same thing for Otto as Inga's brain does for her. Their *functions* are identical. The only difference seems to be *material* and *spatial*: Inga's brain is made of brainstuff and is located inside her cranium; Otto's

⁷ *Dispositional beliefs* stand opposed to *occurrent beliefs*. The difference between them is that the latter are actively held in consciousness, while the former are not.

notebook is made of paper and ink and is located outside of his cranium. It turns out that there is an assumed theory of mind grounding these intuitions about the primacy of function over material constitution. It should come as no surprise, then, that it is called *functionalism*. I explicate this theory in the following section – now that Oscar and Otto have awoken us from our dogmatic internalist slumber – with an externalist morning-call: meaning is “out there,” and so is the mind.

1.2: Why Minds are Like Mousetraps: Functionalism as a Theory of Mind

In order to consider even briefly whether a notebook could literally be a part of someone’s mind, it is assumed that *functionalism* is the default theory of mind with which one is working. Functionalism is the theory that mental states are identified by the causal relations they hold to stimuli, behaviour, and other mental states.⁸ It was developed in large part by Hilary Putnam, who exploited the conceptual possibilities of analogizing a Turing Machine with the human mind. It was later developed, defended, and qualified by philosophers of mind such as David Armstrong, Sydney Shoemaker, and David Chalmers.⁹ The stock example of a mental state analysis based on a functional understanding thereof is *pain*. To have the mental state *pain* is to receive a stimulus that causes anxiety or discomfort; to produce avoidant behaviour; and to cause the *desire*, which is itself a mental state, to be out of that state. As demonstrated, it is this causal web that determines the essence of the mental state *pain*, not its identity with some particular

⁸ Ned Block, “Troubles with Functionalism,” in Chalmers, *Philosophy of Mind*, 94.

⁹ See: Hilary Putnam, “The Nature of Mental States,” in *Art, Mind, and Religion*, eds. W.H. Capitan & D.D. Merrill, (Pittsburgh: Pittsburgh University Press, 1967), 37–48; David Armstrong, “The Causal Theory of Mind,” in Chalmers, *Philosophy of Mind*, 80–87; Sydney Shoemaker, *Identity, Cause and Mind* (Cambridge: Cambridge University Press, 1984); David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996).

activation of neurons, or brain state. What is called *multiple realizability* is perhaps the central functionalist insight, for it allows inclusivity of the kinds of material vehicles that can support mental states and the kinds of beings that can have mental states. *Multiple realizability* affirms that internal composition is irrelevant, since there is no privileged material, for example, brainstuff, that is the essential material substrate for the realization of a mental state. Put another way:

If psychological phenomena are constituted by their causal-functional roles, then our terms for mental states, mental processes, and so on, pick out equivalence classes of different material substrates, any one of which might in principle realize the type-identified state or process in question. But of course that means that robots, Martians, and the Ood may all join us in having mental states, just so long as the physical stuff out of which they are made is capable of being organized so as to implement the right functional profiles.¹⁰

The central tenet of functionalism can be explicated with a simple comparison, that of mousetraps and diamonds.¹¹ A mousetrap is entirely defined by what it does, i.e. its function, and its constitutive parts belong to the whole in virtue of how they causally contribute to the function of the whole. Diamonds are another matter, however. Diamonds are difficult objects for functional analysis because it is not what the constituent parts of a diamond *do* that makes a diamond; it is what they are made of and how the parts are structured. The difference between coal, graphite, and diamonds is not one of differing functional attributes, but of differing composition and arrangement. On the other hand, there is more than one way to catch a mouse, and there is more than one way to constitute a mind. So long as the function remains the same, what instantiates the causal roles of a functional state is materially neutral. This, of course, opens up a can of worms straight out

¹⁰ Wheeler, "In Defence," in Menary, *The Extended Mind*, 250.

¹¹ Derek Pereboom, "Why a Scientific Realist Cannot be a Functionalist," *Synthese* 88 (1991): 343.

of science fiction, inviting philosophical thought experiments about Martians, sentient computers, and a man trapped in a room whose only diversion is playing with squiggles and squaggles.¹² The relevance of this talk of mousetraps and diamonds is that minds are assumed to be more like mousetraps and less like diamonds by functionalist thinkers. Why? Two reasons: (a) *multiple realizability*, which is, to use the type/token distinction, the possibility that the same mental state *type* can be *tokened* by different material things;¹³ (b) *mental state inclusivity*, which is an attitude of liberality in attributing inner mental lives to creatures other than fellow human beings despite potential differences in gross physiology. Both of these insights render plausible the possibility of minds made of cheese and Martians with PhDs. In the face of these possibilities, the *extended mind* should be beginning to sound not so strange. It should be noted, however, that multiple

¹² Intuitively, it might seem like some substances are materially privileged or well-suited to being vehicles for a mind. Philosopher John Searle has gone so far as to say that brainstuff is the only material nomologically capable of producing mental activity by virtue of its unique causal-material structure. This assertion is the consequence of a famous thought experiment he put forward as a critique of functionalism, particularly, how functionalism seems to impute thoughts and beliefs to things like thermostats and computers, purely on the basis of functional definitions of mental states. Imagine a man inside a room completely quarantined from the outside world except for two slots in one of the walls labelled “input” and “output” respectively. Through the input slot come Chinese ideographic characters, which are nothing but squiggles to the man since he cannot understand a lick of Mandarin or Cantonese. Inside the room there is a box of said ideographs and a large book filled with instructions of the type “if this squiggle comes in, put this squaggle out.” The conclusion of the thought experiment is that being able to deal appropriately with representations according to a syntax is not sufficient for concluding that there is any understanding of the meaning of those representations. This a general critique of functionalism in that the room is functioning as a native Chinese speaker, yet doesn’t understand Chinese. What is lacking in the room lacking that a human understander of Chinese possesses? Searle’s answer is a brain, which, in virtue of its fine-grained causal profile, is contingently sufficient for human cognition. The room is just too far afield in its functional profile to ever be comparable to a human being with a brain. “Mental states are biological phenomena. Consciousness, intentionality, subjectivity, and mental causation are all a part of our biological life history, along with growth, reproduction, the secretion of bile, and digestion.” See: John Searle, “Can Computers Think?” in Chalmers, *Philosophy of Mind*, 675.

¹³ The *type/token distinction* is a relation between an abstract general kind of thing (*type*) and particular concrete instances of that kind of thing (*token*). For example, my copy of Plato’s *Republic* is a token of the type “Plato’s *Republic*.”

realizability and mental state inclusion are matters of degree, a fact which has not escaped the attention of critics. Questions raised in the next section include: “What exactly does ‘functional similarity’ mean?” or, more colloquially, “How different is *too* different?”

1.3: The Path is Wide, but the Gate is Narrow: Limitations on Multiple Realizability

Now with a functionalist background theory of mind in place, perhaps Clark and Chalmers aren’t out of their mind when they say the mind is out of the head. In a rough and ready sense, objects out in the world *could* do the same sort of thing as the stuff inside your head. The notebook does serve as a reliable repository of belief for Otto by being the same sort of thing as Inga’s brain. But is “roughly the same thing” enough to ground the radical claim that the spatial boundaries of your mind do not lie at the skin and skull? It is no mistake to point out that there are many differences between a notebook and a brain, besides their respective location. So maybe one can accept that spatial location is really not an essential property of the mind and its constituents, while at the same time argue that there *is* some special qualitative difference between what the stuff inside the head does and what the stuff outside does. This is the line that critics of the *extended mind* thesis, Fred Adams and Ken Aizawa, take in their transparently-titled articles “Why the Mind is still in the Head”¹⁴ and “Defending the Bounds of Cognition.”¹⁵ They give two reasons for why, when considering if something is part of the mind, “roughly the same thing” is not good enough. The first critique is that putative external

¹⁴ Fred Adams and Ken Aizawa, “Why the Mind is Still in the Head,” in *The Cambridge Handbook of Situated Cognition*, eds. Murat Aydede & P. Robbins (Cambridge: Cambridge University Press, 2009), 78–95.

¹⁵ Fred Adams and Ken Aizawa, “Defending the Bounds of Cognition,” in Menary, *The Extended Mind*, 67–80.

vehicles of the mind lack the so called “mark of the cognitive,” arguing that there *are* relevant functional differences between notebooks and the human brain, or in other words, the brain *does* instantiate some unique set of properties that “marks” it as the sole claimant to mental state constitution. The second critique is what Adams and Aizawa refer to as *the coupling/constitution fallacy*. Both the “mark of the cognitive” criterion, which has been variously critiqued as too stringent or ambiguous, and the causal/constitution fallacy have received many replies in the *extended mind* literature.¹⁶ Of Adams and Aizawa’s twin critiques, the coupling/constitution fallacy is antecedent to the mark of the cognitive argument and must be explicated first. Adams and Aizawa put it like this: “The fallacious pattern is to draw attention to cases, real or imagined, in which some object or process is coupled in some fashion to some cognitive agent. From this, one slides to the conclusion that the object or process constitutes part of the agent’s cognitive apparatus or cognitive processing.”¹⁷ To hammer the point home, they provide this clarifying addendum: “coupling relations are distinct from constitutive relations, and the fact that an object or process X is coupled to object or process Y does not entail that X is part of Y.”¹⁸ Remember that on a functionalist view, parts constituting a mental state are determined by the causal role they play in appropriately connecting stimuli, behaviour, and other mental states. It is this conception of what a mental state is that pumped the intuition that Otto’s notebook is just as much a vehicle of Otto’s belief as Inga’s biomemory is of hers. Adams and Aizawa are flatly denying that causal coupling is a

¹⁶ See: Andy Clark, “Coupling, Constitution and the Cognitive Kind: A reply to Adams and Aizawa,” in Menary, *The Extended Mind*, 91; See David Michael Kaplan, “How to Demarcate the Boundaries of Cognition,” *Biology and Philosophy* 27 (4) (2012): 550.

¹⁷ Adams and Aizawa, “Defending,” 70.

¹⁸ Adams and Aizawa, “Defending,” 70.

sufficient condition for parthood. Yet strangely enough, Adams and Aizawa advance a necessary condition for inclusion in a cognitive process based on a specific *causal* profile that the mind supposedly possesses: “Cognition is constituted by certain sorts of causal processes that involve nonderived content.”¹⁹ By “derived content” Adams and Aizawa mean those mental representations that possess meaning by virtue of social convention and public use. For example, traffic lights, police tape, and the Roman alphabet, are examples of derived representations or sets thereof. By “nonderived representations” is meant those publicly inaccessible mental states such as experiences, thoughts, and beliefs, which do not derive their meaning derivatively either from social convention or from more primitive representations. They are *original* in what they are about, and thus possess *original intentionality*.²⁰ Bring this to bear on the case of Otto and his notebook and we can see how the derived/nonderived distinction characterizes Inga’s biomemory as different from Otto’s notebook memory. In order to encode information in his notebook Otto must deal entirely with derived representations, i.e. English words and strings of letters that have meaning only in virtue of their conventional use by a group of speakers; Inga does not. Her biomemory deals in internal representations the meaning of which is primary, non-derived, and supposedly, a result of the specific causal laws of the brain that produces them.

Though Adams and Aizawa avoid the supposed mistake of equating causality with constitution, they express the functionalist spirit that they clearly criticize.²¹ They

¹⁹ Adams and Aizawa, “Defending,” 71.

²⁰ See John R. Searle, “Minds, Brains and Programs,” *Behavioral and Brain Sciences* 3 (1980): 417–457.

²¹ The relationship between constitution and causality is not always clear; dynamical systems based on continuous mutual interaction will be considered later in this paper as a possible instance of when causal relations can be constitutive. Adams and Aizawa are wrong on two counts: what

say that what makes something a part of the mind is its *ability* to deal with nonderived content, that is, what it is made up of may be variously formed, so long as it does the right kind of thing. This certainly sounds like an echo of the functionalist credo, for Adams and Aizawa are clearly emphasizing that it is what parts of the mind *do* that determine whether they are parts of the mind at all, rather than what they are made of or where they are located. Here is their argument:

P1: As essential to their function, all mental states must possess nonderived content.

P2: The only thing nomologically capable of doing so is the human brain.

C: Therefore, as a matter of contingency, the mind just happens to be in the head.

1.4: Functionalism and Mental State Inclusivity

Besides the clear tension between the causal/constitution fallacy and their own functionalist sympathies, there are, I think, at least three reasons why Adams and Aizawa have presented a formulation of functionalism that is too *fine-grained*.^{22,23} Adams and Aizawa's requirement of dealing solely in nonderived content is too stringent a condition for any putative vehicle of the mind. In fact, building a necessary condition for cognition on nonderived content is like building a castle in the sky: there is no widely accepted or

they call a "simple fallacy" turns out to be metaphysically complex and not necessarily fallacious. This is equally bad news for Adams and Aizawa, as for functionalist thinkers who use coupled systems arguments, but, as I will argue later on, for different reasons.

²² Adams and Aizawa need to distance themselves from a functionalist frame of mind if they want their causal/constitution critique to have any bite. For consider what a mind is according to functionalism. A mind is a collection of mental states; a mental state is identical with a functional state; functional states are made up of, i.e., constituted by, the *causal* relations between stimulus, behaviour, and other mental states. Therefore, causality is sufficient for constitution.

²³ Functional *grain* has to do with deciding which psychologically relevant causal factors identify a given mental state.

endorsed theory of nonderived meaning in the philosophy of mind.²⁴ Further, even if the mind as a *whole* ultimately needs to deal in nonderived representations, it does not follow that any *parts* of the mind must.²⁵ Averring otherwise commits the fallacy of division, which is the invalid line of reasoning that moves from the premise that some part of thing *X* has the property *y*, to the conclusion that thing *X in toto* has the property *y*, e.g., the pulmonary arteries, which are part of the circulatory system, bring oxygenated blood from the lungs to the heart; therefore the entire circulatory system itself brings oxygenated blood from the lungs to the heart. This relieves Otto's notebook from the requirement of having to instantiate properties that the whole mind does, and thus renders the nonderived content criterion unable to redraw the boundary between mind and world at the skull. Finally, the demand for dealing in nonderived content does wrong by the functionalist programme of mental inclusivity that it seems *designed* to support.²⁶ Adams and Aizawa assert that in order to support nonderived content, vehicles must possess a unique causal profile, a *natural kind*. There are "brain natural laws that are not identical to physical, chemical, biological or neurophysiological covering laws spanning the cranium."²⁷ They point to the peculiarities of the human mind and its faculties, which appear to exhibit properties that no notebook ever would: Weber's law, primacy and recency effects on memory, "negative-transfer," etc. However, does this not beg the question? Adams and

²⁴ "Tying a demarcation criterion for cognition to non-derived content appears to require nothing less than a complete theory of content, something that continues to elude philosophers of mind after several decades of effort." Kaplan, "How to Demarcate," 549.

²⁵ Robert A. Wilson and Andy Clark, "How to Situate Cognition: Letting Nature Take its Course," in *The Cambridge Handbook of Situated Cognition*, eds. Murat Aydede and P. Robbins (Cambridge: Cambridge University Press, 2009), 69.

²⁶ Mark Sprevak, "Extended Cognition and Functionalism," *Journal of Philosophy* 106 (9) (2009): 513.

²⁷ Adams and Aizawa, "Still in the Head," 80.

Aizawa assume that everything properly called “cognitive” must exhibit all the fine-grained causal qualities of the human brain, such as negative transfer or Weber’s law, despite the fact that what is under dispute is that very question. The appeal to fine-grained-ness as a necessary condition for cognition can be shown to beg the question; it merely *assumes* that any process deserving the label “cognitive” must match actual human cognizing, which exhibits such peculiar causal regularities as Weber’s law and negative transfer. Appealing to fine-grained-ness amounts to *assuming* what needs to be proved – that all mental states must display a high degree of causal similarity to the vehicles typical of actual human cognizing, i.e., brain states.

Andy Clark thinks that such a condition does wrong by a specific intuition, call it *the Martian intuition*, and therefore fails to be a necessary condition that putative vehicles of mental states must meet, contra Adams and Aizawa.²⁸ Andy Clark asks whether an intelligent-seeming Martian needs to display the *fine-grained* functional properties typical of human psychology in order to be included and recognized as a sentient being. For example, on the topic of *pain*, is wincing an essential behavioural product of pain? Perhaps not, for one can imagine a race of Martians who laugh hysterically when they stub their toe. Is bodily harm an essential stimulus to having the mental state *pain*? In most cases, yes, even the Martian who laughs hysterically at having stubbed its toe did so because of harm to its body. One must choose which causal roles are not essential to the characterization of a particular functional state, and thereby able to be ignored, and further, one must decide upon the essential causal roles which are to form a more general kind. Otherwise, if one indexes every single distinct instance of things which cause

²⁸ Andy Clark, “Memento’s Revenge: The Extended Mind, Extended,” in Menary, *The Extended Mind*, 46.

human beings pain, the resulting list would be potentially infinite: stubbing one's toe; stubbing one's toe on a Tuesday; stubbing one's toe on a Tuesday at five o'clock. It would be infinitely briefer to abstract what all members of the list have in common and elevate that property to the status of a *general kind*, such as, in the case of things which cause *pain*, the general kind *bodily injury*.

Consequently, Martians can suffer *pain* not by virtue of being able to suffer every specific instance of *bodily injury*, but by virtue of being able to experience *pain in general*. Put more concretely, Martians may not be able to experience heartburn, a specific and typical source of pain in human beings, but this does not preclude their being able to experience the broad general kind *bodily injury*, which they share with human beings as an essential causal role in the functional state *pain*. To be sure, there is an obvious gradient or *grain* to the characterization of functional states, from specific to general, and the more specific the brand of functionalism, the less inclusive it will be about who or what can have mental states, and what can count as a vehicle of a mental state.

So, the intuition is that Martians need not exhibit the exact psychological subtleties that general human psychologies do. Thus, we should not restrict the functional-grain so much that possible intelligent Martians, despite their differing material or structural make-up, do not count as cognitive beings because they laugh instead of cry at stubbing their toe. This is a direct rebuttal of Adams and Aizawa, whose critique of the *extended mind thesis* was based on the fine-grained functional *dissimilarities* between, say, Otto's notebook and Inga's hippocampus.

If we are functionalists about belief and accept multiple realizability, we lack an adequate motivation to individuate functional roles in a very fine-grained way. With fine-grain individuation of the functional roles,

Otto's and Inga's states have different functional roles, and hence even if Inga can be awarded the belief, Otto perhaps cannot. However, individuated in a coarse-grained way, the differences between the two functional profiles disappear. Otto has a state that plays the same functional role as Inga's belief that MoMa is on 53rd street plays. Hence, we should award the belief to Otto.^{29,30}

It is untenable to cut the functional grain so finely without falling into a kind of neurological chauvinism, for chauvinism is what it is when it is assumed that our mental profile sets the gold standard for mentality in general. If one is committed to the possibility of mental-state-having Martians, or that your dog can have the belief that *the ball has been thrown*, then "roughly the same thing" is in fact good enough for some external object to count as a vehicle of some mental state. Why? Because, as Mark Sprevak argues, if the functional grain is set coarse enough to allow the possibility of intelligent Martians, then it allows for minds with vehicles that lie outside the head.

1.5: Brains, Notebooks, Ectoplasmic Slime – What's the Difference?

The bridge from functionalism to the *extended mind thesis* is completed by Sprevak's handling of the Martian Intuition:

One could imagine a Martian whose memory, instead of being stored in patterns of neural activity, was stored internally as a series of ink-marks. If the Martian wished to store new information, it would activate a process that would create new ink-marks in its storage system. If it wished to retrieve information, it would activate a process that would make a mental image of ink-marks appear in its visual system. It seems wrong to say that simply because a Martian stored its memories this way, we should deduce that it had no mental life, or lacked genuinely mental memory states or processes. In principle, there seems no bar to such a Martian having beliefs and mental states (provided, for example, it exhibited the relevant coarse-grained features of memory, and caused

²⁹ "MoMa" stands for the "Museum of Modern Art," a place customarily referenced in discussing the Otto example.

³⁰ Katalin Farkas, "Two Versions of the Extended Mind Thesis," *Philosophia* 40 (3) (2012): 441.

occurrent beliefs and desires in an appropriate way). Such a creature would have internal states with the causes and effects typical of the notebook in the Otto–notebook system. But just because a creature used ink-marks rather than neurons to store information, we would not conclude that it must thereby lack mental states.³¹

Sprevak argues that any functional characterization of mental states that allows for the possibility of intelligent Martians implies not only the *possibility* of extended minds, but the *actuality* of such. Since it has already been argued that fine-grained renditions of functionalism are faulty for at least three reasons, coarse-grained functionalism will be the working theory of mental state characterization hence forward, thereby implying that intelligent Martians *are* in fact possible.³² However, how does one get from a simple modal statement about extraterrestrials to the bold claim that our minds are actually extended? Sprevak offers one simple argument: “My claim is that if the grain parameter is set at least coarse enough to allow for intelligent Martians, then it also allows in many cases of extended cognition. The justification for the claim is that cases of extended cognition are at least as similar to cases of internal human cognition as possible Martian thought processes.”³³ The crux of Sprevak’s argument is that any species of functionalism that preserves the Martian intuition allows cases of possible Martians whose minds are as strange or stranger than that of Otto and his notebook. If we allow cases of Martians who store beliefs via ectoplasmic slime, or who count using fleshy tubes inside their head, to count as functionally identical or at least similar to actual

³¹ Sprevak, “Extended Cognition,” 512.

³² The reasons being: (a) that fine-grained functionalism is too exclusive in mental state attribution, (b) that it begs the question by assuming that actual human cognizing sets the universal bar for how *all* cognizing must be, (c) that it demands some form of intrinsic suitability of putative material vehicles, such as “dealing in nonderived contents,” for which there is little reason to think is the case.

³³ Sprevak, “Extended Cognition,” 515.

human minds in their ability to have a particular type of mental state, then the case of Otto and his notebook is really not so out of this world after all. But has the case been successfully made for not only the *possibility* of externalism about vehicles, but the *actuality* thereof? The Martian Intuition is powerful in campaigning for the functionalist demand for mental state inclusivity and general realizability. Unfortunately, its power pushes these demands to absurd conclusions about functionalism as a theory of mind.

Chapter 2: The Problem of Cognitive Bloat

2.1: Reliable, Accessible, and Trusted: Nice but not Necessary Conditions for Mental Constitution

As an answer to the question “Why would anybody think *extended mind theory* is true?” I have emphasized the intuitive pull of functionalist reasoning inherent in the capstone quotation of the theory, the *parity principle*. This principle states that “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.”³⁴ The intuitiveness of mental state inclusivity provided by functionalism, cashed out in terms of Martian thought experiments, counteracts the initial oddity of considering that one’s mind might be made of more than just brainstuff. For any given task involving the manipulation of material vehicles bearing propositional content, the boundary of the skull marks no difference in suitability or capability as to what can and cannot serve as a vehicle for thought in some cognitive process. The spatial properties “external” and “internal” merely mark out some contingent relation of a mental vehicle relative to the functionally irrelevant boundary formed by the skull. What the parity principle attests is that being internal or external are really properties of no major functional importance. Deciding that they are *not* was what talk about possible Martians and coarse-grained functional parameters was all about. It is chauvinistic and question-begging to assume that what goes on inside of the head of *homo sapiens sapiens* is the gold standard for mental constitution. The grain of characterization

³⁴ Clark and Chalmers, “Extended Mind,” in Chalmers, *Philosophy of Mind*, 644.

of functional states must be broad enough to ignore the specificity of actual human cognition, but not so broad that mental state conditions are so easy to satisfy that thermometers end up having beliefs about room temperature. Functionalism in moderation is required to establish a mean between extreme mental state exclusivity and extreme inclusivity. At any rate, we get to the *extended mind* on any version of functionalism that allows the possibility of Martians with differing material profiles, because “cases of extended cognition are at least as similar to cases of internal human cognition as possible Martian thought processes.”³⁵ This means, for example, that there is some possible Martian whose mind is structured just like the extended mind of Otto and his notebook. Coarse-grained functionalism entails the actuality of extended minds by permitting the possibility of Martian minds, which is a very strong and surprising modal claim as demonstrated by this argument:

P1: Otto and his notebook actually exist.

P2: There possibly exists a Martian with a functional profile similar enough to Otto and his notebook such that the Martian has an extended mind.

P3: If the possible Martian has an extended mind, then Otto and his notebook, being sufficiently functionally similar, also have an extended mind.

C: Otto has an extended mind.

So the recipe list for an extended mind is simple: a principle stating that internal and external objects and processes are equal claimants to the status of “mental vehicle”; a kind of functionalism that is coarse-grained enough to allow the possibility of intelligent Martians with a different material constitution; and an intuition-pump that shows that

³⁵ Sprevak, “Extended Cognition,” 515.

extended cognition is an actual phenomenon. However, there is a certain special ingredient needed to prevent *cognitive bloat*, which is the rampant and unwarranted mental extension of the mind into the world. Cognitive bloat occurs when mental state inclusion conditions are too easy to satisfy. The problem is that certain external objects that are candidates for being a mental vehicle are ridiculously easy to find. Remember that the work being done in constituting a mind is done entirely by the relevant functional parts, i.e. vehicles, which are those processes and/or objects which causally contribute to an identifiable mental activity or state. It is this functionalist insight that Clark and Chalmers ran with when flying the flag for *extended mind theory*, leading them to formulate the oft-quoted *parity principle*, which just states that the brain/world boundary does not pick out the spatial constraints of the mind. If there are external processes or objects that play the right kind of causal role, and this is all that matters in identifying parts of a functionally-defined system such as a mind, then their being located beyond the skull is a mere contingent fact about them – a fact which by no means diminishes their claim of being constitutive parts of the mind.

Without special conditions and restrictions, the books in a library, the mass majority of which I have never read, and their manifold of content, become part of my mental economy just in virtue of the causal role they can play in supporting a mental state such as *belief*. Clark and Chalmers anticipate some untenable and radical implications of claiming that minds literally extend beyond the skull to incorporate external objects and processes. Consequently, they provide some criteria to be met by such putative external mental vehicles, if they are *really* to be part of one's mind. Known as the "trust and glue" conditions, they are as follows:

1. That the resource be reliably available and typically invoked.
2. That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
3. That information contained in the resource should be easily accessible, as and when, required.³⁶

Why these three conditions in particular? Consider this: merely by picking up a book, you come to believe everything that is written therein. Why is this so? Well, without the condition that any information thus retrieved be more-or-less automatically endorsed, you would essentially have a repository of innate, heretofore unexamined beliefs whenever you interact with a book. The book is functionally coarse-grained enough to conceivably serve as a repository of mental content, just like your hippocampus, and thus would count as part of your (extended) mind. Further, without the condition that the resource be readily available and typically invoked, one could not appeal to the temporariness of the “book-as-memorybase” for a reason to discount it, along with all the information it contains, from being a part of your mental content base. In a similar vein, the lack of applicability of a third condition, that information contained in the resource should be easily accessible, as and when required, means that the scope of things that can constitute your extended cognitive system expands even wider. You have externally instantiated beliefs in your copy of the *Encyclopedia Britannica*, which sits on your shelf in New York, while you vacation in Hawaii. With these additional conditions

³⁶ Clark, “Memento,” 46.

in place, it is harder for an external object to satisfy mental *vehicle-hood*. The contents of a library would no longer count as part of my mental economy, for a library is neither readily available nor is the information contained therein automatically endorsed. Otto's notebook on the other hand, satisfies all three conditions for Otto. The apparent confirmation of the pet intuition-pump of the *extended mind thesis*, Otto and his notebook, along with the simultaneous inoculation against cognitive bloat, leads one to wonder about the adequacy of the "trust and glue" conditions. The question to which I turn now is: "Are these genuine principled conditions or are they merely *ad hoc* maneuvers, tailor-made to increase the plausibility of *extended mind theory*?"

2.2: It's not What's on the Inside that Counts: Mental State Constitution is a Fair Game

I agree with Sprevak in seeing the "trust and glue" conditions as post-facto attempts by *extended mind* theorists at damage control over an absurd consequence of their view – cognitive bloat. Sprevak has us consider Clark and Chalmers' own *parity principle* whose purpose was to level the playing field as to which kinds of players can play ball in the game of mental state constitution. External objects are just as suitable as inner objects so long as they play the appropriate causal role in constituting a given mental state. Yet, in order to anticipate and counter a looming absurd consequence of their position, cognitive bloat, Clark and Chalmers preemptively provide a list of conditions that external vehicles must meet, but that internal vehicles ostensibly need not. If external vehicles are just as functionally relevant as internal vehicles in being able to constitute mental states, then why the differential treatment of stuff outside the head? For something called the "parity principle," these additional criteria are quite prejudiced. If the outer *really were* as

suitable to being the vehicle base of a mental state as the inner, then there would be no need to posit these special criteria that effectively imply that internal vehicles set the norm of suitability, and that putative external vehicles must copy them.

The case against the “trust and glue” criteria is even stronger than just pointing out the tension between them and the *extended mind* keystone, the *parity principle*. Actual cases of *internal cognition* often violate these requirements. For example, it is true that not all of our internally held beliefs are automatically endorsed or typically invoked. We sometimes deliberate over them, doubt them, and even reject them. Neither is it the case that they are easily accessible. All of us have unconscious biases, and possess significant belief structures below the narrow threshold of consciousness, much of which might remain uncovered by all attempts at introspection short of self-hypnosis. Even availability is not guaranteed for in-the-head vehicles. The brain is a material thing; it is thus liable to suffer pathologies and damage, distorting the information it contains – just as water causes ink to bleed and blur the words contained in the pages of a book.

Even worse, parts of the world can meet all three conditions and still be considered dubious as vehicles of a person’s mental states. Here is Katalin Farkas with an example:

Suppose someone, call her Lotte, always carries an electronic reading device. She downloads a 37 volume history of Europe with a quick search function, from a source she trusts completely. If any question about the history of England comes up, Lotte consults the book. The device satisfies the criteria of easy accessibility, reliable availability, and it is subject to automatic endorsement – these were, according to Clark and Chalmers, the main criteria of something counting as a belief. Does Lotte thereby acquire all the beliefs, and hence does she become an expert on the history of England? To say that she does is at least *prima facie* problematic. We do want to distinguish between an

impressively erudite scholar of English history who mastered the subject, and Lotte, who simply stores a file on a device.³⁷

In sum, the “trust and glue” criteria are insupportable for three reasons: (a) they clash with the *parity principle*, which is the prime motivation of the entire *extended mind* movement; (b) they effectively postulate that there are some intrinsic properties, i.e. accessibility, availability, and immediate endorsement, that only internal vehicles possess, and it is the duty of external vehicles to copy internal ones as best they can; (c) all three conditions can be satisfied and yet not be sufficient for constituting a mental state with external vehicles. For a theory that claims that a vehicle’s being inside or outside the head makes no functional difference, the “trust and glue” conditions imply that there is quite a disadvantage to being on the outside of the skull looking in. Further, it is a matter of empirical fact that internally held propositional states are not readily available, easily accessible, or automatically endorsed by the thinker. To hold that *all* instances of external cognition must adhere to these conditions, while internal ones occasionally (perhaps *routinely*) violate them, is just another inconsistency espoused by the view that internality and externality have nothing to do with mental vehicle suitability. The rejection of the “trust and glue” conditions means that cognitive bloat remains a live issue, leaving *extended mind* theorists without an obvious response.

2.3: Raindrops are Falling on my Mind: The Proper Parts of Mental States

Vehicle externalism is radical in that it suggests an “anything goes” attitude towards what can be a part of the mind. Mental vehicles can come in all shapes and sizes, e.g. they can

³⁷ Farkas, “Two Versions,” 444.

be made of paper, brain matter, web pages, etc. It might have been the right move to subvert the skin and skull hegemony of the traditional vehicle internalism that went unchallenged in philosophy of mind for decades, but with the failure of the “trust and glue” conditions, *extended mind theory* stands to fall into a mereological nightmare with no way of telling what is a part of one’s mind and what is a part of one’s environment:³⁸

“How far can we extend the mind? [. . .] If the mind is spatially extended at all – in the sense that mental features have a physical basis which is spatially extended – then there seems to be no general reason to confine the extension inside the skull. Similarly, there seems to be no reason to limit the extension once we are outside the skull.”³⁹ *Extended mind theory* is currently without boundaries; the fence-posts have been removed from the traditional range of the “cognitive,” the skull, but have not been replaced at any other non-arbitrary property line.

The spatially scattered parts of the mind need a unifying property in virtue of which they can be distinguished from the environment at large. It may be true, as the *parity principle* attests, that some parts of our mind *can be* located in the environment, but the principle remains vague as to which ones, in particular, are suitably poised to constitute a mental state. However, if the question is “What are the *proper parts* of a mental state?”, then there is a teleological story to tell, which may hold the answer to the question of cognitive bloat. Here is Andy Clark using the notions of “design” and “selected for” as if they make *the* difference between an external object that is part of the mind and one that is not:

³⁸ Mereology is the study of the relations between parts and wholes.

³⁹ Farkas, “Two Versions,” 444.

Suppose the rhythmic pulse of rain on my Edinburgh window somehow helps the pace and sequencing of a flow of thoughts. Is the rain now part of my cognitive engine? No. It is merely the backdrop against which my cognizing takes shape. But this, I submit, is not because the rain is outside the bounds of skin and skull. Rather, it is because the rain is not part of (it is not even a side-effect or a “spandrel” within) any system selected or maintained for the support of better cognizing. It is indeed mere (but as it happens helpful) backdrop. Compare this with a robot designed to use raindrop sounds to time and pace certain internal operations essential to some kinds of problem solving. Such a robot would be vulnerable to (non-British) weather. But it is not clear, at least to me, that the whole drop-based timing mechanism is not usefully considered as one of the robot’s cognitive routines.⁴⁰

Perhaps some theory of *proper functions* can be utilized to save the *extended mind* position from its boundary conundrum as Andy Clark seems to intimate above.⁴¹ In general, proper functions work like this: to determine if something is a particular kind of thing, for example the kind *mental vehicle*, the specific etiological history of all mental vehicles, both external and internal, would have to be considered.⁴² History is relevant to kind specification because it is what members of the same kind did in the *past* that explains what members of the kind do *now*. Further, what members of the kind *do* explains the survival and continued reproduction of the members of the kind. To illustrate this idea consider what makes a heart, a heart. What is essential to being a heart is its function, which is to pump blood.⁴³ But not all hearts pump blood; some hearts are dysfunctional or malformed, and there are some things which are not hearts but *can* pump blood, such as hand-pumps. What this means is that function alone is neither sufficient

⁴⁰ Andy Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (Oxford: Oxford University Press, 2011), 130.

⁴¹ Ruth G. Millikan, *Language, Thought and Other Biological Categories* (Cambridge: MIT Press, 1984), 17.

⁴² *Etiological* referring to the specific chain of causes and effects that comprise a thing’s or a kind of thing’s being.

⁴³ Millikan, *Biological Categories*, 17.

nor necessary for determining what something is essentially. There must be an appeal to *history* in order to hone in on what makes a heart, a heart, and a pump, a pump. This is what the notion of *proper function* delivers: today's hearts pump blood in virtue of the fact that ancestor hearts did and in so doing passed on this fitness-increasing trait to subsequent generations.⁴⁴ Therefore, the proper function of a heart is to pump blood even if a certain heart doesn't actually do so or is unable to, while the proper function of a hand-pump is not, even though it *could* be used to. Though talk about "proper" functions sounds normative, it can be cashed out in a descriptive and naturalistic way; through the framework of natural selection, there exists a simple means of explaining why the presence of predecessors with a trait is the reason current members have the trait: it increased the fitness of the past members of the kind, who passed it down to subsequent generations via the heritability of the trait.

Andy Clark seems to suggest that mental vehicles can be defined in a similar manner, i.e. that any mental vehicle without the right selective history is no vehicle at all. Mental vehicles *could* pump blood, but that surely isn't essential to their kind; mental vehicles are not of the kind *hearts*. The history essential to the kind *mental vehicle* is something more like this: the proper function of a mental vehicle is to be disposed to realize mental states. There is no doubt that the most uncontroversial mental vehicle is the brain, but it is possible that parts of the environment, especially artifacts, can be amenable to the same kind of proper function explanation as the brain. For example, one could say that Otto's notebook possesses the proper function of retaining information for later

⁴⁴ "Fitness" is the evolutionary notion that describes the overall capacity of an organism or, more specifically, its genes, to deal with the challenges of its environment, thereby conferring increased probability of its occurrence in subsequent generations (if and only if it is heritable.)

access because this is the property that ancestor notebooks possessed which explains not only the “reproductive success” of notebooks, but also Otto’s usage of the notebook as an externally-located memory bank; it was selected for the very purpose of being a vehicle of Otto’s mental states, particularly his belief that *the art exhibit is on 53rd street*. So it is a part of Otto’s mind in virtue of its having the appropriate proper function of being a mental vehicle for human mental states. Members of the kind *notebook* are selected for the purpose of realizing mental states just as the human brain is.

Maybe Andy Clark is right: conceivably the mind is made up of those objects and processes, either internal or external, that belong to the kind *mental vehicle* which are identified by their historical and adaptive success in realizing mental states. If we want to know whether something is a part of the mind, just reconstruct its selection history, derive what its proper function is likely to be, and see whether what it was selected for is relevant in explaining the occurrence of a certain mental state or cognitive process. Proper functions look like the saving grace for the *extended mind thesis* in the face of cognitive bloat. It does provide a desperately needed necessary condition for inclusion into the kind *mental vehicle*. Unfortunately, there is one simple thought experiment that wholesale discredits proper functions as at all relevant to mental constituency.⁴⁵ The *swampman* thought-experiment owes its conception to philosopher Donald Davidson who imagined a scenario in which he was reduced to his most basic material constituents by a lightning-bolt. Simultaneously, another bolt of lightning strikes a nearby tree miraculously combining matter in such a way as to constitute a perfect clone of the original, but now molecularized, Davidson. The implication of the thought-experiment is that Swamp-

⁴⁵ Sean Allen-Hermanson, “Superdupsizing the Mind: Extended Cognition and the Persistence of Cognitive Bloat,” *Philosophical Studies* 164 (3) (2013): 800.

Davidson lacks any of the causal history of the original, yet is physically (but not numerically) identical in every regard (including possession of mental states) and, therefore, causal history is irrelevant to function.

In an applied version of Davidson's *swampman* argument, Sean Allen-Hermanson spares no effort in questioning the significance of Clark's raindrop-timing robot:

Take *swamp-robot*, a perfect physical duplicate of its counterpart actually designed (or evolved) to spit water droplets. Swamp-robot's behavior and internal processes are just like the real thing, but it did not evolve, and was not designed, and so lacks a recruitment history. So, Clark, implausibly, would have to deny that swamp-robot's spitting-raindrop mechanism is part of its cognitive process. Yet it falls under the same causal-explanatory kind as the original spitting-robot, in virtue of the fact that exactly the same laws and models apply when it comes to explaining how the timing-mechanism (or swamp-mechanism) boosts performance.⁴⁶

By Clark's proposed etiological criterion, we would be forced to deny the swamp-robot mental states and cognitive processes while at the same time attributing them to the original *despite* the complete physical identity between the two. I think even the example of Martians can be deployed against Clark for showing that causal history is irrelevant to function. Imagine a world covered by swamps and bombarded by a perpetual continent-sized lightning storm. Given enough time, enough lightning strikes, and faith in the law of large numbers, an eventual probabilistic miracle occurs! One Martian is spontaneously generated, then another, then another, until there is a small population of Martians. One might observe that they are *behaving* civilly and intelligently, appearing to possess desires, share beliefs, and publicly express thoughts. But only a behaviourist or Andy

⁴⁶ Allen-Hermanson, "Superdupersizing," 800.

Clark would stop at behaviour and say that they didn't possess any mental states or have minds of their own. A functionalist would say that they most certainly do, despite the complete lack of selective history.

2.4: The Extended Mind as a Coupled System

The terms “designed” and “selected for” in Clark's raindrop illustration do nothing to find the sorely-missed boundaries of the mind; cognitive bloat remains a live issue. However, going back to the roots of the *extended mind* program, Clark and Chalmers did present an alternative framework for the understanding of an extended mind. Consider the notion of *coupled systems*: “The central feature of these cases is that the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in such a system play an active causal role, and together they govern behavior in the same sort of way that cognition usually does. Remove the external component and the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head.”⁴⁷

The key aspect here is the notion of “two-way interaction.” Orestis Palermos singles out this feature of coupled systems as the solution to the problem of cognitive bloat and advances two arguments, both reliant on *dynamical systems theory* (DST), in favour of his conclusion.⁴⁸ Dynamical systems theory mathematically models the

⁴⁷ Clark and Chalmers, “Extended Mind,” 644.

⁴⁸ Orestis S. Palermos, “Loops, Constitution and Cognitive Extension,” *Cognitive Systems Research* 27 (2014): 26.

continuous mutual interactions (CMI) between two systems which are engaged in a two-way relation of reciprocal and simultaneous causation. Systems are “sets of interdependent elements, objects, entities, or items standing in interrelations on the basis of specific processes in which they take part and to which they give rise, thereby forming a unified whole.”⁴⁹ In this case, the unified whole is a coupled system. Arguments for the existence of a coupled system over and above the two subsystems engaged in CMI can be made in two independent ways. The first, called *the Systemic Properties Argument*, is the postulation of a coupled system to explain the emergent behaviour of two systems bound by a relation of continuous mutual interaction. If the interactive behaviour of two such systems cannot be explained with reference to the workings of either system individually, then one must posit a coupled system over and above either individually, which explains the new and unique behaviour. Further, the resulting coupled system cannot be reduced to either subsystem, for it alone instantiates behaviour which is beyond the capacity of either subsystem to produce individually. In short, behaviour produced by two independent systems conjoined by a relation of continuous mutual interaction cannot be explained by anything *but* a coupled system that contains both as parts; in other words, “the postulation of coupled systems is necessary with respect to the explanation of certain systemic properties, which we would otherwise be at a loss how to account for.”⁵⁰

The second argument Palermos refers to as *the Ongoing Feedback Loops Argument*. *Feedback* is the simultaneous and reciprocal affecting on one system by another, influenced in real time by how each respective system affects and is affected by the other. Feedback, or feedback loops, can be contrasted with traditional input/output

⁴⁹ Palermos, “Loops and Extension,” 29.

⁵⁰ Palermos, “Loops and Extension,” 32.

linear relations between systems; that is, there is no input/output phenomenon between systems but rather loops of continuous mutual affecting happening in real-time.

Consequently, feedback loops are essential to the identity of coupled systems because they make it impossible to decompose any real coupled system into two separate systems on the basis of the linear input/output relations that would stand between them. Two subsystems in a coupled system simply do not stand in such a linear relation but instead are linked via a relation of mutual and simultaneous interaction, where the causal effect of one subsystem upon the other is partially determined by the affected system's own affecting. Consequently, there are no discrete inputs and outputs between systems. Each system receives feedback from affecting the other, which is determined by the way the system *itself* has affected its sister system. Without discrete inputs and outputs between systems, we are faced with an indecomposable coupled system, whose parts are engaged in causal feedback loops which generate systemic properties attributable to neither of the subsystems. Such systems are currently posited to exist in explaining and modeling the complex interactions between two systems engaged in real-time mutual causation, including "two mutually interconnected pendulums, the watt governor and a rotation engine, and, possibly, cognitive agents and their epistemic artifacts."⁵¹

In the case of cognitive agents and their artifacts, we identify a coupled system constituted by a cognitive agent and the appropriate environmental factors by asking if the agent, as our starting system, has any direct effect on the environment such that the affected environmental aspects *in return* affect the agent's cognitive performance.⁵² If not, then the acted-upon environmental aspects can be safely relegated to background

⁵¹ Palermos, "Loops and Extension," 31.

⁵² Typically, the human organism.

conditions, which are causally relevant, but fail to be a systemic part of the cognitive agent. Conversely, if the environmental aspects *do* demonstrate an ability to affect the agent's cognitive performance in a continuous and feedback-oriented way, then there is a coupled system over and above the organism-centered agent.

Palermos' criterion can be applied to a few cases of putative cognitive extension. In the previously mentioned case where the raindrops pounding against the window generate a cadence which is conducive to philosopher Andy Clark's writing process, Clark is right to say that they are not a part of his cognitive routine. The reason in this case being that there is no *continuous mutual interaction* between himself and the raindrops. There is a linear causal relation between the raindrops and Clark: the raindrops are causally affecting his writing process, but Clark is in no way, shape, or form, affecting the raindrops, least of all in a feedback-oriented fashion necessary for the constitution of an extended coupled system. In the case of the robot designed to use raindrops to facilitate a cognitive process, such as adding a sum, there is no continuous mutual interaction either; only the similar linear causal relation that also runs between the raindrops and Clark.

In the most-famous case of cognitive extension, namely Otto and his notebook, Palermos' reasons that there is no coupled system after all. "Otto does not continuously interact with his notebook. Instead, when he needs some old information he looks it up and when he needs to store some new information he writes it down. It might, therefore, be true that Otto's relationship with his notebook is best described as a case of one-way dependence and thus, by the lights of the CRC criterion, does not count as a case of

cognitive extension.”⁵³ Though Clark and Chalmers’ hallmark case of cognitive extension fails to satisfy the CMI test, Palermos’ assures the reader that “it is not in danger of being so strict as to render cognitive extension a metaphysically possible, yet so far unrealized phenomenon.”⁵⁴ In pursuit of this point Palermos offers examples of actually extended cognitive phenomena including an astronomer using a telescope to gather information about celestial bodies: “Consider, for instance, telescopic observation. Making observations through a telescope is a dynamic process that requires a great deal of experience in operating the artifact. Moving the telescope around while adjusting the lenses, generates certain effects (e.g., shapes on the lens of the telescope), whose feedback drives the ongoing cognitive loops along. Eventually, as the process unfolds, the coupled system of the agent and his telescope is able to identify – that is, see – recognizable objects in space (e.g., stars, planets, comets, galaxies et cetera).”⁵⁵

All in all, Palermos gives two arguments to stem the tide of untenable cognitive extensions threatened by cognitive bloat. The systemic properties argument exploits the fact that two systems reciprocally affecting one another generate a set of properties not ascribable to either system individually. Consequently, there must be a coupled system constituted by the two subsystems which grounds all the properties generated by the two subsystems through their activity of engaging in continuous mutual interaction. For this reason, coupled systems are explanatorily and even *ontologically* necessary in the description of phenomena whose operation cannot be understood through decomposition into component parts. The ongoing feedback loops argument prevents the decomposition

⁵³ Palermos, “Loops and Extension,” 38.

⁵⁴ Palermos, “Loops and Extension,” 39.

⁵⁵ Palermos, “Loops and Extension,” 38.

of coupled systems into parts engaged in linear, output/input relations. Such relations do not hold between parts of a coupled system for the following reason: the causal relation between the systems is reciprocal and partially self-determining, as is typical of feedback loops. The way one system is affected by the other is always partially determined by how the affected system is influencing the affecting system. In this way there is no distinct input/output between systems since all such feedback-based interaction is a product of *joint* determination. Jointly the two arguments above ground the fact that coupled systems are necessary for the explanation of super-systemic properties, that coupled systems are ineliminable by decomposition, and that they efficiently describe the boundaries of possible and actual extended cognitive systems by providing a necessary and sufficient condition that is neither too strict nor too lax, i.e., continuous mutual interaction.

2.5: One of These Things is not Like the Other: Where is the Mind in a Cognitive System?

What Clark and Chalmers introduce and Palermos fleshes out is a principled working model for an extended cognitive system based on the coupling between environmental systems and the cognitive core of human neurological systems. Coupled systems have the advantage of falling under a successful explanation scheme, dynamical systems theory, which already describes and predicts the holistic workings of things like “two mutually interconnected pendulums, the watt governor and a rotation engine.”⁵⁶ Further, coupled systems preserve the functional parity intuitions that motivated the *extended mind thesis* in the first place: a coupled system is multiply realizable, not bounded by the spatial or

⁵⁶ Palermos, “Loops and Extension,” 31.

material requirements of vehicle internalism. Lastly, coupled systems so described have access to a much needed limiting condition on what exactly the vehicular constituents of a cognitive agent's mental economy are: only those that are engaged in a reciprocal, non-linear relation of causality. Tellingly, Otto and his notebook do not cut it as a coupled system predicated on continuous mutual interaction. Does this mean Palermos' criterion is too strict? No. If the only alternatives to delimiting a *cognitive system* are vehicle internalism or coarse-grained functional parity, then Palermos strikes a nice middle ground between arbitrary exclusiveness and inclusiveness respectively.

However, Palermos has given only conditions for the possibility of extended *cognitive systems*, not of *extended minds*. An extended mind is an entity with higher metaphysical hurdles; *extended mind theory* is the claim that *mental states*, not cognitive processes, can be extended beyond the confines of the head. Palermos, in fact, notes the difference in magnitude, and admits that the case for the *extended mind* is a harder one to make than his argument from continuous mutual interaction between members of a coupled system allows.⁵⁷ An account of cognitive activity that acknowledges the explanatorily constitutive status of external objects and processes is really not so strange. Palermos depicts a coupled system whose parts engage in continuous mutual interaction as an enlightening and precise way to understand how human intelligence can depend on factors besides the brain. *Extended mind theory* construed as an empirical hypothesis about the "proper unit of cognition" and the description of "extended cognitive systems" requires no radical metaphysical commitments or reconceptualizations of our notions of selfhood, identity, and knowledge attribution. Extended cognitive systems are not *minds*. I

⁵⁷ Palermos, "Loops and Extension," 25.

think that Clark and Chalmers conflate two very different things, both in their explanatory requirements, and in their claims about ontological status, when they appeal to coupled systems as a model for an extended mind.⁵⁸ What Clark and Chalmers have described is a model for understanding “the right unit of analysis in cognitive science” by arguing that “understanding cognition requires paying attention to the neural, organismal (i.e., extra-neural) and environmental (i.e., extra-organismal) elements of the cognitive system.”^{59,60} But what Clark and Chalmers started with by introducing Otto and speaking extensively about the possibility of externally realized mental states was the claim that “it is possible to point to cases in which the particular network of causes and effects that (according to functionalism) makes up a given mental state is at least partially realised by something in the organism’s external environment.”⁶¹ Clark and Chalmers’ use of coupled systems in their arguments for the *extended mind thesis* are intended to support a different claim from that supported by the Otto thought experiment. The coupled systems argument is essentially an attack on modern cognitive science for its tendency to neglect the rich and complex ways extraneural objects and processes contribute to cognition as we experience it. Clark and Chalmers’ more famous and inventive argument, the Otto thought

⁵⁸ Here is Drayson with the distinction: “What is mind? Mental states are traditionally thought of as the ascriptions of ‘folk-psychology’: the commonsense everyday states we use to describe, explain, and predict our fellow human beings’ actions. Prime examples of mental states are experiential states (such as pains and other sensations), and thoughts: propositional attitudes such as beliefs, desires, hopes, fears, expectations, and intentions.” Regarding cognitive science she says: “What is cognition? The cognitive sciences – including psychology, artificial intelligence, linguistics, robotics, and neuroscience – aim to understand minds and intelligent behavior by studying its mechanisms (where these are usually understood to be computational). When cognitive science tries to explain traditional mental phenomena, it can end up positing different processes to those of our everyday understanding.” Zoe Drayson, “Extended Cognition and the Metaphysics of Mind,” *Cognitive Systems Research* 11 (4) (2010): 374–375.

⁵⁹ Armin W. Schulz (2013), “Overextension: The Extended Mind and Arguments from Evolutionary Biology,” *European Journal for Philosophy of Science* 3 (2) (2013): 243.

⁶⁰ Schulz, “Overextension,” 243.

⁶¹ Schulz, “Overextension,” 242.

experiment, is the claim that functionalism entails a recognition that the vehicles which realize mental states might not be exclusively what happens in the head. The claims belong to two completely different domains with different aims: the coupled systems argument concerns the proper unit of analysis for cognitive scientists, i.e., what the boundaries of the human information processing unit are; the Otto thought experiment concerns the metaphysical possibility of vehicle externalism, i.e., where the mental states of belief, thought, and desire, are locationally realized.

In sum, the possibility of *extended cognitive systems*, but not of minds, can be cashed out in a way that (a) allows it to fall under the rubric of a successful scientific theory, (b) does right by functionalist demands for multiple realizability, and (c) possesses a limiting condition such that it is neither too inclusive nor too exclusive in recognizing which parts of the environment are relevant in constituting an extended cognitive system. I contend, therefore, that only a weaker conception of the Clark and Chalmers project can avoid the problem. Cognitive processes can span brain, body, and world in the sense that the members of systems that engage in the right kind of coupling relation are materially neutral: an astronomer charting the sky with a telescope is an extended cognitive system; a mathematician armed with a calculator is an extended cognitive system; perhaps even a thermostat closely engaged in monitoring room temperature is an extended cognitive system.⁶² Not only is the original claim that mental states can have vehicles that lie

⁶² Thermostats deal in representations about the temperature of their environment in a continuous and mutual way. The temperature of the room, and the thermostat's affecting of room temperature form a feedback loop indicative of a coupled system, as we have seen. The system may be extended, but is it cognitive? It involves information processing so in a broad sense, yes. The more interesting question would be "Is there a mind present with beliefs and desires concerning the ideal room temperature?" It is this difference between cognitive processes and minds that lead me to focus on the latter in the context of folk psychological states and vehicle externalism.

outside the head merely side-stepped by reframing the problem to be about cognitive systems and not minds, but also, there is something very important that gets lost in all this abstract talk of systems and feedback loops: the *person* whose mental states are being described. Clark and Chalmers argue that persons match the extension of their minds. “The information in Otto’s notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto himself is best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity. Far better to take the broader view and see agents themselves as spread into the world.”⁶³ I think Clark and Chalmers are dead wrong. If a coupled system is not sufficient to constitute an extended mind, it is far from being sufficient to constitute a person. Even if one can buy the claim that there is an extended cognitive coupled system comprised of Otto and his notebook, Otto is not identical to the system, nor is he ontologically demoted from personhood status by becoming a part of the system. The reasonableness of this is rendered clear by a consideration of *persistence conditions*. Otto can surely survive losing his notebook; the coupled system might not be so hardy. Otto can also surely survive using his notebook, thereby becoming a part of a coupled system that includes the notebook, without fear of disappearing as a person. In fact, there are many qualities of mindedness that no coupled system will ever capture: “Several more general things can also be said about minds: the concept of mind is entwined with both our everyday and our philosophical concepts of selfhood and persons, of rationality and responsibility, of free-will and morality.”⁶⁴ In my

⁶³ Clark and Chalmers, “Extended Mind,” in Chalmers, *Philosophy of Mind*, 650.

⁶⁴ Drayson, “Metaphysics of Mind,” 374.

final chapter, I will argue that vehicle externalism is a nightmare for our notions of personhood and personal identity.

Chapter 3: The *Extended Self*

3.1: Where am I?

The hallmark argument for the *extended mind thesis* holds that if some external object functions in such a way that, if the function were performed in the head, we would consider the object a part of the mind, then it *is* a part of the mind. Right off the bat a problem presents itself: parthood is a transitive relation. So, if this external object or process is part of my mind, and my mind is a part of me, then this external object or process is a part of me. Applied to the well-worn example of Otto and his Notebook, we are committed to the notebook being not only a part of his *mind*, but a part of *him* in virtue of the fact that it is a part of his mind. Am I, then, to be charged with criminal negligence causing bodily harm if I spill my glass of orange juice on Otto's precious notebook? It is a part of him after all. This is weird and unintuitive at best, but I argue that it is worse – it is absurd. It is absurd because it implies “extended persons,” the existence of which infringes on the notoriously messy notions of personhood and personal identity.

Take the most popular current theory of personal identity, psychological continuity theory, which takes persons to be identical with a continuous series of mental states unified by memory. This implies that what we are made of are mental states, related in the appropriate way via memory. This further implies a mereological corollary relevant to making an extended mind seem an absurd notion: mental state internalism, which says that I am located (not entirely) wherever any of my mental states are. Say my belief, which is a mental state, that all men are mortal is written down in my first-year philosophy notebook. If the conjunction of (a) mental state internalism and (b) vehicle externalism (*extended mind*) is true, then I am literally located wherever my notebook is

located. But this implies that I am *not* located in my feet, my elbows, or my hair. Why? Because none of these things are (usually) a vehicle for any of my mental states, hence, according to the psychological continuity view of personal identity, they are not parts of me.

Externalism about the self is what leads Daniel Dennett to ask, “Where am I?”⁶⁵ Through a fanciful thought experiment involving his envatted brain controlling his ambulatory body, which houses his point of view, he illustrates that “you” can be extended across space and, even more provocatively, across *non-overlapping* volumes of space – much like how a remote control is related to an RC car. On the *extended mind theory*, this intuition is pumped for all its worth, spreading the minds of individuals from the brain to notebooks, the internet, and even social institutions, with no hope of reasonable containment.⁶⁶ The question I consider here is an echo of the question Dennett asks. The *extended mind thesis* provides some real interpretive difficulties concerning self-constitution, despite being clearly and resolutely endorsed by its authors. Clark and Chalmers query:

What, finally, of the self? Does the extended mind imply an extended self? It seems so. Most of us already accept that the self outstrips the boundaries of consciousness; my dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then these boundaries may also fall beyond the skin. The information in Otto’s notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto himself is best regarded as an extended system, a coupling of biological organism and external resources.⁶⁷

⁶⁵Actually, Dennett’s real purpose is to argue for a *reductio* of the self. I remain a realist about the self but agree that the concept is spatially flexible and materially vague, as Dennett’s thought experiment shows. See Daniel C. Dennett, “Where am I?” *Brainstorms* (Cambridge: MIT Press, 1981), 333–347.

⁶⁶One can still avoid cognitive bloat, but only at the cost of endorsing the weaker claim about *extended cognitive systems* discussed previously.

⁶⁷Clark and Chalmers, “Extended Mind,” 650.

In a nutshell, Clark and Chalmers' argument for the *extended self* is this:

P1: A person is essentially a collection of mental states.

P2: Mental states can be realized in vehicles that are spatially removed from the person's organismic body.

C: Therefore, a person can have parts external to their organismic body.

This simple argument requires some big assumptions to make it work. First, Clark and Chalmers assume psychological continuity as their theory of personal identity, which may be the theory most widely adhered to, but is still deeply flawed. Second, Clark and Chalmers are committed to what Eric T. Olson calls *mental state internalism*, which is the assumption that a person is partially located *only* wherever one of their mental states is located and, as I suggested at the beginning of this chapter, this commitment provides ample ammunition for critiques from mereological principles.⁶⁸

3.2: Locke, Stock, and Personal Identity as Psychological Continuity

To explain the first assumption of personal identity as a continuum of psychological states, there are two competing kinds of continuity which are most often stipulated as essential to the identity of persons over time: spatiotemporal continuity or psychological continuity. Today, psychological continuity is the dominant theory, owing much of its formulation to the work of John Locke, especially his famous "prince and cobbler" thought experiment.⁶⁹ Locke argues that a person is not identical to the body they possess in space and time, but to a continuity of psychological states unified by memory. To

⁶⁸ Eric T. Olson, "The Extended Self," *Minds and Machines* 21 (4) (2011): 484.

⁶⁹ Robert A. Wilson, and Bartłomiej Lenart, "Extended Mind and Identity," in *Handbook of Neuroethics*, eds. Jens Clausen and Neil Levy (2014), 424.

illustrate his argument, Locke conjures up a thought experiment: a certain prince and cobbler go to bed one night (not together, but in their respective homes) and wake up finding they have switched bodies. That is, the whole range of psychological states belonging respectively to the prince and to the cobbler, including memories, personality traits, and beliefs, have wholesale switched from one body to another. Now, say the prince has committed some crime prior to the transmigration of his psychology. After the swap, agents of the king come to take “him” to trial. But who is “he” at this point? Locke argues that, while they take away the body of the Prince, they have actually taken the Cobbler as a *person*, and intend to punish him for what the Prince has done. Clearly this is a gross miscarriage of justice! For Locke, and anyone who believes that personal identity is tied to psychological continuity, they got the right body, but the wrong *person*.

The role of memory in unifying the continuity of mental states is essential in Neo-Lockean traditions of personal identity. Memory unifies by allowing an autobiographical narrative to emerge out of a series of psychological states. In this sense, we all are bundles of memories, sorted into a narrative with which we, *qua* persons, are identical. Now, to repeat Clark and Chalmers’ argument in fuller detail, the identification of persons with a continuity of mental states unified by memory, also known as *the autobiographical self*, gives their argument credence by the lights of a widely supported theory of personal identity.⁷⁰ In short, the argument from an extended mind to an extended self is as follows:

P1: Mental states that are unified by memory into an autobiographical whole are what a person is.

P2: Such states can be extended.

⁷⁰ Wilson and Lenart, “Mind and Identity,” 431.

C: Therefore, persons can be extended.

Unfortunately, however, the central role of memory in providing continuity and coherence to the succession of mental states that constitute a person is, I think, *the* fatal flaw of all Neo-Lockean personal identity traditions, and Clark and Chalmers get caught in its dredge net. Any theory of identity based on memory faces at least two major conceptual issues: a) Intermittent existence, and b) the circularity of memory.⁷¹ If your continued existence through time is based on an autobiographical string of memories, then there are bound to be gaps in that string. In fact, gaps will occur whenever you are unconscious, such as when you sleep or become intoxicated. When the continuity of memory is broken by such lapses, you effectively do not exist whenever you are not conscious. Your existence is intermittent; your life a series of temporary deaths and resurrections. Further, any account of identity based on memory appears to be circular. In order to know that the person who visited the zoo with his grandma when he was five was me, I need to already know that my memories are veridical before I can use them as proof of my identity with this five-year-old boy.⁷² In order for memory to prove sameness of person, they must be *my* memories. However, in order to be *my* memories, we must assume sameness of person – hence a circularity. So although marrying autobiographical identity and *extended mind theory* would lead to some interesting metaphysical and ethical conclusions, there are conceptual problems at a deeper level concerning memory as the unifying function of the self through time. Consequently, Clark and Chalmers only inherit old and insufficiently addressed problems by endorsing the Neo-Lockean position.

⁷¹ Eric T. Olson, “Personal Identity,” *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), ed. Edward N. Zalta, <https://plato.stanford.edu/archives/sum2017/entries/identity-personal/>.

⁷² By “veridical” here is meant “corresponding with how the world actually was/is.”

3.3: Mental State Internalism

As mentioned earlier, the second big assumption that Clark and Chalmers make in support of the *extended self* position is *mental state internalism*: “A being’s mental states can never extend beyond its own boundaries. We might call this *mental state internalism*, as it says that a being’s mental states must be located entirely within it.”⁷³ This assumption of Clark and Chalmers owes its formulation to Eric T. Olson who begins with distilling what the *extended mind thesis* says in its strongest terms: any psychological being (a being with mental states) could extend beyond its skin. Further, since mental states are parts of a psychological being, parts of a psychological being can lie outside its skin. So Otto, being a psychological being, by virtue of having mental states instantiated by the notebook, has parts outside of his skin. However, consider the basic mereological principle *an object is located wherever any of its parts are located*. This implies that Otto, *qua* subject of mental states, is at least partially located in the notebook. Given that this is the case, Otto gets bigger when he picks up the notebook. But no *biological organism* would actually get bigger by doing so. What does this mean? According to Olson, it means that Otto, or any psychological being, is *not* an organism. The only relevant organism in the story is Otto’s body, and it certainly doesn’t get bigger by picking up a notebook. It follows then that Otto is not an organism. He neither was nor ever will be since he has the disposition of having externally located mental states, a property *never* instantiated by any organism.

So, “if (owing to the truth of the extended-self thesis) Otto’s possession of external mental states implies that he is never an organism, then no one who has external mental states at some time is ever an organism. Having external mental states at some

⁷³ Eric T. Olson, “The Extended Self,” 484.

time and being an organism are incompatible properties.”⁷⁴ It is not only the *having* of external mental states that is incompatible with being an organism, but also the *possibility* of having them. For it is true on the strongest interpretation of the *extended mind thesis* that any mental state *could* be externally realized. That any human being could pick up a notebook and extend themselves to contain external mental states leads Olson to say, “In fact, none of us is an organism.”⁷⁵ It sounds strange enough on the surface that no person is an organism, but what are the actual implications of such a position? Right away, the problem of “too many thinkers” is raised by Olson: “If I am not an organism and my animal body is something other than me, and that organism thinks just as I do, then I am one of two beings thinking these thoughts and writing these words. More generally, there are two thinking beings wherever we thought there was just one: a person who is not an organism and an organism that is (presumably) not a person.”⁷⁶

There is a *person* capable of thought, but there is also an *organism* capable of thought. Both have a working brain, after all. In fact, it’s the same brain shared between them. This is a strange consequence, that there are two entities, numerically distinct, supervening on the same material base. It seems that this paper is currently being co-authored by both a psychological being and an organism when this whole time I thought there was only one thing writing it. Olson calls this the *too many thinkers* objection. Effectively, this objection doubles the number of intentional agents in the world, for, as long as there are organisms that are capable of having mental states, there is a person in addition to the organism which co-owns those states. Further, this troubling question now

⁷⁴ Olson, “Extended Self,” 486.

⁷⁵ Olson, “Extended Self,” 486.

⁷⁶ Olson, “Extended Self,” 487.

arises: which of these thinkers am *I*? Am I the person or the organism? I cannot be both since they are numerically distinct, and one thing (I) cannot be identical to two things (the organism and the person). Introducing *extended beliefs* further complicates matters, because, as per mental state internalism, an organism cannot have mental states beyond its boundaries. This means that Otto can have a belief concerning the location of the Art Exhibit because Otto's notebook *does* lie within his boundaries *qua* extended psychological being. However, Otto *qua* organism cannot have such a belief because notebooks never become parts of organisms. Further, Otto *qua* organism cannot have *any* mental beliefs given the conjunction of mental state internalism and the mere possibility of extended mental states since, due to *parity principle* considerations, *any* mental state *could* be realized beyond organismal boundaries. It follows that no organism ever has any mental state. For any psychological being, there is the property *possibly having a notebook as a part* which applies equally to all such beings and to all possible mental states. Since organisms never have such a property, organisms cannot have mental states because they could in principle be realized by something outside of their boundary.

What follows from the conclusion that no psychological being could be an organism? Firstly, the problem of *too many thinkers* disappears simply because no organism can be a thinker since no organism can have mental states, as concluded above. Our thoughts and other mental states need no longer share authorship between a person and an organism. However, in the place of the *too many thinkers* objection, another problem arises. This position commits us to *substance dualism* – a view which many find unattractive.⁷⁷ “This is no mere dualism of properties, distinguishing mental from

⁷⁷ Olson, “Extended Self,” 488.

physical or biological properties. Nor is it merely the claim that psychological properties or descriptions are somehow irreducible to physical ones. It is full-on substance dualism: a dualism of psychological and biological beings.”⁷⁸ *Dualism* is the view that there are two fundamental kinds or categories into which different properties, substances, or beings fall. Descartes uses this categorization scheme to divide the world into two kinds of stuff or substance: mind and body. The respective categories of mind and body capture properties exclusive and exhaustive of all possible kinds of substance. For example, whereas mind is indivisible, body is infinitely divisible; whereas body is extended in time and space, mind is non-extended. By antonymically defining these two fundamental categories, the world, in principle, falls neatly and completely into a dualism of two substances: something is either made of mental substance or of bodily substance.

However, Olson argues that it is not mind and body that the world is divided into, but mind and *life*. It is impossible for a thinking thing to be a living thing on pain of violating the principle of mental state internalism. Therefore, there must be two fundamental kinds of things, each with their own set of exclusive and exhaustive properties. If the *extended mind thesis* necessitates the conclusion that the world must be divided along dualistic lines of living things and psychological things, it is far from a feather in the cap for friends of the *extended mind*. The reason why is that dualism is a thoroughly problematic and contentious position in its own right. Specifically, dualism fittingly faces two kinds of questions. The first kind are the ontological questions: What exactly are the two kinds of substances? What are their properties? Is one a subclass of the other? Are they completely distinct? The second kind are the causal questions: How

⁷⁸ Olson, “Extended Self,” 488.

do different substances affect one another, if at all? Is the causal relation between them reciprocal or one-way? What is the mechanism by which inter-substance interaction is accomplished? By being forced to adopt a kind of dualism as a consequence of their view, *extended mind* sympathizers add a bushel of knotty questions over and above “Does the mind extend beyond the head?”

Yet, are Clark and Chalmers forced to accept dualism at this juncture? Olson anticipates an objection to the premise that *an organism can never have a notebook as a part*. There is something of a complementary view to the *extended mind thesis* in biology called the *extended phenotype thesis*. It is, simply put, the proposal that extra-organismal structures and products which contribute to the reproductive success of that organism are literally part of that organism. Examples include the extended organism of the spider and its web or the beaver and its dam.⁷⁹ Here the purported extended organism is Otto and his notebook. However, this is not to say that there is still something which does *not* get bigger by including the notebook as a part. Call this the *core organism*. Otto may very well become an extended organism by virtue of using the notebook, but the entity definitionally demarcated by Otto’s biological boundaries, i.e., the core organism, never gets bigger by using a notebook. This means that the initial problem stands: there is a dualism between an entity which can have mental states, Otto *qua* extended organism, and an entity which cannot, Otto *qua* core organism. Olson argues that this appeal to the *extended phenotype* is a mere label-swap; it is still the same dualism of psychology and life under different names.⁸⁰

⁷⁹ Schulz, “Overextension,” 245.

⁸⁰ Olson, “Extended Self,” 489.

3.4: The Scattered Self

Shifting focus, how does our conception of the self change under the substance dualism that Olson argues we must accept if we take the *extended mind thesis* on board? One thing is that we would have to accept that we are much smaller than we think we actually are. Remember that, according to mental state internalism, a psychological being is only where its mental states are. Since we have no mental states in our hands, feet, or teeth, none of these things are actually a part of us. We seem to be a “neural net”: “I am a highly exiguous net of tissue scattered across the parts of my brain that light up during CAT scans when I am mentally active.”⁸¹ Otto is composed of such a thing and a notebook.”⁸² Strictly speaking, we psychological beings, being constituted by mental states, are scattered objects existing partially wherever one of our mental states is partly realized. Candidates for mental realization are first and foremost the brain and nervous system, but by *parity principle* considerations, we are able to get bigger or smaller depending on which functionally isomorphic structures and processes we recruit into our repository of mental vehicles, be they notebooks, computers, or even other psychological beings. The most disconcerting thing about being constituted purely by things that realize mental states is that those things which we thought were part of us, namely most of our organismal body, are in fact not part of us, but part of our environment. “That would make us rather peculiar material things, if indeed we are material at all. It would also leave us wondering why a psychological being must be located only where its mental states are. Why should thinking be incompatible with having feet as parts?”⁸³

⁸¹ Olson, “Extended Self,” 492.

⁸² Olson, “Extended Self,” 493.

⁸³ Olson, “Extended Self,” 494.

There are two ways one can go about answering “Why should thinking be incompatible with having feet as parts?” One can deny mental state internalism, thereby allowing organisms, feet and all, to have mental states despite their being beyond their organismic boundary. Or you can determine that psychological beings are not *material* beings, but a sort of *bundle* of psychological states. Take the first option of denying mental state internalism and you are left with the familiar problem of *too many thinkers*. Without boundary considerations to qualitatively distinguish persons from organisms, every mental state would belong to both a person and an organism. To take the second option is to embrace substance dualism wholesale: psychological beings are immaterial bundles of mental states, separate in substantial *kind* from the organism on which we happen to supervene. This would save the *extended mind thesis* from being committed to the “neural net” model of the self, scattered across the brain, body, and world where, and only where, the material that constitutes a mental state resides; a bundled self is not scattered in space because it does not exist in space at all: it is *immaterial*.

Olson notes how Clark and Chalmers’ commitment to psychological continuity and mental state internalism evades the scattered object version of the self, but the bundled self does leave us with two mysteries: (a) Why is it metaphysically impossible for an organism with a working brain to think? (b) How is it that an immaterial bundle of mental states with no brain *can* think? Remember that we cannot say that organisms, in addition to persons, are thinkers without running into the *too many thinkers* objection. Yet the alternative, that organisms with complex nervous systems and highly developed brains are incapable of thought, is highly unintuitive and unsatisfying. Further, to say it is not the organism who authors the mental states, but the bundle of mental states itself, appears to be a confusion of act and actor: “Saying that what thinks is not the organism of which my

thoughts are states but rather the bundle of those thoughts itself is like saying that what moves across the stage is not the dancer but rather her movements.”⁸⁴ In conclusion, then, *extended mind* theorists face a disjunctive choice: either they accept the inference from the extended mind to the extended self, as argued initially by Clark and Chalmers, and commit themselves to a bundle view of the self along with its metaphysical mysteries concerning what follows from a dualism of life and mind; or they simply reject the inference to an extended self from the extended mind: “Otto’s memories may or may not extend into the notebook he keeps in a drawer. But either way, Otto himself is right here.”⁸⁵

⁸⁴ Olson, “Extended Self,” 494.

⁸⁵ Olson, “Extended Self,” 495.

Concluding Thoughts

Surely mental state inclusivity has been pushed too far when we start talking about notebooks and hippocampi as if they were the same kind of thing! As we have seen, Adams and Aizawa express such reservations about the Otto example and other cases of what they see as over-inclusivity generated by too loose a definition of what defines a mental state.⁸⁶ They argue that the human mind is unique in two different respects: one, the mind has the unique capacity to generate *nonderived content*; two, there are specific causal laws that describe the workings of the mind that do not apply to things like amnesiacs using notebooks. Adams and Aizawa conclude that since the brain is the only thing that can meet both above mentioned criteria, the mind contingently remains in the head (because that's where the brain is). On top of their account of what the necessary "mark of the cognitive" is, Adams and Aizawa generate a critique of *extended mind* arguments, the so-called *causal-constitution fallacy*. Adams and Aizawa charge *extended mind* defenders with equating causal contribution of external objects and processes with constitutive or "parthood" status in the material make-up of the mind. Adams and Aizawa make a multifarious case against vehicle externalism but, as I have shown, their critiques have been adequately deflected by defenders of the *extended mind thesis*.

As a bridge between functionalism and the *extended mind thesis*, I used the Martian Intuition, which was just the thought that differing physiological profiles is not sufficient to deny mentality to a creature. Martians can want to go to the New York Museum of Modern Art as well!

⁸⁶ Adams and Aizawa, "Bounds of Cognition," 70.

However, if we extend the olive branch of mental state inclusivity to *Martians*, then what motivation do we have to balk at Otto's having the belief about the art museum just because his mental state profile has a notebook as a part? That's the question Sprevak asks and I think it is a good one.⁸⁷ For any case of an extended mind, there is a possible Martian whose functional profile is functionally identical with it. Since there are adequate motivations to keep functional state definitions *coarse-grained*, the plausibility of the *extended mind theory* is increased enormously. But surely there must be *some* boundary to the mind, some set of limiting conditions that determine what is and what is not a mental vehicle. Clark and Chalmers come up with some *prima facie* good ones, known as the "trust and glue" conditions. The external resource must have three attributes: (a) it must be readily available when needed, (b) it must have been previously endorsed and not subject to scrutiny or doubt, (c) it must be easily accessible and the information contained therein retrievable. Viewing the Otto example in light of these conditions we saw that the notebook satisfies all three for Otto and therefore ought to be considered an extension of his mind.

Unfortunately for Clark and Chalmers (and Otto), these criteria could not stand up to scrutiny without appearing to be completely *ad hoc*. On the one hand, counter-example scenarios exist which show that even if all three criteria are satisfied, the *extended mind thesis* is still extremely unintuitive. What happens when Otto, or anyone, uses a telephone?⁸⁸ All phones come with a directory service, usually comprised of an index of phone numbers and a director assistance operator if need be. Both of these resources are

⁸⁷ Sprevak, "Extended Cognition," 515.

⁸⁸ Robert D. Rupert, "Challenges to the Hypothesis of Extended Cognition," *Journal of Philosophy* 101 (8) (2004): 402–403.

(a) readily available when needed, (b) have been previously endorsed and not subject to scrutiny or doubt, (c) easily accessible and the information contained therein retrievable. Satisfaction of all three “trust and glue” criteria means that Otto has beliefs about the phone numbers of innumerable people just by having a phone as a readily available, trusted, and accessible resources in his daily life. And in our age, where nearly everyone has a phone, we all have much more dispositional knowledge of phone numbers and people than we ever dreamed of (or wanted). Another way in which the “trust and glue” conditions are not necessary conditions for the possibility of extended mental states is the empirical fact that many cases of *internal* mental states do not meet them. It is simply not true that every mental state you have is readily available – consider repressed memories or cases of unconscious biases, for example. It is not true that every internal mental state is immune to scrutiny and doubt, one can easily doubt if your memories are veridical, or wonder if the pain you are experiencing is real or psychosomatic. It is also not true that all internal mental states can be easily accessed, most of our beliefs are below the threshold of consciousness and sometimes, in order to remember something, one must use mnemonics or some other associative process to get the relevant information for the relevant belief. So why do internal mental states get a free pass while external ones have to jump through “trust and glue” hoops? Wasn’t the internal/external boundary supposed to be deemed irrelevant through functional *parity*? There is significant tension between Clark and Chalmers’ “trust and glue” criteria and functional parity. There just is not equal opportunity between inner and outer vehicles if one must meet all of these additional criteria while the other kind need not.

So what did the failure of the “trust and glue” conditions mean? It meant that if being in the head is no longer a necessary condition (or even a sufficient condition!) for

some putative part (vehicle) of a mental state, then what is? It is the lack of a principled answer to this question that leads to what is known as *cognitive bloat*. The problem, as we have seen, is that without a bare necessary condition, such as *being in the head* or *being made out of brainstuff*, the boundaries and proper parts of the mind are effectively lost, leading to an inclusion of objects and processes that seem erroneously to be part of one's mind or cognitive process based on the merest causal contribution. Inclusion is *too* easy, and coarse-grained functional parity is *too* vague to give a principled demarcation of the mind. Appeals to Proper Functions and appropriate etiological history of putative parts have been demonstrated to be irrelevant through *swampman* style arguments. Finally, appeals to *coupled systems* based on reciprocal causation were dismissed as an answer to another question entirely, for they speak of some abstract "cognitive system" which I argue is not identical to the mind of a *person*, let alone persons *themselves*. In fact, the system contains the mind of an individual cognizer as a subsystem, so it is not the case that the mind is extended; rather there exists some system such that it contains a cognizer and the relevant environmental objects and processes engaged in a relation of two-way causation. The actions of this coupled system may explain certain cognitive processes more accurately and completely, such as using a pen and paper to do a math problem, but it is far afield from the question of whether a person's mind can have parts of the world as vehicles of their mental states.

Even if there were some principled way to demarcate the mind from the environment, the bare fact that external objects and processes can be mental vehicles would wreak havoc on our notions of selfhood and identity. I am located (not entirely, of course) wherever a part of me is; parthood is a transitive relation; vehicle externalism implies that objects outside of my head could be parts of my mind, and my mind is a part

of me. Thus, these external vehicles are a part of me. Imagine pressing charges against someone for intent to commit bodily assault just because they threatened to doodle in your philosophy notebook which just happened to be the external vehicle responsible for your belief that *All men are mortal*. I do not think I am exaggerating with this example because *extended mind* theorists are being literal when they say parts of the world can be parts of your mind. But remember *mental state internalism* which says that I am wherever my mental states (and their parts) are. Further, it is an implied corollary of psychological-continuity theories of personal identity, a very popular stance in the philosophy of personal identity. If both mental state internalism and *extended mind theory* are true, this would imply that I am where my philosophy notebook is, but not where my feet are, because my notebook is a vehicle of one of my mental states but my feet (not usually, at least) are not a vehicle for any of my mental states. Am I really to believe that a book I bought at the dollar store is more essential to my identity than the feet that I was born with, to which I have been attached my whole life?

The mere possibility of external vehicles makes the line between mind and environment hopelessly vague. Just as asking “What is the exact hair one must lose in order to be bald?”, asking “What are the boundaries of my mind?” raises some very perplexing metaphysical questions. If we are functionalists we define the mind as a collection of mental states each one of which is an instance of a particular type of functional state. If we believe in the possibility of mental Martians with a significantly differing physiology, then we are committed to coarse-grained functional states. If we define our functional state types coarsely, then our conception of what it is to have a belief is liberal enough to allow a notebook to play the appropriate causal role in the functional state belief. If beliefs and other mental states can be constituted by the causal

roles played by external objects, then the mind includes external objects as parts. If one endorses a psychological theory of identity, and, therefore, one is one's mind, then there is no boundary between oneself and one's environment. But there is a clear sense in which I am separate from my environment. Therefore, something has to give. One possible angle is to be a reductionist about the whole business about minds, and persons, and folk-psychological mental states. There are only cognitive systems, units of information processing, some of which contain human organisms and notebooks as parts. This has a name, Hypothesis of Extended Cognition (HEC), and really does not sound so strange. How are external objects and processes causally contributory, and in that sense, explanatorily constitutive, to human problem solving and human information processing is an interesting empirical question and there are corollary debates in other fields of science, such as the unit of selection debate in biology. It's not a terribly interesting philosophical question, however. The interesting philosophical question here is, "What happens to our notions of mind and person when we lose the gooey brain center?" To go whole hog on the HEC train of thought is to answer that question with "Who cares? Minds and persons are not explanatorily relevant to human cognition." At any rate, I will leave the unit of cognitive analysis debate for cognitive scientists to worry about. What I want to worry about is the locational conundrums faced by an extended view of the mind and the self.

Like most people, I want to say things like "I am right here" and not "I am partially realized here and partially in a notebook I left on my desk at home." I do not want to have to take my friend literally when she says, "Get your mind out of the gutter," because I accidentally dropped my phone in the street. And yet I find functionalist arguments for the possibility of external mental vehicles plausible, for reasons discussed

at length. Can I bring coherence to the common-sense conception of the self as individual-bounded and functionalist sympathies for Martians and men with notebooks containing beliefs about the locations of art museums? Short of abandoning either the individual self or the extended self, one might unhelpfully point out that maybe there is some “magic membrane” separating the *extended individual* from her environment. There are several candidates for what could count as a demarcating principle, but all of them come with costs. One could limit those mental states constitutive of a mind, and transitively, of a person, to only those that are currently held in consciousness. Otto does not know the Art Museum is on 53rd Street unless he is consciously reading the page of the notebook upon which it is written. Inga does not know it either, unless she is consciously thinking about it. The obvious downside to limiting minds to conscious states is that people can only entertain a few states at a time. This means that we are all incredibly ignorant. Nobody knows that the Earth is round unless they are consciously thinking about it. But while they are thinking that the Earth is round, they do not know that water is wet. Either the public education system has failed horribly or limiting minds and persons to conscious-attributable states is far too strict. It is just too extreme to discredit a person of all dispositional knowledge. Of *course* I know that the Earth is round and that water is wet even when I’m not consciously entertaining either proposition.

Since the problem of reconciling extended minds with the common-sense belief that “you are here” is a result of barrier vagueness between the environment and the mind/self, a possible means of compatibilizing these two concepts is a retreat into willful ignorance about the exact boundary between the mind and the world. One must be an

epistemicist about the extension of the mind.⁸⁹ One can make beliefs about mental extension and individual selves/minds coherent only at the cost of not knowing the exact boundary between the mind and the self via an argument such as this:

P1: The mind exists.

P2: It can include extraneural vehicles.

P3: It is distinct from its environment.

C: There must be a boundary between those vehicles that constitute a mind and the environment at large.

The question is “Will we ever figure out where that boundary lies?” A true epistemicist would cling to its unknowability, averring that it is a question epistemologically equivalent to asking at what exact grain of sand did a scattering of sand become a pile? Is the mental boundary quandary a result of the same vagueness woes that generate the sand problem? I do not think so. The sand problem is generated by the arbitrariness of pointing out a single grain of sand as *the* grain that somehow made the difference between a scattering and a pile. If grain #322 made the difference, then are we really to believe 321 grains of sand is not a pile? It is just inconceivable that a single grain can make the difference. However, the vagueness of mind/environment demarcation is not generated by arbitrariness; it is generated by a metaphysical barrier between mind and environment that does not translate neatly onto any natural physical barrier. The *extended mind* protagonist *knows* that the mind is separate from the environment, but also believes that it isn’t picked

⁸⁹ Epistemicism is an attitude towards vagueness that allows one to reconcile the existence of the “magic” grain of sand that made a pile, but only at the cost of professing that the exact grain is unknowable. I want to suggest that maybe there is a parallel option available to defenders of *extended mind*, namely that one can be committed to there being a boundary between mind and world, but not being able to know it.

out by the neat natural boundary of the skull. What's to be done? One of two things: one can either throw up their hands and just say that there is *some* physical boundary which maps onto the metaphysical boundary between mind and world; we either just do not know it *yet* or cannot know it *in principle*. The second thing that can be done is to retreat into reductionist thinking and endorse a unit of cognitive analysis that contains causally or explanatorily relevant members as all there is to say on the topic of extended minds.

I think that both options are lackluster. An epistemicist position is unnecessary because professions of such profound ignorance on the matter of mind location is intellectual laziness. It is not clear that one can take the easy way out, like the epistemicist who says, "Beats me" when asked which grain of sand made the difference between a scattering and a pile. Is knowledge impossible as to the physical realization of the metaphysical boundary between the mind and world? Or is it only that we have not come by it yet? Is there even such a boundary? It is certain that we occupy a position of epistemological uncertainty, for we do not even know whether there is such a boundary, let alone where it might lie. Of course, boundary demarcation is easier if we substitute *cognitive system* for *mind*. But, as I have argued, a reductionist tack on the subject of extended minds is anemic in the sense that all our notions of self, identity, and knowledge get lost in the language and ontology of "dynamical systems" and "continuous mutual interaction."

So what is an *extended mind* theorist to do? I do not think that they can have their cake and eat it too. The only options as I see it are either to downgrade their metaphysically daring theory to a humble empirical theory about the explanatory contribution of bodily and environmental props in describing actual human cognition accurately (*hypothesis of extended cognition*) or keep the philosophical gusto their

position entails only at the expense of never knowing the true word-count of their page-bound minds. It would be an interesting world if extended minds were actual, but the tension between “I” as a more or less clearly bounded individual spatially centered on my body and the entailment of *extended selfhood* from the *extended mind* means it is time to decide what to believe. In making a coherent structure of belief, it is not a bad idea to be conservative and aim to adopt new beliefs only when they do not threaten to overturn beliefs you already endorse. Unfortunately, adopting the postulates of the *extended mind thesis* is extremely disruptive to many common-sense notions regarding personal identity, selfhood, spatial location, and knowledge attribution. Consider, as well, the moral and legal implications an *extended mind* theorist would have to face: what counts as assault if external objects are a part of me? Who owns the beliefs about the contents of a popular textbook? Is it a medical emergency, akin to suffering a concussion, if I drop my phone and it loses some functionality? And so on. I suggest we all “remember what the Dormouse said: *feed your head*” – because that is where the mind is.

Bibliography

- Adams, Fred and Aizawa, Ken. "Why the Mind is Still in the Head." In *The Cambridge Handbook of Situated Cognition*, eds. Murat Aydede and P. Robbins, 78–95. Cambridge: Cambridge University Press, 2009.
- . "Defending the Bounds of Cognition." In *The Extended Mind*, ed. Richard Menary, 67–80. Cambridge: MIT Press, 2010.
- Allen-Hermanson, Sean. "Superdupersizing the Mind: Extended Cognition and the Persistence of Cognitive Bloat." *Philosophical Studies* 164 (3) (2013): 791–806.
- Armstrong, David. "The Causal Theory of Mind." In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, 80–87. Oxford: Oxford University Press, 2002.
- Baker, Lynne Rudder. "Persons and the Extended-Mind Thesis." *Zygon* 44 (3) (2007): 642–658.
- Block, Ned. "Troubles with Functionalism." In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, 94–98. Oxford: Oxford University Press, 2002.
- Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.
- Clark, Andy. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press, 2011.
- . "Coupling, Constitution and the Cognitive Kind: A reply to Adams and Aizawa." *The Extended Mind*, ed. Richard Menary, 81–100. Cambridge: MIT Press, 2010.
- . "Memento's Revenge: The Extended Mind, Extended." In *The Extended Mind*, ed. Richard Menary, 43–66. Cambridge: MIT Press, 2010.
- Clark, Andy and Chalmers, David J. "The Extended Mind." In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, 643–651. Oxford: Oxford University Press, 2002.

- Dennett, Daniel C. "Where am I?" In *Brainstorms*, 333–347. Cambridge: MIT Press, 1981.
- Drayson, Zoe. "Extended Cognition and the Metaphysics of Mind." *Cognitive Systems Research* 11 (4) (2010): 367–377.
- Farkas, Katalin. "Two Versions of the Extended Mind Thesis." *Philosophia* 40 (3) (2012): 435–447.
- Hurley, Susan L. "Varieties of Externalism." In *The Extended Mind*, ed. Richard Menary, 101–154. Cambridge: MIT Press, 2010.
- Kaplan, David Michael. "How to Demarcate the Boundaries of Cognition." *Biology and Philosophy* 27 (4) (2012): 545–570.
- Millikan, Ruth G. *Language, Thought and Other Biological Categories*. Cambridge: MIT Press, 1984.
- Palermos, Orestis S. "Loops, Constitution and Cognitive Extension." *Cognitive Systems Research* 27 (2014): 25–41.
- Pereboom, Derek. (1991) "Why a Scientific Realist Cannot be a Functionalist." *Synthese* 88 (1991): 341–358.
- Putnam, Hilary. "The Meaning of 'Meaning'." In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, 581–596. Oxford: Oxford University Press, 2002.
- . "The Nature of Mental States." In *Art, Mind, and Religion*, eds. W.H. Capitan & D.D. Merrill, 37–48. Pittsburgh: Pittsburgh University Press, 1973.
- Olson, Eric T. "The Extended Self." *Minds and Machines* 21 (4) (2011): 481–495.
- . "Personal Identity." *The Stanford Encyclopedia of Philosophy* (2017), ed. Edward N. Zalta. <https://plato.stanford.edu/archives/sum2017/entries/identity-personal/>.
- Rupert, Robert D. "Challenges to the Hypothesis of Extended Cognition." *Journal of Philosophy* 101 (8) (2004): 389–428.

Schulz, Armin W. "Overextension: The Extended Mind and Arguments from Evolutionary Biology." *European Journal for Philosophy of Science* 3 (2) (2013): 241–255.

Searle, John. "Can Computers Think?" In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, 669–675. Oxford: Oxford University Press, 2002.

Shoemaker, Sydney. *Identity, Cause and Mind*. Cambridge: Cambridge University Press, 1984.

Sprevak, Mark. "Extended Cognition and Functionalism," *Journal of Philosophy* 106 (9) (2009): 503–527.

Wheeler, Michael. "In Defence of Extended Functionalism." In *The Extended Mind*, ed. Richard Menary, 245–270. Cambridge: MIT Press, 2010.

Wilson, Robert A. and Clark, Andy. "How to Situate Cognition: Letting Nature Take its Course." In *The Cambridge Handbook of Situated Cognition*, eds. Murat Aydede and P. Robbins, 55–77. Cambridge: Cambridge University Press, 2009.

Wilson, Robert A. and Lenart, Bartłomiej. "Extended Mind and Identity." In *Handbook of Neuroethics*, eds. Jens Clausen and Neil Levy, 423–439. Springer, 2014.