From cognition's location to the epistemology of its nature

Matthew J. Barker

Mount Allison University

Email: mbarker@mta.ca

Telephone: (506) 364-2339

Fax: (506) 364-2645

Department of Philosophy

63D York Street

Mount Allison University

Sackville, NB

E4L 1G9

Canada

Abstract

One of the liveliest debates about cognition concerns whether our cognition sometimes extends beyond our brains and bodies. One party says Yes, another No. This paper shows that debate between these parties has been epistemologically confused and requires reorienting. Both parties frequently appeal to *empirical considerations* and to extra-empirical *theoretical virtues* to support claims about where cognition is. These things should constrain their claims, but cannot do all the work hoped. This is because of the overlooked fact, uncovered in this paper, that we *could never* distinguish the rival views empirically or by typical theoretical virtues. I show this by drawing on recent work on testing, predictive accuracy, and theoretical virtues. The recommendation to emerge is that we step back from debate about *where* cognition is, to the epistemology of *what* cognition is.

Keywords

Extended cognition; situated cognition; extended mind; empirical test; theoretical virtue; individuation

1. Introduction

Sometimes, possibilities escape us. Until Richard Dawkins (1982) showed at book length that phenotypes can extend beyond the organisms they belong to, most of us had not conceived this possibility. Clark and Chalmers' 1998 paper "The Extended Mind" had a similar effect. Until then, only a few had seriously entertained the possibility that cognition extends beyond the brain.

But we are well beyond that now. Even the harshest critics of Clark and Chalmers concede it is possible, even nomologically possible, that cognitive processes extend beyond the brain (e.g., Adams and Aizawa 2001; 2008). But *does* cognition extend? Debate on this issue, stemming from the Clark and Chalmers' paper, has led to a battle between advocates of two views. Shortly I'll precisely characterize these views, but we can start with loose descriptions. Proponents of the *Thesis of Extended Cognition* (TEX) hold that some cognitive processes grow and shrink frequently, protruding outside of, and then

retracting back into, the skull as needed.[1] On this view cognition is *based* within the skull, but as an octopus confined to a cage, with its legs lashing out between the bars, then back in again. Several scientists carry out their research under a TEX banner.[2] To make TEX seem less radical, philosophers have compared extended cognitive processes to processes in other domains that extend beyond the bounds of their bearers. This has seemed especially plausible for feedback processes, where an agent's intracranial cognition shapes extracranial interactions, the products of which are then fed back into and influence other intracranial cognition. Many fish use similarly extended feedback loops to swim, as do turbo engines to drive vehicles.

Opposing this view are advocates of the *Thesis of Embedded Cognition* (TEM), who attempt to capture the importance of features external to the brain without taking the final TEX step of claiming those features are constitutive of cognition. For TEM authors, external features are causally crucial for cognition, and many of our explanations should appeal to them, but cognition is contingently brain-bound.[3]

To keep clear in mind which view goes with which acronym, remember that the acronym that reminds you of the relatively big state of Texas refers to the view that cognition is "bigger" (it extends beyond the brain and body) than most of us have thought. The other acronym refers to the view that keeps cognition within the skull.

Leading authors on either side of the TEX-vs-TEM debate frequently appeal to empirical considerations and theoretical virtues to evidentially favor their preferred view, rather than merely to sway intuitions. They attempt to show, for example, that their preferred view has superior "empirical

---

[1] Philosophical proponents of TEX include Clark (1997, 2007, 2008), Clark and Chalmers (1998), Rowlands (1999), Wilson (1995, 2004), and Wilson and Clark (2009).

[2] See Glenberg (1997), Glenberg and Kaschak (2002), Hutchins (1995), Kirsh (2009), McNeill (2005), and Sawyer and Greeno (2009).

[3] TEX and TEM roughly correspond, respectively, to the views that Rupert (2004, p. 389, p. 393) calls the Hypothesis of Extended Cognition (HEC) and the Hypothesis of Embedded Cognition (HEMC). Shortly I clarify the correspondence and make TEX and TEM more precise than the HEC and HEMC counterparts.

power" (Rupert 2004, p. 407).[4] My chief contribution is to show that this is an epistemological mistake and that we need to reorient the debate accordingly. Specifically, I argue for a strong modal thesis. It is not just that TEX and TEM are presently at an empirical stalemate, one that we could resolve, or one we could reasonably live with in virtue of favoring TEX or TEM via extra-empirical theoretical virtues. Rather, I draw on recent work on testing and predictive accuracy to uncover the fact that we *could never* distinguish TEX and TEM empirically (Section 3) *or* by theoretical virtues (Section 4).[5]

As we will see, my thesis leaves open that disagreement between proponents of TEX and advocates of TEM is reasonable; my thesis does *not* imply that the disagreement is merely verbal, for example. My thesis also leaves open that the disagreement is important; for instance, TEX and TEM may motivate distinct methodologies despite being indistinguishable in the respects I discuss. Indeed, on the assumptions that disagreement about TEX and TEM is both reasonable and important, my thesis motivates us to reorient debate between TEX and TEM. We should steer some attention from empirical considerations and theoretical virtues, to the prior task of determining what methods we could and should use to successfully favor one view over the other. In closing (Section 5), I suggest we do this by turning from trendy questions about where cognition is, to urgent epistemological and methodological questions about what it is. This is to confront the epistemology of individuation. To start though, let me clarify TEX and TEM.

2. Theses of Extended and Embedded Cognition

To formulate TEX and TEM precisely, let *brain features* refer only to matter, activity, interactions, states, structures, etc., of the brain or wholly realized or implemented by the brain. Let *external features* refer only to matter, activity, interactions, states, structures, etc., external to the brain or at least partially

[4] Guilty TEM advocates include Rupert (2004) and Adams and Aizawa (e.g., 2001; 2008); on the TEX side there is Clark and Chalmers (1998) and Wilson and Clark (2009).

[5] I have benefited from quite divided responses to this thesis while circulating drafts of the paper. Some readers have said that they're not at all surprised to hear that we cannot distinguish TEX and TEM empirically or by theoretical virtues; I am happy that they agree with me. Others have been surprised by my thesis, and convinced there must be something wrong with my argument for it; I am also happy with this response, because it helps show that the argument is important.

realized or implemented outside the brain. In these terms, some TEXers carefully imply that external features are proper parts of the *minimal supervenience base* of cognitive processes, rather than simply proper parts of cognitive processes, and that they *physically* constitute cognitive process, rather than simply constitute those processes. But we can conveniently and harmlessly recognize these qualifications implicitly, and define TEX as follows:[6]

> *TEX:* External features often help produce and explain cognitive phenomena, and in some of these cases the external features are partly constitutive of cognition.

Exactly when external features play this role, let's say they *constitutively shape* cognitive processes that help produce and explain cognitive phenomena. Next:

> *TEM*: External features often help produce and explain cognitive phenomena, but they in fact never partially constitute cognition and brain features in fact always exhaustively constitute cognition.

According to TEM, precisely when external features help produce and explain cognitive phenomena, they do so by virtue of their relations to the cognitive parts of the processes that produce those phenomena. Let's say that when external features do this, they *causally shape* cognitive processes.[7]

We can now clarify how TEX and TEM relate to what Rupert (2004) calls the hypothesis of extended cognition (HEC) and the hypothesis of embedded cognition (HEMC), respectively. My TEX roughly corresponds with his HEC, and my TEM roughly corresponds with his HEMC. Rupert initially characterizes the two views he discusses as follows:

---

[6] You might think the qualifications would be misleading anyway, because you are an identity theorist about cognition. But if you are, the debate about extended cognition remains open. It will be a debate about whether some of the parts of the physical features with which cognition is identical are outside the brain. All my claims in this paper could be translated into this context without loss.

[7] In the next note I distinguish two ways of understanding this claim.

HEC (cf. TEX): "human cognitive processing literally extends into the environment surrounding the organism, and human cognitive states literally comprise—as wholes do their proper parts—elements in that environment" (p. 389).

HEMC (cf. TEM): "cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place" (p. 393).

Take the correspondence between TEX and HEC first, as it is a simpler correspondence than that between TEM and HEMC.

My definition of "TEX" very modestly sharpens Rupert's initial characterization of HEC, through quantifiers. Rupert's initial characterization of HEC is ambiguous between the view that all human cognitive processes extend and the view that only some do. His subsequent discussion clarifies that it is the latter view that is in play, as parties to the debate intend (e.g., Wilson and Clark 2008). My definition of "TEX" simply makes this explicit. The definition of "TEX" also makes the controversial appeal to constitution more conspicuous.

My definition of "TEM" avoids the appeals to the vague notion of "depends heavily" and the agent-relative notion of "unexpected ways" that we find in the initial characterization of HEMC; it also clarifies the quantifiers, for easy comparison to TEX. Perhaps more importantly, the initial characterization of HEMC is compatible with the initial characterization of HEC, but TEX and TEM are clearly incompatible. TEX implies that cognitive processes sometimes extend beyond the brain, and TEM implies they do not. Technically then, TEM is a sharpened version of HEMC *plus* the denial of the constitution claims in TEX/HEC. I advance to this incompatibility straightaway because it is the one that all parties to the debate argue about. Indeed Rupert's own view is not simply HEMC, but TEM. He denies TEX/HEC and characterizes HEMC in a compatible way only so that he can briefly entertain the

possibility that HEMC follows from HEC (p. 393ff.). After quickly dismissing this possibility, he goes on to argue for TEM, not simply HEMC.[8]

Now to clarify the general idea of *process* extension, and the dispute about *cognitive* extension in particular, consider a biological case, then a cognitive one.

Adams and Aizawa (2001) concede an extensionist interpretation of digestion in some spiders. Before ingesting their prey, some spiders inject the prey with enzymes that liquefy the prey's innards. Thereby, the spiders control and anchor feedback processes of digestion that extend beyond their own bodily bounds: they produce substances that they manipulate to interact outside their bodies with other substances, and they then exploit the products of those external interactions to shape their internal digestion. Adams and Aizawa allow that these external interactions *constitutively* shape the spider's digestion process, so that its digestion extends beyond its bounds.

But Adams and Aizawa, and TEM proponents generally, resist extensionist interpretations of cognitive cases. When teachers show others how to solve a math problem, the teachers often describe the sequence of mathematical operations involved in the solution by gesturing with their hands towards math symbols they have drawn. As shown in more detail below, data suggests the teachers' gesturing lightens their cognitive loads while they explain their math solutions: wholly brain-based cognitive interactions help produce the external gestures, which then feedback into the brain so that working memory is freed. On TEM interpretations, gestures merely *causally shape* the cognitive process that produces the given teacher's math explanation. However, on Clark's (2007; see also 2008) TEX interpretation, the gestures *constitutively shape* that cognitive process. Let us say that both TEX and TEM advocates claim that some cognitive processes are *transbrain* processes. But according to TEX, some *extrabrain* parts of transbrain

[8] I apologize for introducing new acronyms into the debate. If you do not think that the differences just mentioned in the text warrant new acronyms, consider that "TEX" and "TEM" are easier for you to pronounce when talking about the views. They also avoid association with the term "hypothesis" (in favor of "thesis"), which has certain empirical connotations that I am trying to steer us clearer of in this debate.

cognitive processes are *cognitive parts* of those processes, while according to TEM the extrabrain parts are always *un*cognitive parts of the transbrain cognitive processes.[9]

Diverging claims about necessity underlie this dispute. TEM advocates propose conditions that must hold for a part of a process to be cognitive, and TEX advocates deny those claims to necessity. The TEM claims to necessity typically appeal to similarity. Rupert (2004), for instance, notes that the uncontroversial cases of cognitive process parts are wholly brain-based; he then claims these are so dissimilar from extrabrain parts of transbrain cognitive processes that we should conclude that the extrabrain parts are not cognitive. Adams and Aizawa (2001; also see 2008) are more specific. They identify two necessary conditions for a process part's being cognitive: it must involve *non-derived content* and feature the *same kind of processing* as other cognitive processes. The extrabrain parts of transbrain processes do not, they argue, meet these conditions. They leave open just what non-derived content is, and what kind of processing marks cognitive processing. But they do review leading accounts of such content and processing, and note that extrabrain parts of transbrain cognitive processes would not satisfy the content or processing conditions on any of these accounts (see also Fodor 2009). A concrete example we can use throughout the paper is Adams and Aizawa's preferred account of non-derived content and processing, a computational-representational view of cognition descended from Fodor (1987; 1990). On this view, cognitive processing corresponds to computations over representational language-of-thought-symbols (LOTS).[10] Were Adams and Aizawa to insist on this specification of their general view, their two necessary conditions for a process part's being cognitive would be that the part a) contains LOTS, which b) computations operate over.

---

[9] Alternatively we could say that, according to TEM, external and brain features are related in ways important to the production of cognition, but nonetheless do not form a single process. On this view there is a wholly brain-based cognitive process and a distinct external non-cognitive process, and these merely relate crucially for cognition. For my purposes, the differences between this interpretation of TEM and the one I use in the text, as well as any related differences in process individuation criteria, are irrelevant.

[10] These symbols have non-derived content in that they have non-derived meaning. A symbol 'X''s content X is its meaning, and this is non-derived, only if that meaning does not derive from conventions and social practices. According to Fodor it is *non-derived* "if it is the case that there is a law connecting Xs to 'X's, and all other laws connecting non-Xs to 'X's exist only in virtue of the existence of the X to 'X' law" (Adams and Aizawa 2001, p. 48). Processing of such symbols is *computational* in that the processing manipulations of the symbols are governed by algorithmic laws that relate their syntactic and semantic structures.

TEX advocates reject these general and specific claims to necessity. Wilson and Clark (2009) suggest that even if any given cognitive process necessarily involves non-derived content, not all *parts* of the process must involve such content to be cognitive parts; a part must involve non-derived content *or* be properly related to other parts with non-derived content.[11] Wilson and Clark vie for *hybrid* cognitive processes, where it is precisely the complementary dissimilarities between tightly integrated parts of the processes that make those parts cognitive and allow for the production of cognitive phenomena.

Having clarified the dispute between TEX and TEM, let me now show that we could never discriminate between TEX and TEM empirically or by theoretical virtues.

3. Empirical Indistinguishability

Typically, empirical data only favors a given view *with respect to a competing view* (Royall 1997; Sober 2008). In this contrastive framework there are two ways we might try to empirically discriminate between TEX and TEM. First, our empirical data could provide bases for *discriminating tests* between them. Second, our estimates of the *predictive accuracy* of each might differ.[12] It is best to step through these two possibilities with some light technical machinery. This will afford precision that clarifies the strength of the argument. For evaluators of the argument, the precision will also helpfully illuminate which claims are doing what work.[13,14]

3.1 Testability.

[11] Wilson (2008) provides a less ecumenical rejection of the TEM appeal to non-derived content.

[12] Below I suggest that other means of empirically distinguishing theories are relevantly like one or both of these two ways, such that these two ways represent all the options for empirical distinguishability.

[13] The machinery and associated precision also help correct, as we will see, key under-developed claims that TEX and TEM authors make and which are of a sort more thoroughly addressed in the philosophy of science using some machinery. For a flavor of the under-development, see Rupert's 2004 and 2009 appeals to empirical power, conservativism and simplicity, as well as Clark's 1997, 2007 and 2009 appeals to explanatory unification, empirical bets, and empirical considerations, respectively.

[14] So long as we agree that in most contexts empirical support is essentially contrastive, my discussion of testing will apply whether you think tests between hypotheses provide reasons for thinking one hypothesis is more probably *true* than the other (as most likelihoodists and Bayesians do), or instead reasons for *not rejecting* one and rejecting the other (where this question is independent of truth). Thus it will be harmless when, for convenience, I write as though taking one of these two testing perspectives rather than the other. When discussing predictive accuracy in Section 3.2 it will be more important to adopt an instrumentalist perspective.

Let's adopt a necessary condition of testability from Sober (2008) and deploy it in a fuller discussion of the gestures case that Clark (2007) heavily leans on in a recent defense of TEX:[15]

> *Testability condition*: For any two hypotheses H_1 and H_2, we could discriminate H_1 and H_2 by test only if there could be some true observation statement O and set of independently attested auxiliary assumptions A, such that $\Pr(O|H_1\&A) \neq \Pr(O|H_2\&A)$.

The probabilistic terms in this condition are likelihoods, e.g., $\Pr(O|H_1\&A)$ is the likelihood of $H_1\&A$. As such, $\Pr(O|H_1\&A)$ is not the probability of $H_1\&A$ at all, but rather the probability of the observation or data to which O refers, given H_1 and A. Testability requires that $H_1\&A$ and $H_2\&A$ differentially probablify the data.[16]

In her 2003 book, and with colleagues (2001), Goldin-Meadow reports the leading empirical studies of gesturing that Clark interprets as supporting TEX. The researchers asked whether gesturing lightens cognitive load, where cognitive load refers to burdens on an agent's working memory processes, especially while she is attempting to learn or instruct (Sweller 1988; 1994). Researchers indirectly measure effects on cognitive load during, say, instruction, by assessing performance on other related cognitive tasks.

To measure the effect of gesture on working memory processes implicated in explaining math solutions, Goldin-Meadow et al. (2001) quantified performance on a recall task performed simultaneously with the explanations of math solutions. Before subjects began explaining a math solution (which they had already completed themselves), the experimenters gave them a list of words (in the case of children) or pairs of letters (in the case of adults). Subjects were to hold the list in mind while explaining the math solution, then recall the list for experimenters after they had finished explaining the math solution. In the

---

[15] This condition is a piece of Sober's full definition of testability. His main innovations over past (and notorious) definitions are to define testability contrastively and probabilistically.

[16] Notice that this condition is free of any Bayesian*ism* or appeal to Bayes's Theorem: only *likelihoods of hypotheses* are involved, not prior or posterior *probabilities of hypotheses*. Likelihoods of hypotheses and probabilities of hypotheses are quite different things (see Sober 2008) and generally it is the probabilities, not the likelihoods, which are controversial.

control condition, subjects were *permitted* to gesture while explaining their solutions, which were on a chalkboard.[17] In the manipulation condition, they were *not* to gesture while sitting at a table and explaining their solutions.

Consider two tests one might run with this experimental design. The first is a test Goldin-Meadow and colleagues ran, the second is a hypothetical test between TEX and TEM.

In the actual experiment, the auxiliary assumptions, *A*, that researchers coupled with hypotheses to make predictions involved assumptions about potential experimental error and the effectiveness of controls.[18] Let's presume these assumptions are true. Two of the hypotheses then proposed were:

H_1: Gesturing lightens the cognitive load of the working memory system involved in explaining the math solution, such that it *shapes* the cognitive process(es) underlying the explanation of that solution.

H_2: Gesturing is epiphenomenal with respect to the cognitive load of the working memory system involved in explaining the math solution, such that it is *epiphenomenal* with respect to the cognitive process(es) underlying the explanation of that solution.

These hypotheses generate the following competing predictions:

H_1&*A Prediction*: Subjects instructed not to gesture perform *poorer* on the recall task than subjects allowed to gesture.

H_2&*A Prediction*: Subjects instructed not to gesture perform about the *same* on the recall task as those allowed to gesture.

[17] The authors do not say whether permitted gesturing was explicitly or implicitly permitted. But they appeal to the permitted cases to rule out certain confounds in a way that suggests permitted gesturing was implicitly permitted (Goldin-Meadow 2001, p. 521).

[18] Controls were to show the following: remembering not to gesture does not tax cognitive load; gesturing does not diminish with increasing difficulty of math problems explained; math expertise does not affect recall performance; gesturing does not save explanation time; significant changes in recall performance between test conditions cannot be attributed to gesture shifting cognitive load from one cognitive system to another (rather than lightening load on a single system); the explanation and recall tasks involve the same working memory system; and so on.

Generally, researchers observed that non-gesturers performed poorer on the recall task than gesturers. For clarity, consider the researchers' observation of children who were to recall the longer lists of words:

O: Subjects instructed not to gesture perform poorer on the recall task than those allowed to gesture.

It is clear from the above predictions that H_1 confers a higher probability on O than H_2 does. That is, $\Pr(O|H_1\&A) > \Pr(O|H_2\&A)$. This entails $\Pr(O|H_1\&A) \neq \Pr(O|H_2\&A)$. The testability condition is satisfied.

The results of the test between H_1 and H_2 would please both advocates of TEX and of TEM, but not discriminate between TEX and TEM. For such discrimination, we must pit competing TEX and TEM hypotheses about gesturing against each other:

H_{TEX}: Gesturing lightens the cognitive load of the working memory system involved in explaining the math solution, such that it *constitutively shapes* the cognitive process(es) underlying the explanation of that solution.

H_{TEM}: Gesturing lightens the cognitive load of the working memory system involved in explaining the math solution, such that it *causally shapes* the cognitive process(es) underlying the explanation of that solution.

When conjoined with A from above, these hypotheses generate predictions about the gesturing data:

H_{TEX}&*A Prediction*: Subjects instructed not to gesture will perform *poorer* on the recall task than subjects allowed to gesture.

H_{TEM}&*A Prediction*: Subjects instructed not to gesture will perform *poorer* on the recall task than subjects allowed to gesture.

These predictions appear identical. However, the shared allusion to "poorer" performance by non-gesturers is vague. Do these predictions, and so the hypotheses from which them come, differ at a more fine-grained level by suggesting differing degrees of poorer performance by the non-gesturers?

No. H_{TEX} and H_{TEM} are predictively equivalent because the difference in the natures they attribute to the shaping relation—and the implied difference in whether external features are cognitive—is undetectable by observation of performance on the recall task. But nothing special about the recall task makes this true. The problem is more general. Both advocates of TEX and those of TEM claim that some external features hook up with brain-based cognition in crucial, productive, explanatory ways. But from our observational perspective, *nothing* about the transbrain productive processes thus formed, or the cognitive phenomena produced by those processes, depends on whether only the brain-based parts of those processes are cognitive or instead some extrabrain-based parts of those processes are also cognitive. More generally, there is just *one* thing about the transbrain processes that depends on which of the two possibilities is the case: which of the contested process parts are cognitive and which are not. But none of our empirical tests between competing TEX and TEM hypotheses could be sensitive to this lone difference. So for *any* TEX hypothesis, H_{TEX}, *any* competing TEM hypothesis, H_{TEM}, and *any* of our observations O (appropriately coupled with auxiliary assumptions A), it is the case that $\Pr(O|H_{\text{TEX}}\&A) = \Pr(O|H_{\text{TEM}}\&A)$.

Talk of process parts fills in this argument. Let P be any cognitive phenomenon that any transbrain process T produces. T will include some external feature F as a proper part, since it is a transbrain process. Further, some portion of T will be cognitive; call this C. One way to articulate the difference between F's constitutively shaping C, and F's causally shaping C, is to say that F constitutively shapes C just when F is not only a proper part of T, but also of C; otherwise, F merely causally shapes C.[19] But in this context of *processes*, the proper part relation is similar to the causal relation in ways that preclude the testability we seek. Both relations are diachronic and productive. If F *causally* shapes C, F produces an

[19] For this latter claim to hold when there is some part of T intermediate between F and C, we may have to assume causation is transitive.

effect *E* that is a part of *C*, and *F* is temporally prior to *E*. But *F* also does this if it *constitutively* shapes and is a proper part of *C*. And in either case, *F* helps produce *P* in virtue of producing *E* in *C*. The only relevant difference between the two possibilities is this. In one case the line between what is cognitive and what is not falls *between F* and *E*, so that *E* is part of *C* (is cognitive) and *F* is *not* (is not cognitive); in the other case the line falls *before F*, so that *both F* and *E* are parts of *C* (both are cognitive). Alternate this line's location between the two possibilities and you get no corresponding detectable difference. *F* remains temporally related to *E* in the same way, there is no change in production that we could infer, no change in any laws we could discover, no change in any entity that we could detect. There must be some such change if we are to discriminate the two possibilities by empirical test, but there is not. *F* can be as (dis)similar to the brain-based parts of *C* as you like, and the shaping relation between *F* and *C* can be as integrated as you wish, and all of this will still hold.

Thinking in term of cognitive states, instead of processes and their parts, may have concealed this conclusion from parties to debate about extended cognition. Going astray in this way is especially easy when we pair our thinking about states with a common understanding of the constitution relation that is in play. Typically, we think of the constitution relation as a synchronic relation of unique determination between states (or properties etc.) such that if *B* constitutes state *C*, then *B* at time *t* is necessary and sufficient for *C* at *t*. Put this consequent differently: iff *B* at *t*, then *C* at *t*. We then seem to have a basis for empirically testing between TEX and TEM claims, at least in principle. Let *C* be some state that both sides agree is a cognitive state of some system of ours. Let *B* be some brain feature of ours. If we introduce *E* as some (perhaps variously tokened) type of external feature in our environment, it is natural to think that some competing TEM and TEX claims take the following forms, respectively:

TEM Hypothesis: Iff *B* at *t*, then *C* at *t*

TEX Hypothesis: Iff *E*+*B* at *t*, then *C* at *t*

TEM Hypothesis, but not TEX Hypothesis, says that *B* is sufficient for *C*. So can't we test between the hypotheses by allowing *B* but not *E* to obtain? If we do this, isn't it true that if *C* occurs we favor TEM over TEX, and if *C* doesn't occur we favor TEX over TEM?[20]

Unfortunately this won't work. The hypotheses just introduced mischaracterize TEX and TEM, and when we correct for this the testability vanishes. The hypotheses concern *synchronic* instantiation of a state, but TEX and TEM disagree about whether external *diachronic* portions of transbrain processes are cognitive.[21] Although TEM Hypothesis is "iff *B* at *t*, then *C* at *t*," TEX can also furnish this hypothesis. A TEXer can reasonably say, "sure, once we get to time *t*, I'm happy to say that if *B* obtains at that time, then so does *C*. This is because I'm concerned about the status of *E* at some earlier time *t*-1. I'm claiming that *E* is so intricately and systematically related to *B*—and thereby so integral to the production of the cognition in question—that we should count *E* (as we do *B*) as a cognitive part of the process that produces *C*. This fits perfectly well with the further claim that once *E* leads to *B*, then *B* guarantees *C*." This TEX clarification spoils the test. Because TEX agrees with TEM Hypothesis, both views predict that allowing *B* to obtain without *E* preceding *B* at *t*-1 will result in *C*. So observing *C* does not favor TEM over TEX as hoped.

We should come at the test from the other side to fully appreciate that it is spoiled. Because TEXers can agree with the TEM Hypothesis claim that *B* at *t* is sufficient for *C* at *t*, they can deny being saddled with the TEX Hypothesis claim that *E*+*B* at *t* is necessary for *C* at *t*. Indeed, TEXers *should* deny this, because it is incoherent to say that *E*+*B* does or could occur at any one *moment t*; *E*+*B* cannot do this because it is a process, it is extended in time. Because TEXers deny the necessity claim in the hypothesis that we have mistakenly attributed to them, if we allow *B* to occur and *C* does *not* obtain, this does not favor TEX over TEM. Our route for favoring TEX fails, as did our route for favoring TEM. The testability has disappeared.

---

[20] Thanks to an anonymous reviewer for stating this line of reasoning in these clear terms.

[21] Below I entertain competition of TEX and TEM in synchronic terms.

Converting TEX to a claim about synchronic coupling cannot save the testability that I extinguished by emphasizing the diachronic nature of the disputed cases. In the cases we have considered, part of the problem for testability is that the very external features that TEXers seize upon will be the same ones that TEM insists, in its promotion of the embedding of cognition, are causally and explanatorily important for cognition: empirical testing is sensitive to causal explanatory features, but in disputed cases, TEX and TEM will both agree that the external features in question are among the explanatory causes of cognition. The form of the problem remains the same if we switch TEX from a view about external features that is partially motivated by how those features diachronically relate to brain features in tightly integrated and productive ways, to a view about external features that is partially motivated by how those features synchronically relate to brain features (and cognition). This is because we must continue to respect TEM's insistence on the embedding of cognition, by having it also recognize any explanatorily impressive synchronic relations between external and brain features (and cognition). In effect, both views can insist that external features sometimes lawfully and synchronically couple with brain features in ways that help explain cognition. The lawful relations they *both* could recognize would be testable, but irrelevant here precisely because both TEX and TEM accept them. The only difference between TEX and TEM would be that TEX makes the additional claim that the external features are constitutive of cognition, while TEM would deny this addition. But empirical testing would be insensitive to this disagreement, as in the diachronic case. For example, TEX Hypothesis and TEM Hypothesis would still mischaracterize TEX and TEM, though in slightly different ways. We would flip from TEXers agreeing with TEM Hypothesis (and denying the hypothesis we attributed to them), to TEMers agreeing with TEX Hypothesis (and denying the hypothesis we attributed to them). In the disputed cases, both proponents of TEX *and TEM* would accept claims of the form "iff E+B at t, then C at t." And so the results of tests in which we fix B and manipulate E could not discriminate between the views. The remaining and untestable disagreement would be about the interpretation of the necessity of E at t for C at t. TEX proponents would interpret this as suggesting that E is a *constitutive* part of C's realization. TEM proponents would interpret

it as suggesting that *E* is a part of the needed and important context within which the real constitutive parts of *C*'s realization to do their work of determination.[22]

So, returning from our foray into synchronic cases, imagine that our technologies enjoy stunning advances and we are able to *observe* (not merely test for) all physical inner workings of cognitive processes, and all their products. We observe many transbrain cognitive processes and (to the chagrin of Fodor's critics) discover it is a fact that all and only the brain-based parts of those transbrain processes involve LOTS that are computed within those processes. This could leave the dispute between TEX and TEM entirely open. For instance, given the discovery, TEM advocates such as Adams and Aizawa might make the step from insisting that cognitive process parts necessarily involve non-derived content and certain processes, to further specifying that those parts necessarily involve computations over LOTS. But TEX advocates could insist that although all and only brain-based parts of the transbrain processes *do* (we are supposing) involve computed LOTS, computed LOTS are *not necessary* for a process part's being cognitive. For instance, the minds of Martians might involve cognitive processes with no computed LOTS. Granted, Martians may have something sufficiently like computed LOTS. But TEX advocates could also maintain that what is necessary (or sufficient) for a process part's being cognitive is not simply containing computed LOTS or something sufficiently like them, but rather containing such things *or* being properly related to a process part that does contain them. Both disjuncts in this disjunctive TEX condition would predict the occurrence of the computed LOTS that we are supposing we have discovered in all observed instances of cognition, just as the specified TEM condition would predict. The discovery would not, could not, be a test that discriminates between the TEX and TEM conditions. We could never empirically discriminate by *test* between such competing claims, in any TEX-TEM dispute. Let me now turn to close off the only other option for attempting to *empirically* distinguish TEX and TEM.

3.2 Predictive Accuracy Estimation.

[22] For more detail, see Wilson's (2004) inclusion of external features among cognition's realization, and Rupert's (2009) opposing relegation of these features to the surrounding context.

Estimations of predictive accuracy for two hypotheses can differ, and thereby empirically distinguish those hypotheses, even as data *piles up without distinguishing them*. This is because estimations of predictive accuracy are a function not only of fit-to-data, but also of simplicity. To see this, briefly consider the notions of predictive accuracy and simplicity.

When predicated of theories, simplicity is typically an unclear notion without epistemic significance (Sober 1996; 2002; 2006). However, in the statistical field of *model selection studies* simplicity has a concrete formulation and epistemic significance. Models are *composite* hypotheses: they have at least one adjustable parameter and are equivalent to disjunctions of multiple *simple* hypotheses that are each distinguished by the values given to the adjustable parameters in them. A model M_1 is simpler than another model M_2 if M_1 has fewer adjustable parameters than M_2 (Forster and Sober 1994). As Hitchcock and Sober (2004, p. 11) note, "it is a well-confirmed fact" that models that are more complex in this sense are often worse predictors of new data than are simpler ones. Complex models typically fit old data better, in virtue of their greater number of parameters, but the tighter fitting weakens their capacity to capture new and different data. This renders them less predictively accurate. The predictive accuracy of a model is its *average* performance on predicting new data, given values of its parameters that were estimated to maximize fit to previous data (Forster and Sober 1994). The most predictively accurate models tend to strike the right balance between fit-to-data and simplicity. Doing so makes a model useful (for future prediction), regardless and often quite independent of whether it is true (Sober 2008).

Remarkably, Hirotugu Akaike (1973) proved a theorem that describes and even explains (Hitchcock and Sober 2004) how predictive accuracy trades off fit-to-data against simplicity. The theorem describes how to achieve unbiased estimates[23] of the predictive accuracy of models; within an instrumentalist framework we can then compare competing models in terms of their predictive accuracy. We empirically distinguish them if they score differently in these terms, and so long as one is simpler than the other they can score differently even as tests repeatedly fail to discriminate between them. So despite the results of

[23] Here, an estimator is unbiased *iff* the *expected value* of its result equals the true predictive accuracy of the model in question (Sober 2008, p. 86).

Section 3.1, could we empirically distinguish any competing TEX and TEM models in terms of predictive accuracy?

No, and this is clear given the *reasons* for the results in Section 3.1. Every method of estimating predictive accuracy relies *just* on measures of fit-to-data and simplicity (Hitchcock and Sober 2004). My foregoing argument implies that any competing TEX and TEM models will score precisely the same on fit-to-data. But less obviously it also implies those models will score precisely the same on simplicity too, because they will have the same number of adjustable parameters. Such parameters are posited in order to track and accommodate lawful and/or production relations between phenomena modeled. But as we saw, no such relations within the transbrain productive processes that TEX and TEM advocates argue about will depend on whether only the brain-based parts of those processes are cognitive or instead some extrabrain parts of those processes are also cognitive. Thus the dispute about what counts as cognitive will not give rise to differing numbers of parameters between competing TEX and TEM models. Given also the equivalence in fit-to-data, any competing TEX and TEM models can enjoy identical estimates of predictive accuracy. The reasons I have given for this ensure that we cannot empirically distinguish TEX and TEM by related methods either (more on this in the next section). And since the only other means we have for distinguishing them empirically (by empirical test) is also unavailable, we can never empirically distinguish TEX and TEM.

For all this, there may remain a metaphysical fact of the matter about whether external features that shape cognitive processes do so causally or constitutively. Nothing I have said rules this out.[24] Rather, if there is such a fact, I have ruled out our empirical access to and consequences of it. It is also important to emphasize that this is a claim about *our* empirical access. Perhaps there is a possible world with fantastical creatures who observe metaphysical process part-whole relations, or who in any case empirically discriminate between the claim that some external features that help produce cognitive

[24] Some authors such as Alan Sidelle (1989) might disagree. Partially from compelling claims about imagination-based approaches being our only promising approaches to the epistemology of necessity, Sidelle draws the conclusion that issues of necessity—such as the one underlying the TEX-TEM dispute—are conceptual and conventional, not metaphysical. I am unsettled about his conclusion. One worry is that the world may determine the truth values of claims to necessity even if we cannot see this past our concepts and conventions. The outstanding issue is whether we have good reason to think the world does this.

phenomena are cognitive, and this claim's negation. But we are not these creatures, and presumably never could be.[25]

4. Explanation and Extra-Empirical Indistinguishability

Parties to the TEX-TEM debate may feel I have missed some things. Their claims to empirical power for their preferred views are sometimes couched in terms of explanation (e.g., Clark 1997; Clark 2007; McNeill 2005; Wilson 1995; Adams and Aizawa 2001; Rupert 2004). In particular, TEX advocates have suggested that we should endorse their view because it is *simpler* (e.g., Clark and Chalmers 1998), more *unified* (e.g., Wilson 1995, p. 86), and more conducive to *understanding* (Clark 1997, p. 112) than TEM rivals. Authors attach these supposed virtues to both explanations and theories, and perhaps they attach differentially to TEX and TEM.

But a dilemma here faces any TEX or TEM advocate. Either the virtue will clearly be epistemically significant, yet attach to both TEX and TEM equally, or the virtue's epistemic significance will be entirely unclear (at best) or nonexistent (at worst). Merely sketching out this argument will suffice here.

One epistemically significant virtue of many scientific explanations is their capacity to generate empirically testable *predictions*. But we have seen that any competing TEX and TEM explanation will score equivalently on this virtue.

Appeals to *simplicity* in the TEX-TEM debate are uniformly brief and vague. We now know that remedying this by elaborating simplicity just in terms of numbers of adjustable parameters will bear no epistemic fruit in this debate. Authors have argued that certain other notions of simplicity are epistemically significant in some contexts, as Sober (2002) summarizes. But *these* notions will also bear no epistemic fruit here because any competing TEX and TEM hypotheses will again be equivalently simple on them.

[25] There is no need to rest this modal hunch on appeal to unfashionable intrinsic essences of our species (though see Devitt 2008 for resurrection of this fashion). It can rest instead on more fashionable versions of the homeostatic property cluster (HPC) view of kinds (see Wilson, Barker and Brigandt 2007) or wholly relational essentialism (see Okasha 2002), both of which are compatible with the consensus view that species are historical entities.

Most appeals to *unification* in the TEX-TEM debate are also brief and vague. An exception is Andy Clark's (1997, pp. 110-113) claim that some "emergent" TEX explanations are more unified than TEM alternatives. According to Clark, these TEX explanations appeal to "collective variables," which track patterns resulting from interactions among multiple features. These features may include both external and brain features, in which case the system to which the emergent explanation appeals extends beyond the brain. But TEM advocates can take all this onboard. They can appeal to the *same* collective variables and corresponding extended systems. In doing so their explanations would differ only by not deeming any external features cognitive. If epistemically significant unification depends on differences in number of variables, there is no epistemically significant difference in TEX and TEM unification here. This same line of skepticism would also afflict any TEX or TEM advocates' attempt to utilize Kitcher's (1989) well-known account of unification, or similar accounts of unification revised to overcome the well-known problems with Kitcher's view (see Woodward 2003). On the only other accounts of epistemically significant unification that I know of, unification is a function of, or tracks, epistemically significant simplicity (see Sober 2003). But the above problems for the appeal to simplicity follow TEX or TEM appeals to such unification.

Finally, TEX advocates sometimes suggest their view *enhances understanding* more than TEM does (e.g., Clark 2007). Unfortunately, on the only rigorous attempts to explicate such enhancement as an epistemically significant virtue, enhancement stems from unification (e.g., Friedman 1974). But perhaps there is a promising idea in the vicinity, one best considered in terms of theories rather than explanations: either TEX or TEM may better help us generate hypotheses. Even though all the same predictions and discoveries appear available to TEX and TEM, one of these views may provide a more inspiring or clarifying (and so perhaps more expedient) vantage point from which to study cognition.

Clark sometimes hints that this is ultimately what he has in mind, such as when he calls TEX a mere "guiding idea" (p. 13) that helps structure research pursuits. Sometimes he even grants that the virtue of hypothesis fertility may favor TEM over his own view in some contexts: "as cognitive scientists we can (and should) practice the art of flipping between these different perspectives, treating each as a lens apt to

draw attention to certain features, regularities and contributions while making it harder to spot others, or to give them their problem-solving due" (p. 19). This passage locates the virtue of hypothesis fertility clearly within the *context of discovery*, not the *context of justification* that has concerned me here (Reichenbach 1938). It may be worth pursuing this basis for choosing between TEX and TEM within the context of discovery. But the pursuit will be steep because there is so far no systematic discussion of such a basis.

5. Conclusion and Debate Relocation

There may be good pragmatic reasons for privileging one of TEX or TEM with, say, more research resources than the other. But this paper has concerned epistemology and metaphysics, not pragmatics. It has shown that we could never distinguish TEX and TEM empirically or by theoretical virtues. TEX and TEM remain strictly incompatible though. Are there additional epistemological tools we could enlist to discriminate these views? Thought experiments and imagination, perhaps. But if we enlist these tools, we should first (at least temporarily) relocate the debate.

Currently the debate centers on the truth of competing TEX and TEM claims that are agreed to be contingent. Everyone agrees that our cognition could extend beyond us; the issue is *where*, in fact, cognition is. Imagination-based thought experiments are notoriously poor arbiters for competing contingent claims. But they may be better arbiters, or relevant arbiters among others, when asking a different question: *what* is cognition? Even many empiricists would agree that empirical considerations alone cannot answer this. And the answer to this question may well, in conjunction with empirical facts, imply that one of TEX and TEM is favored over the other.

Certainly, some TEX and TEM advocates have used the debate about where cognition is as backdoor for advancing at least implicit views about what cognition is. But this paper suggests that we need to more directly and explicitly debate the issue of what cognition is. Specifically, my argument directs us to the disagreements that I suggested underlie the dispute between TEX and TEM. These are the disagreements about the necessary and sufficient conditions for cognition, which include disagreements about *whether*

there even are conditions of both sorts. But to settle these disagreements we need to address prior and urgent epistemological and methodological questions. How could and should we tell whether there are both necessary and sufficient conditions for a thing's being cognition? Assuming or deciding that these conditions exist, how could and should we tell what they are? We need to start with these tough questions to avoid spinning our wheels further in the debate about extended cognition.

Acknowledgements

I am grateful to Frederick Adams, Kenneth Aizawa, Michael Goldsby, Larry Shapiro, Elliott Sober, Shannon Spaulding, Peter Vranas, Rob Wilson, and a careful anonymous reviewer for helpful comments.

References

Adams, F. & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology, 14,* 43-64.

Adams, F. & Aizawa K. (2008). *The Bounds of Cognition*. Malden, MA: Blackwell Publishers.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. Petrov & F. Csaki (Eds.), *Second international symposium on information theory (pp. 267-281)*. Budapest: Akademiai Kiado.

Devitt, M. (2008). Resurrecting biological essentialism. *Philosophy of Science, 75*, 344-382.

Clark, A. (2009). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind, 118,* 963-993.

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

Clark, A. (2007). Curing cognitive hiccups: A defense of the extended mind. *The Journal of Philosophy, 104*, 163-192.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. & Chalmers D. (1998). The extended mind. *Analysis, 58*, 7-19.

Dawkins, R. (1982). *The Extended Phenotype*. Oxford: Oxford University Press.

Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1-55.

Glenberg, A. & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558-565.

Goldin-Meadow, S. (2003). *Hearing Gesture: How our Hands Help Us Think*. Cambridge, MA: The Belknap Press of Harvard University Press.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens cognitive load. *Psychological Science, 12,* 516-522.

Fodor, J. (2009, 12th February). Where is My Mind? *London Review of Books*, *31*.

Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fodor, J. (1987), *Psychosemantics*. Cambridge, MA: MIT Press.

Forster, M. & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science, 45*, 1-36.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy, 71,* 5-19.

Hitchcock, C., & Sober, E. (2004). Prediction versus accommodation and the risk of overfitting. *British Journal for the Philosophy of Science, 55,* 1-34.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition (pp. 264-306)*. Cambridge: Cambridge University Press.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Minnesota Studies in the Philosophy of Science (pp. 410-505)*, vol. XIII. Minneapolis, MN: University of Minnesota Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: Chicago University Press.

Okasha, S. (2002). Darwinian metaphysics: Species and the question of essentialism. *Synthese, 131,* 191-213.

Reichenbach, H. (1938). *Experience and Prediction*. Chicago, IL: Chicago University Press.

Rowlands, M. (1999). *The Body In Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.

Royall, R. M. (1997). *Statistical Evidence: A Likelihood Paradigm*. Boca Raton, FL: Chapman and Hall.

Rupert, R. D. (2009). *Cognitive Systems and the Extended Mind*. New York: Oxford University Press.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy, 101,* 389-428.

Sawyer, K. R. & Greeno, J. G. (2009). Situativity and learning. In P. Robbins & M. Aydede (Eds.). *The Cambridge Handbook of Situated Cognition (pp. 347-367)*. Cambridge: Cambridge University Press.

Sidelle, A. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Ithaca, NY: Cornell University Press.

Sober, E. (2008). *Evidence and Evolution*. Cambridge: Cambridge University Press.

Sober, E. (2006). Parsimony. In S. Sarkar & J. Pfeifer (Eds.). *The Philosophy of Science: an Encyclopedia*. New York, NY: Routeledge.

Sober, E. (2003). Two uses of unification. In F. Stadler (Ed.). *The Vienna circle and logical empiricism: re-evaluation and future perspectives* (vol.10 of *Vienna Circle Institute Yearbook, pp. 205-216*). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Sober, E. (2002). What is the problem of simplicity? In H. Keuzenkamp, M. McAleer, & A. Zellner (Eds.). *Simplicity, inference, and modeling (pp. 13-32).* Cambridge: Cambridge University Press.

Sober, E. (1996). Parsimony and predictive equivalence. *Erkenntnis, 44,* 167-197.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295–312.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science,* 12, 257–285.

Wilson, R. A. (2008). Meaning making and the mind of the externalist. In R. Menary (Ed.). *The Extended Mind (pp. 167-188)*. Cambridge, MA: MIT Press.

Wilson, R. A. (2004). *Boundaries of The Mind*. Cambridge: Cambridge University Press.

Wilson, R. A. (1995). *Cartesian Psychology and Physical Minds*. Cambridge: Cambridge University Press.

Wilson, R. A., Barker, M. J. & Brigandt, I. (2007). When traditional essentialism fails: biological natural kinds. *Philosophical Topics, 35,* 189-215.

Wilson, R. A. & Clark, A. (2009). How to situate cognition: letting nature take its course. In P. Robbins & M. Aydede (Eds.). *The Cambridge handbook of situated cognition (pp. 53-77)*. Cambridge: Cambridge University Press.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.