

The Extended Mind and its Theoretical Vices

Lane DesAutels

Abstract: It is not enough to show that the extended mind account accommodates the observable evidence just as well as its internalist rival. The widely accepted underdetermination of theories by their evidence shows equivalence in empirical adequacy to be trivial (and uninteresting). To show that the extended mind account is the *better* theory, it must be shown that it enjoys support from other theoretical virtues. In what follows, I compare the extended mind account to its internalist rival using several of the traditional theoretical virtues as comparative dimensions. Appealing to predictive fruitfulness, parsimony, unification, and explanatory power, I aim to show that there is no available analysis according to which the extended mind account comes out on top of the internalist orthodoxy on these measures. As such, I conclude that (on balance) the extended mind account is more theoretically vicious than it is virtuous—and should, therefore, be rejected.

1. Introduction

When dividing one large number by another, many reach for paper and pencil, or better yet, a calculator. Gamers have an easier time determining where Tetris shapes best fit by rotating them on a computer screen before them, rather than in their heads. A notebook detailing the locations of frequently visited places, for an Alzheimer's patient, can prompt behavior much the way a memory does for a healthy person.

On one recent and influential account in the philosophy of mind, examples like those above are instances of something quite striking. These are times when the mind *extends* into the world. Things in the world like notebooks, pencils and paper, computer screens, and calculators, when coupled in the correct way with cognitive agents, quite literally become *part of* their minds. Those holding such a view accept what has been called the *extended mind thesis*: the mind and the cognitive processes that constitute it extend beyond the boundary of the skin. Clark and Chalmers, in their 1998 *Analysis* paper *The Extended Mind*, write "...we will argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes. If so, the mind extends into the world" (Clark & Chalmers, 1998, p. 33). Those holding the contrasting position, which I call the *internalist orthodoxy*, maintain that the mind and the cognitive processes that constitute it *do not* and *cannot* extend beyond the boundary of the skin. The extended mind thesis, in various forms, has been taken up by several prominent philosophers of mind (Logan 2007, Clark 2008, Hurley 2010,

Wilson 2010, Sutton 2010, Menary 2010, Wheeler 2010), and heartily criticized by others (Adams and Aizawa 2008, 2010, Rupert 2009, 2010, Preston 2010).

In this paper, I undertake a novel re-examination of the extended mind debate: one that brings important considerations from the philosophy of science to bear. I propose, first, that it is not enough to claim that the extended mind account captures the relevant observable phenomena just as well as an internalist account of the mind does. The widely accepted underdetermination of theories by their evidence shows equivalence in empirical adequacy to be trivial (and uninteresting). In order to motivate the acceptance of the extended mind account over the internalist orthodoxy, rather, it must be shown that it is a *better* theory than its internalist rival. The only way to do this is to show that the extended mind account enjoys more support from the theoretical virtues, virtues like: predictive fruitfulness, parsimony, unification, and explanatory power.

I argue, in what follows, that no such case can be made. In other words, I suggest that there is no available analysis according to which the extended mind comes out on top with respect to the traditional theoretical virtues listed above. In short I aim to show that, on balance, the extended mind account is more theoretically vicious than it is virtuous.

Here is the plan: in Section 2, I show that the extended mind account and its internalist rival are both capable of accommodating the observable phenomena equally well (but this fact is neither interesting nor useful). In Section 3, I recount an instance where the extended mind account seems to have generated novel, true predictions, but I suggest that a sophisticated internalism about the mind would have been equally predictively fruitful. In Section 4, I distinguish between three kinds of parsimony: explanatory, ontological, and process. I then argue that only the latter two should matter for theory comparison, and on these measures, the internalist account wins the day. In Section 5, I contend that no sense can be made of the claim that the extended mind account better achieves explanatory unification—at least not in the sense that philosophers of science care about unification; if anything, considerations from evolutionary biology show that it is less unified with our best science than internalism is. In Section 6, I distinguish between 1st and 2nd wave extended mind theories and suggest that the 1st wave extended mind approach, by virtue of its commitment to coarse-grained functionalism, has less explanatory power than its internalist rival. By virtue of its rejection of functional comparisons, on the other hand, the 2nd wave approach sacrifices the intuitive pull motivating the extended

mind project to begin with. I conclude, in Section 7, that (on balance) the extended mind account is more theoretically vicious than it is virtuous, and should thereby be rejected.

2. Empirical Adequacy and Underdetermination

A theory is empirically adequate just in case it accounts for—in the sense that it explains, captures, or predicts—the relevant observable phenomenon. Darwin’s theory of evolution by natural selection accommodates the multifarious evidence for the mutability of species.

Traditional creationist accounts, by virtue of their commitment to the designed, immutable nature of species, cannot accommodate this evidence; Darwin’s theory is empirically adequate; creationism is not.

What are the observable phenomena that theorists are trying to make sense of in the extended mind debate? What counts as the evidence to be accounted for by the rival theories? There appear to be a few different sorts. First, there are the types of considerations mentioned at the outset of this paper: people tend to rely on pencil and paper when dividing large numbers; and people are better at playing Tetris when they rotate the shapes on a computer screen in front of them rather than in their head. Call these cases of *cognitive offloading*: instances where people tend to do better at cognitive tasks when they can employ objects in their surroundings. There is also Clark and Chalmers’s imagined case of Otto: an Alzheimer’s patient who logs all his “memories” and “beliefs” in a notebook (Clark & Chalmers, 1998, p. 33). When Otto needs to get to the Museum of Modern Art (MoMA), he automatically pulls out this notebook, finds that he recorded the location of MoMA, and walks to 53rd street. The sentences in Otto’s notebook are “functionally poised” to produce just the same sort of behavior that memories and beliefs do for healthy modern art enthusiasts. There also seem to be analogous cases from dynamic systems studies. A fish has “an evolved capacity to couple its swimming behaviors to the pools of external kinetic energy found as swirls, eddies, and vortices in its water environment” (Clark & Chalmers, 1998, p. 32). Just like long-dividers, Tetris players, and Otto, reference must be made to aspects of the fish’s environment in order to fully explain its capacity to swim.

With regard to the issue of empirical adequacy, the question becomes the following: are the forgoing observable phenomena *only* able to be accommodated by the extended mind advocate? Put another way: is the extended mind framework *uniquely* able to capture, explain, or predict these sorts of phenomena?

Before answering this question, it is necessary to further elucidate just what sort of position is available to the internalist. Recall that internalism about the mind is simply the view that the mind and the cognitive processes that constitute it do not (and cannot) extend beyond the skin. However, *and this is key*, the internalist need not ignore the complex, causal reciprocity that environmental factors play on the mind and the cognitive processes that constitute it. And in paying careful attention to these complex causal relationships, the internalist is still perfectly capable of maintaining that these environmental entities are *separate* from the minds of those interacting with them. On the basis of these considerations, I characterize the kind of internalism sketched here as *internalism (+)*. The added '(+)' is meant to denote the special emphasis a sophisticated internalist is free to place on the complex, dynamic, and reciprocal interaction the mind can have with its (external) surroundings while still remaining an internalist.

So now returning to the question of whether the extended mind is more empirically adequate than its internalist rival. I contend that, if internalism (+) is adopted, then it is clear that both positions are equally empirically adequate. The extended mind theorist explains people's increased abilities to divide large numbers with paper and pencil as the result of their minds extending over the paper and pencil. The internalist, on the other hand, equally well explains this phenomenon by appealing to the causal-reciprocal relationship that the mind has with the paper and pencil—environmental entities that nevertheless remain external to the mind. Both theories account for the phenomenon in question. In the Tetris case, the story goes the same. For the extended mind theorist, the computer screen becomes part of the mind, but for the internalist it remains an external entity that merely has a certain causal relationship to the mind. How about Otto? The extended mind advocate accounts for Otto's ability to get to the MoMA by counting the sentence in his notebook as one of his beliefs. The internalist, on the other hand, explains that Otto merely reads the sentence in the notebook (an external, non-mental artifact) which then prompts an internal, mental representation (a belief) which he uses to get to the MoMA. A similar account can be given of the analogy to the dynamic nature of fish swimming. A fish's capacity to swim may depend on vortices in the surrounding water, but referencing these factors amounts to adding an explanation of the external, causally relevant factors playing a certain role in the fish's ability to swim.

The above analysis indicates that there seems to be nothing about these observable phenomena that the extended mind account can accommodate that the internalist (+), in practice, cannot. The two accounts simply accommodate them differently.

The defender of the extended mind might respond by saying something like this. Fine; the extended mind account does not capture any phenomena that the internalist cannot. But since they both accommodate the evidence equally well, aren't we equally justified in believing in either? My response to this is to point out that philosophers of science have long recognized the underdetermination of theories by their evidence (Duhem 1914, Quine 1975). Given any body of observed phenomena, there are an *infinite* number of theories that can accommodate it. The simple adding of irrelevant disjuncts or multiple negations to any going account shows this to be trivially true. Underdetermination shows that the process of theory selection in science would never get off the ground if we had to equally believe every theory that accommodates the evidence. That would mean equally believing in an infinite number of theories; and this is epistemically untenable. As such, there just is not anything interesting or useful about claiming that both the extended mind and internalist accounts equally accommodate the evidence. To motivate the acceptance of one empirically adequate theory over another, appeals must be made to other theoretical virtues.

3. Predictive Fruitfulness and New Interdisciplinary Perspective

Consider the following historical puzzle. In Shakespearean England, the repertory theatre system demanded its exclusively male actors to perform, what Evelyn Tribble describes as, “a staggering number of plays... with relatively infrequent repetition, and with the additional demands of putting on a new play roughly every fortnight” (Tribble, 2005, pp. 135-136). According to Tribble, it was not uncommon for leading actors to memorize and retain command of over seventy different roles. Demands on modern day thespians are nowhere near this level. How were these Shakespearean actors able to accomplish such a cognitive feat? The answer, Tribble discovered, is that actors used a diverse array of artifacts such as offstage scripts called “sides”, as well as some large documents called “plots” which were most likely hung on walls out of view of the audience. These offstage artifacts served, the story goes, to prompt the actors' spoken lines in a way that made the mastery of such a staggering number of roles possible.

Why does this historical puzzle matter? It becomes relevant for us because Tribble came across the solution to it *from* the extended mind perspective. Inspired by Hutchins (1995), Tribble sees these offstage cues as constitutive of the actors' extended cognitive systems: their minds. Just like long-dividers, Tetris players, and Otto, these environmental artifacts are functionally poised to play just the right sort of role to support the actor's cognitive task of remembering their lines. For an extended mind theorist, this qualifies them to count as part of the mind. Indeed it was precisely because Tribble was looking for such cognition-supporting artifacts, artifacts to be subsumed by these actor's minds, that she generated her (later confirmed) predictions about the existence of offstage cues. If a theory generates novel, true predictions—predictions not generated by its rivals—then this should count as a theoretical virtue of the theory. The extended mind advocate might, hereby, argue thusly: (1) if the extended mind account generates novel, confirmed predictions, whereas its internalist rival generates no such predictions, then we should favor the extended mind theory. (2) Tribble's research is an instance of the extended mind account generating novel, true predictions that internalism did not. (3) Therefore, we should favor the extended mind account on the dimension of predictive fruitfulness.

My response to the above argument for the predictive advantage of the extended mind is this. Predictive fruitfulness should only count in favor of one theory over another when it can be shown that either (A) the rival theory, in principle, *could not* have made the same novel, true predictions—or (B) if its rival theory *could* generate the same novel, true predictions, it could only do so in a dubiously *ad hoc* fashion. Einstein's general theory of relativity predicts that the path of a light beam will bend in the presence of a strong enough gravitational field. Newtonian physics' ignorance of general relativity makes it such that it *could not*, in principle, make such a true prediction. Thereby, in accordance with (A), we should favor the Einstein's general relativity over the classic Newtonian theory on the measure of predictive fruitfulness. With regard to (B), consider an example from developmental psychology. According to a *mindreading account* of the behavior of infants, infants are able to successfully demonstrate knowledge of another person's mental states (like beliefs and desires) because they can somehow sense them in the other person's mind. This account contrasts with *behavior-rule accounts* of infant behavior. On a behavior-rule account, infants know a set of behavior-rules like "people search for things where they were last in their line of sight." However, on such an account, no determinate set of

rules is specified. When proponents of the mind-reading account amass evidence for infant behavior that appears to fall outside of a specified rule, the behavior-rule theorists merely tacks on a new rule to accommodate the evidence and allow for making the same good predictions.ⁱ This is an example where a theory *can* make the same good predictions as its rival—but only in an *ad hoc* fashion. As such, despite its *ability* to make the same good predictions, the fact that the behavior-rule account can only do so in a problematically *ad hoc* fashion shows that the measure of predictive fruitfulness should still be seen to count in favor of the mindreading account.

So how does internalism (+) fare with respect to (A) and (B)? It strikes me as obviously true that an internalist (+) *can*, in principle, make the same predictions Tribble did. In fact, an internalist (+), by the very way I have defined it, *would have had every reason* to look to environmental artifacts in order to generate an explanation of the Shakespearian actors' remarkable feat of memorization. It just so happens that no sophisticated internalists took to the study of this particular puzzle: a fact that likely has more to do with sociological factors about the sorts of interests extended mind theorists tend to have. Given the internalist (+)'s ability to generate the same good predictions as Tribble's, the extended mind account cannot be seen to meet condition (A), and as such, the extended mind account cannot be seen to outcompete on the measure of predictive fruitfulness. What about (B)? Is the only way an internalist could have made these good predictions via some dubious *ad-hoc-ery*? My answer is no. There is no reason to think that the only way an internalist (+) would have come to the same predictions that Tribble did is in an *ad hoc* fashion. In fact, I think there is every reason to expect that the internalist (+), by virtue of her focus on the complex, causal role of environmental artifacts, would have come to these predictions straight away—just as Tribble did. Therefore, with respect to condition (B), the extended mind account still cannot be seen to outcompete on the dimension of predictive fruitfulness. Thus, predictive fruitfulness cannot be a decisive factor favoring the extended mind account.

Perhaps the extended mind advocate would grant me the above points, but nevertheless maintain that there is something valuable to the *different, inherently interdisciplinary perspective* an extended mind theorist takes. Such an interdisciplinary perspective, one might argue, is one that gives its adherents a broader scope with which to analyze puzzling phenomena. And this is what leads to better predictions. This is precisely what John Sutton seems to claim when he writes, "Tribble's careful analysis of cognition in the Globe [theatre] demonstrates the practical

value of the extended mind framework for offering new perspectives on cross-disciplinary topics of independent importance” (Sutton, 2010, p. 204).

Though this response seems promising at first glance, I think the internalist (+) can offer a (by now familiar sort of) response. Namely, there is nothing that precludes the internalist from *also* taking an equally interdisciplinary perspective. Indeed, internalists who study cognition already do take such an interdisciplinary perspective. Cognitive science is, and has been since its inception, an interdisciplinary enterprise combining psychology, neuroscience, robotics, computational modeling, artificial intelligence, philosophy, and dynamical systems studies. Whether or not external artifacts are conceptualized as part of the mind has no bearing on whether—and to what extent—other disciplines get considered when deriving predictions in the study of cognition. An internalist about the mind can, and should, consider these diverse disciplinary perspectives when analyzing puzzling phenomena. There is nothing, in particular, that makes the extended mind perspective the only interdisciplinary game in town.

On the basis of these considerations, I conclude (once again) that, if the extended mind approach is to be motivated as the better theory, it must be shown to enjoy support from some other theoretical measure.

4. Parsimony

William of Ockham told us that, all things equal, the simpler (more parsimonious) theory tends to be the right one. Ptolemy and Copernicus were each capable of accounting for how the stars appear to move across the night sky; but the Ptolemaic model could only do so by forming epicycles upon epicycles. The Copernican heliocentric model was much simpler. It was also right. Some defenders of the extended mind account cite parsimony as an advantage that their theory has over its internalist rival. In this section, I explore whether it is; and if it so, whether it matters.

First, an important distinction must be made. Parsimony can be used to evaluate rival theories in (at least) three disparate fashions: *explanatorily*, *ontologically*, or *viz. the number of distinct processes, principles, or laws* to which the theory appeals. Roughly, a theory is more explanatorily parsimonious if it captures the phenomena in question in fewer steps or using less language. A theory is more ontologically parsimonious if it postulates, or is committed to, the existence of fewer entities. And a theory is more parsimonious in the third sense if it appeals to

fewer distinct processes, principles, or laws as it captures a given set of phenomena.ⁱⁱ Clark and Chalmers seem to appeal to the first sort of parsimony in their original paper when discussing how best to explain Otto's ability to get to MoMA.

The alternative [to the extended mind explanation] is to explain Otto's action in terms of his occurrent desire to go to the museum, his standing belief that the Museum is on the location written in the notebook, and the accessible fact that the notebook says the Museum is on 53rd street; but this complicates the explanation unnecessarily... We submit that to explain things this way is to take *one step too many*. (Clark & Chalmers, 1998, p. 34)

Here, Clark and Chalmers suggest that an internalist explanation of Otto's behavior contains too many steps; it is less explanatorily parsimonious. The internalist, after all, has to refer to Otto's occurrent desire to get to MoMA, his standing belief that the location of the MoMA is written in the notebook, and the fact that is written in the notebook. The extended mind account need not cite Otto's standing beliefs about what the notebook contains to explain his behavior.

Referencing Otto's standing belief about what his notebook contains makes no more sense than citing a healthy modern art enthusiast's belief about what her memory contains in order to explain her walking to MoMA. Otto has the desire to go to MoMA, automatically reaches for his notebook to recall the address, and goes. This, Clark and Chalmers suggest, is a simpler explanation. Explanatory parsimony tells us to favor simpler explanations, so we should favor the extended mind account.

My response to the above argument is to shed doubt on whether explanatory parsimony ought to be relevant for theory selection in the first place. Specifically, I argue (I) any explanation can be made simpler by making it more abstract. And (II) if explanatory parsimony were to be a decisive measure, it would follow that more abstract explanations are always the better ones. However, this consequence is unacceptable.

Let us begin with (I). Suppose you witness a tree fall in a windy storm, and you want to explain why it happened. There are many levels of specificity you can undertake in order to do so. You can reference the precise wind speed and direction; you can talk about the tree's age and its root structure; you can describe variables in the soil density surrounding the tree's roots; you can mention vulnerabilities in the tree's trunk incurred from insect infestation; and on and on. On the other hand, you can cite the most abstract explanation available: the wind blew it over. This is just to show that, when it comes to explanations, simplicity and strength trade-off. The more

abstract an explanation, the less language it takes to account for the phenomenon in question. The more complicated an explanation is, on the other hand, the more information it contains: the more it tells us about the world. Put another way, increasing explanatory simplicity means decreasing explanatory informativeness, and vice versa. Note, however, that either explanation of the tree falling, whether hyper-specific or fully abstract, is true. But if we take explanatory parsimony seriously, we *have to* favor the explanation with the fewest steps; we *always* have to take the fully abstract approach. I suggest, in accordance with (II) above, that this result is unacceptable. The reason is as follows: explanation is *interest relative*. There are times when we want abstract explanations, and there are times when we want detailed ones. It depends on what we want our explanation to do. That an explanation contains fewer steps, alone, just is not a good reason to accept it. Otherwise we would always have to opt for the most abstract explanations. And those are not always the best ones for the purpose at hand.

Granted, the Copernican move to heliocentrism does seem to be a case where the explanatorily simpler theory was the right one. However, such cases in the history of science are few and far between. In general, scientists (especially life-scientists) seems to be discovering that the truth is often much more complex than we could have ever imagined. As such, many have moved to eliminate explanatory parsimony as a virtue altogether.ⁱⁱⁱ

With respect to selecting between competing theories, then, perhaps the relevant kind of parsimony is ontological. Is the extended mind account more ontologically parsimonious? Or is the internalist account? The extended mind theorist might argue that both accounts are committed to the same number of entities. What is at stake in the extended mind debate, the extended mind theorist might continue, is not how many entities exist, but how to categorize them. Both the extended mind theorist and the internalist are equally committed to the existence of human brains, computer screens, notebooks, papers and pencils, and the like. It is just that, if the conditions are right, the extended mind theorist wants to call some of these things *mental*. On this way of thinking, there is no clear advantage to either theory on the measure of ontological parsimony. Why? The same stuff exists for both accounts.

But this may not be right. Here is why. The extended mind account seems to postulate that the mind *subsumes* environmental artifacts when it gets coupled with them in a certain way. There is a sense in which the mind, on an extended view, quite literally forms a new composite object each time it enters into this tight coupling with a computer screen, a calculator, a pencil

and paper, or a notebook. In other words, on this kind of interpretation of the extended mind account, each time the mind couples in the right way with environmental artifacts, it forms a *different composite thing*. A mind coupled with a calculator, is one thing; a mind coupled with an abacus is another. Since there are, quite plausibly, a vast (if not infinite) number of distinct artifacts with which the mind can couple in this way, the extended mind theorist must be committed to the existence of a vast (if not infinite) number of minds in addition to the one inside your head. This strikes me as a significant ontological commitment. Since the internalist has no such commitment to a vast number of additional composite minds, it postulates far fewer existing entities. As such, it must be seen to be the more ontologically parsimonious theory.^{iv}

The extended mind advocate might respond to the above argument in the following sort of way. I, the extended mind theorist, am not claiming that *a novel kind of mind* comes to exist every time I come into the right kind of contact with a novel artifact. My claim is, rather, that a *novel kind of cognitive process* comes to exist when I am coupled the right way with an artifact in the world. And since a cognitive process is not a new kind of thing (the way a novel kind of mind is), the extended mind should not be seen to have a deleteriously bloated ontology. In response to this I agree that, on this kind of interpretation, the extended mind theorist need not be considered less ontologically parsimonious. However, this interpretation must be considered less parsimonious in the third sense mentioned in the outset of this section: viz. the number of distinct processes, principles, or laws to which a theory appeals. If my coming in the right kind of contact with a novel artifact results in a new and distinct cognitive process, and (once again) there is the possibility of a vast (if not infinite) number of such novel artifacts with which I could couple in this way, then the extended mind account is committed to the existence of a vast (if not infinite) number of distinct cognitive processes. The internalist, on the other hand, need not make any such commitment. Therefore, on the measure of process ontology, the internalist should be seen to win the day.

Perhaps the extended mind outcompetes on another measure.

5. Unification

“[I]t is part of the *job* of a special science to establish a framework in which superficially different phenomena can be brought under a unifying explanatory umbrella.” (Clark, 2010, p. 51)

“...despite the bottom-level physical diversity of the processes that write to, and read from, Otto’s notebook, and those that write to, and read from, Otto’s biological memory, there is a level of description of these systems that treats them in a single unified framework.” (Clark, 2010, p. 50)

"Ultimately, however, I think that issues about what really counts as a belief and about how the term ‘belief’ functions are terminological questions that while interesting, can mask the deeper point...*[E]xplanatory unification is the real underlying point of the extended mind thesis.*” (Clark, 2008, p. 5 [italics added])

Andy Clark, in the above quotations, makes a forceful appeal to another purported theoretical virtue of the extended mind account: explanatory unification. As he sees it, the extended mind theory achieves explanatory unification in a way that the internalist orthodoxy cannot; and since it is part of the job of science to seek unified theories, we should favor the extended mind account.

Just what is explanatory unification? And why is the extended mind supposed to achieve it more than the internalist orthodoxy? Beginning with the first question, some classic examples will help. Before Newton, an apple falling to the ground and the motion of the stars across the night sky were disparate, entirely irreconcilable, phenomena: phenomena that needed to be separately explained. Newton’s theory of gravitation and his laws of motion, however, brought these observable phenomena together under one set of laws. Both were instances of bodies operating under the same forces of gravity and inertia. In a similar vein, electric and magnetic phenomena, before Maxwell, were entirely disparate and irreconcilable. Maxwell’s famous equations, however, showed them to be part of one and the same phenomenon: electromagnetism. Even further, Maxwell’s theory of electromagnetism and Newton’s mechanics were brought together under one explanatory framework upon Einstein’s articulation of special relativity. These examples are all instances of what many take to be an inherently desirable goal of science: bringing together disparate phenomena, phenomena which previously required separate explanatory accounts, under one convergent explanatory umbrella.^v This allows for one explanatory framework when we used to need more than one. This is a good thing.

Now let us ask ourselves, in what sense does the extended mind account do this? In other words, what previously disparate, irreconcilable phenomena does the extended mind account allow us to bring together under one explanatory scheme? I contend there really aren’t any. By extending the mind to include surrounding environmental artifacts, there are no previously

disparate, irreconcilable phenomena that suddenly get unified under one account. On both the internalist and the extended mind account, there exists the realm of the mental and the non-mental. What is at stake in this debate is merely where the line gets drawn. Re-drawing this line to include environmental artifacts as ‘mental’ in no way amounts to the uniting of previously irreconcilable phenomena under one explanatory umbrella. There is nothing the extended mind theory unites that could not be united before.

What the extended mind account *does* do (and this must be what Clark means by ‘unification’) is to provide a novel framework according to which some of what was previously considered non-mental gets subsumed under the concept ‘mental’. And in this sense, previously non-mental entities are *unified* under the umbrella term ‘mental’. But this suggests that Clark must think there is something inherently valuable about subsuming entities under the umbrella of ‘mental’ that were not once thought of as such. But I cannot see how this could be. Subsuming entities previously envisaged as non-mental under the concept, ‘mental’, is no more inherently valuable than the pursuit of subsuming any set of entities under a different general kind than they were once considered to be. Sometimes we should do this, but only if we have an *independent reason*; only if it brings some *explanatory benefit*. We don’t just do it for the sake of doing it.

Perhaps, Clark might respond, the extended mind is more explanatorily unified in the sense that it brings new disciplines to bear on the study of cognition. It unifies disciplines of computational modeling, artificial intelligence, dynamic systems studies, etc, under the heading of cognitive science. But, as I argued in Section 3, cognitive science (even from an internalist perspective) *is already interdisciplinary*. So this does not help.

To go even further, I think we may have good reason to think that the extended mind account is actually *less* unified with our best science. Take the following reformulation of an argument given by Sterelny (2004):

1. Our epistemic tools operate in *contested space*.
2. Things in contested space are apt for sabotage and deception by other agents.
3. Evolutionary biology suggests that, given (1) and (2), a firm boundary between humans’ cognitive systems and their epistemic tools would have been evolutionarily advantageous—to prevent sabotage and deception
4. Extended mind suggests no such boundary exists.
5. Therefore, extended mind is not unified with evolutionary biology.

If we take Sterelny's point seriously, the extended mind account is *at odds* with one of our most evidentially supported sciences: evolutionary biology. Sterelny points out that an organism with a firm boundary between its cognitive system and contested space would plausibly be more suited for survival. Why?—because it possesses a barrier between its precious cognitive system and those who would benefit from its malfunction. With a mind that is bounded behind a skull, a place where would-be saboteurs cannot reach it, an organism can rest assured that his memories and beliefs stay (relatively) impervious to attack. An organism with no boundary between its mind and the environment, on the other hand, is vulnerable to interference and sabotage from its enemies. It is difficult to imagine an evolutionary story accounting for why humans would have evolved in such a way as to allow its memories and beliefs to be vulnerable to interference. So, we might conclude, the extended mind is less unified with evolutionary biology than the internalist orthodoxy.

It should be admitted that there may be, and probably are, reasonable responses to Sterelny's charge on behalf of the extended mind theorist.^{vi} But even if this is so, I take my earlier arguments in this section to show that it remains to be seen just how (and why) the extended mind framework is meant to be *more* explanatorily unified than the internalist account: a claim that (as we've seen) Clark's case squarely rests upon. If Sterelny's objection holds any plausible appeal whatsoever, it just pounds one more nail in the already firmly closed case against the unificationist appeal of the extended mind account.

6. Explanatory Power: 1st Wave vs. 2nd Wave Extended Mind

Having more theoretical resources to effectively explain a puzzling phenomenon is a good thing for a theory. Modern meteorology has the power to tell us why it is raining in a way that is much more filled-out, detailed, and satisfying than the ancient Greek hypothesis that the gods were sad. Modern meteorology, thus, has more explanatory power than Greek mythology. How does the extended mind theory compete with the internalist orthodoxy on this measure?

Before tackling this question, we need to become aware of another important distinction made in the literature. John Sutton, in his chapter 'Exograms and Interdisciplinarity', distinguishes between (what he calls) "two waves of extended mind thinking" (Sutton, 2010, p. 193). *1st-wave extended mind*, as Sutton has it, claims "cognitive states and processes extend beyond the brain and into the (external) world when the relevant parts of the world function in

the same way as do unquestionably cognitive processes in the head” (Sutton, 2010, p. 193). Sutton thinks that this is the approach, for example, that Clark and Chalmers employ when motivating sentences in Otto’s notebook as his beliefs. *2nd-wave extended mind*, on the other hand, is based on what Sutton deems a *complementarity principle*. He writes, “In extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes.” Instead, “...different components of the overall system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting” (Sutton, 2010, p. 194). On a *2nd-wave* approach, the things in the world over which the mind extends need not necessarily have functional similarities to traditional mental stuff. What matters is only that those things enter the right kind of complementary relationship with the mind.

In this section, I argue that the *1st-wave* extended mind approach, by virtue of its commitment to coarse-grained functionalism, incurs a significant loss on the theoretical measure of explanatory power: a burden that the internalist need not shoulder. On the other hand, the *2nd-wave* approach, by virtue of its rejection of the importance of functional similarity, sacrifices much of the intuitive pull motivating the extended mind project to begin with, making it difficult to see why we should want to defend it in the first place.

1st-wave extended mind theorists, recall, offer the following sort of argument. (X) Things like computer screens, calculators, and Otto’s notebook are functionally poised to play precisely the same role that traditional folk-psychological mental states play. (Y) Functionalism tells us that what individuates mental states (as the kind of things they are) is entirely the functional role they play.^{vii} Thus, (Z) when things like computer screens, calculators, and Otto’s notebook effectively take on the function of traditional mental states, they should, thereby, count as mental. It is important to notice, however, that premise (X) of the above argument depends heavily on the idea that environmental artifacts (like computer screens, calculators, and notebooks) and traditional mental states (like beliefs and memories) share a significant degree of functional similarity. Without shared function, no case can be made to count them as part of the mind. However, whether this functional similarity holds crucially depends on the coarseness of grain of the analysis given when comparing them. On a very fined-grained analysis, it becomes much harder to defend complete functional similarity between the roles of mental states and environmental artifacts. When Otto needs to get to MoMA, he grabs his notebook, flips it open

to the right page, and *reads* his memory from off a page. Inga, his healthy counterpart, remembers the location of MoMA without needing to read anything. Furthermore, Inga could fully access her memory of MoMA's location in complete darkness, with her eyes shut, or with her hands tied behind her back. No such luck for Otto. Shouldn't these considerations count as relevant differences between the functional poise of Otto's written sentence and Inga's internal memory? And if so, shouldn't this count against premise (X) of the above argument?

Clark and Chalmers dismiss this worry. They say,

These various small differences between Otto's and Inga's cases are all *shallow* differences. To focus on them would be to miss the way in which, for Otto, notebook entries play just the sort of role that beliefs play in guiding most people's lives (Clark & Chalmers, 1998, p. 37).

This response shows, on the part of the 1st-wave extended mind approach, a commitment to a *coarse-grained functionalism*. It is only by ignoring certain differences between the functional profile of sentences in Otto's notebook and Inga's memories that the two get to be seen as playing an equivalent life-guiding role. I argue that this retreat to a coarse-grained level of analysis amounts to a significant loss of explanatory power. In other words, the 1st-wave extended mind advocate, by virtue of this commitment to this coarse-grained level of analysis, cannot take seriously *the differences* between the way sentences in Otto's notebook function for him compared to the way Inga's memories function for her. The 1st-wave extended mind theorist, as we have just seen, cannot elaborate on Otto's need for light, clear vision, free hands, and the like. These differences must be ignored if any hope is to be had of establishing sufficient functional similarity between Otto's notebook and Inga's memory. Furthermore, these do not seem to me to be shallow, insignificant, or unimportant differences. The only reason why someone might think they were, I venture, is if they already had mind-extending aspirations. A sentence in Otto's notebook exists on a page, must be read, and can only be accessed when it is properly lit and within Otto's reach. These are real, non-negligible, functional differences between sentences in Otto's notebook and Inga's memories: differences that necessarily go unexplained by the 1st-wave extended mind theorist. The 1st-wave extended mind approach must be seen, thereby, to incur a significant loss in explanatory power. And since internalism need not be committed to such a coarse-grained level of analysis, it should be seen to be the more explanatorily powerful account.

What about the 2nd-wave extended mind approach? As Sutton defines it, a 2nd-wave approach need not rely on *any* degree of functional similarity between traditional mental states and the environmental entities over which the mind extends. All that matters, for a 2nd-wave theorist, is that the stuff in the world enters into a certain complementary relationship with the mind. As such, the 2nd-wave extended mind theorist, it seems, need not incur any loss of explanatory power at all.

This seems like a good response. However, I suggest that, in retreating to a 2nd-wave approach, the extended mind theorist encounters a new problem.^{viii} Namely, without requiring any degree of functional similarity between external and internal mental stuff, much of the antecedent motivation for taking on the extended mind theory in the first place disappears. As I have described it, the extended mind thesis initially came about in conjunction with functionalism about mental state individuation. The initial intuitive appeal of (the otherwise jarring) extended mind thesis only came to be when it was pointed out that certain environmental artifacts play a similar functional role in supporting cognitive processes as do traditional mental states. If no such degree of functional similarity is required in order to count artifacts as mental, the intuitive support drawn from functionalism dissolves away. If functional similarity is downplayed to the level of unimportance, the extended mind thesis goes back to being jarring and counter-intuitive. In fact, without the intuitive appeal of functional similarity between traditional mental states and the stuff of the world, I can think of no remaining reason for undertaking the recategorization of environmental artifacts as part of the mind. This is precisely the problem with the 2nd-wave account. By jettisoning its reliance on functional similarity, it also jettisons the very intuitive appeal of the thesis they aim to defend. This, I contend, effectively undercuts any motivation for accepting the 2nd-wave account.

7. Conclusion

In this paper, I have argued that the extended mind account is (on balance) more theoretically vicious than it is virtuous. To do so, I have compared the extended mind framework with its internalist rival using several of the traditional theoretical virtues as comparative dimensions. On the measure of empirical adequacy, I concluded that both the extended mind account and its internalist rival are equally empirically adequate, but the underdetermination of theories by their evidence makes this fact neither interesting nor useful. With regard to predictive fruitfulness, I

have suggested that no advantage goes to the extended mind account because a sophisticated internalist has every reason to generate exactly the same good predictions as the extended mind theorist. Explanatory parsimony, I submitted, is no good for selecting between competing theories because any theory can be made more explanatorily simple by making it more abstract. On the measures of ontological and process parsimony, I argued that the extended mind loses out to the internalist by virtue of its commitment to the existence of either a vast (if not infinite) number of new composite minds or a vast (if not infinite) number of distinct cognitive processes. I also showed that no sense can be made of the extended mind theorist's claim to be more explanatory unified, at least not in the sense that is important to philosophers of science. If anything, considerations from evolutionary biology suggest that the extended mind account is less unified with our best science. Finally, on the dimension of explanatory power, I have concluded that 1st-wave extended mind, by virtue of its reliance on a coarse-grained level of analysis, incurs a significant loss in explanatory power: one the internalist need not suffer. The 2nd-wave approach, on the other hand, only maintains its explanatory power at the cost of sacrificing the initial intuitive appeal of the very extended mind thesis they aim to defend.

If these arguments are convincing, I hope to have done much to show that the extended mind account is no better than good-old-fashioned internalism. In fact, it is worse.

Works Cited:

- Adams, F. & Aizawa, K. (2008). *The bounds of cognition*. Oxford: Blackwell.
- Adams, F. & Aizawa, K. (2010). Defending the bounds of cognition. In R. Menary (ed.), *The extended mind*. (pp. 67-81). Cambridge, MA: MIT Press.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 58,10-23.
- Clark, A. (2010). *Memento's revenge: the extended mind, extended*. In R. Menary (ed.), *The extended mind*. (pp. 43-67). Cambridge, MA: MIT Press.
- Clark, A. (2008). *Supersizing the mind: embodiment, action, and cognitive extension*. Oxford: Blackwell.
- Deutsch, D. (2009). A new way to explain explanation.
http://www.ted.com/talks/david_deutsch_a_new_way_to_explain_explanation.html
- Duhem, P. (1914). *The aim and structure of physical theory*, trans. from 2nd ed. by P. W. Wiener; originally published as *La Théorie Physique: Son Objet et sa Structure* (Paris: Marcel Riviera & Cie.), Princeton, NJ: Princeton University Press.
- Hurley, S. (2010). The varieties of externalism. In R. Menary (ed.), *The extended mind*. (pp. 101-155). Cambridge, MA: MIT Press.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Kitcher, P. (1993). *The advancement of science*. Oxford, Oxford University Press.
- Logan, R. (2007). *The extended mind: the emergence of language, the human mind and culture*. Toronto, Ontario, Canada: University of Toronto Press.
- Menary, R. (2010). Cognitive integration and the extended mind. In R. Menary (ed.), *The extended mind*. (pp. 227-245). Cambridge, MA: MIT Press.
- Preston, J. (2010). The extended mind, the concept of belief, and epistemic credit. In R. Menary (ed.), *The extended mind*. (pp. 355-371). Cambridge, MA: MIT Press.
- Putnam, H. (1960). Minds and machines. In Sidney Hook (ed.) *Dimensions of mind*. New York, NY: New York University Press.
- Quine, W. V. (1975). On empirically equivalent systems of the world. *Erkenntnis*, 9, 313–328.
- Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.
- Rupert, R. (2010). Representation in extended cognitive systems: does the scaffolding of

- language extend the mind? In R. Menary (ed.), *The extended mind*. (pp. 325-355). Cambridge, MA: MIT Press.
- Sober, E. (1990). Let's razor Ockham's razor. In D. Knowles (ed.), *Explanation and its limits*, Royal Institute of Philosophy Supplementary Volume 27, Cambridge University Press, pp. 73-94.
- Sterelny, K. (2004). Externalism, epistemic artifacts and the extended mind. In R. Schantz (ed.), *The externalist challenge: new studies on cognition and intentionality*. Berlin: de Gruyter.
- Sutton, J. (2010). Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (ed.), *The extended mind*. (pp. 189-227). Cambridge, MA: MIT Press.
- Tribble, E. (2005). Distributing cognition in the Globe. *Shakespeare Quarterly*, 56, 135-155.
- Wilson, R. (2010). Meaning making and the mind of the externalist. In R. Menary (ed.), *The extended mind*. (pp. 167-189). Cambridge, MA: MIT Press.
- Wheeler, M. (2010). In defense of extended functionalism. In R. Menary (ed.), *The extended mind*. (pp. 245-271). Cambridge, MA: MIT Press.

ⁱ This example comes from personal correspondence with Peter Carruthers.

ⁱⁱ This thirds sense of parsimony certainly overlaps with ontological parsimony, but by no means reduces to it.

ⁱⁱⁱ See Sober (1990)

^{iv} It should be noted that there might be a particularly radical version of the extended mind claim that would deny internally-individuated mental states altogether. On a view like this, the extended mind advocate could respond by claiming not to have added any more new mental states than the internalist is committed to existing internally. However, even if such a radical version of the extended mind thesis were adopted (I do know any who do), what I am concerned with is not just the number of individual mental states that the theories are committed to, but the number of *minds*. And it is on this measure that I suggest the extended mind theorist loses out.

^v See especially Kitcher (1993)

^{vi} Clark tries a few responses in his chapter 'Memento's Revenge' (2010, 58-61) —some of which work better than others—but none of which I take to ultimately dismiss the spirit of Sterelny's charge.

^{vii} See Putnam (1960) for the original formulation of functionalism.

^{viii} Not to mention their frequently noticed slide from causal complementarity to constitution—deemed by Adams and Aizawa as the 'coupling-constitution' fallacy (2008, 2010). Staying silent, for the purpose of this paper, on whether this truly is a fallacy, it is worth noticing this potential additional problem for the 2nd wave approach.