Is it possible to experimentally determine the extension of cognition?

Michael Baumgartner · Wendy Wilutzky

Abstract

Various analytical tools originally developed for theories of mechanistic explanation have recently been imported into the ongoing debate on the hypothesis of (extracerebrally) extended cognition (HEC). One such tool that appears particularly relevant to that debate is Craver's mutual manipulability account of constitution (MM), most of all, because it promises to settle the debate on experimental grounds. This paper investigates whether it is possible to deliver on that promise. We first find that, far from grounding an experimental evaluation of HEC, MM is conceptually incompatible with both internalist and externalist accounts of cognition. Next, we propose a suitable modification of MM, *viz.* MM*, but it turns out that MM* presupposes rather than produces clarity on the extension of cognition. Moreover, subject to MM* the inference to constitution is radically empirically underdetermined. Finally, we argue that our results can be generalized and conclude that, for principled reasons, it is impossible to experimentally determine whether cognitive processes have extracerebral constituents. Determining the extension of cognition is an inherently pragmatic matter.

1 Introduction

For well over a decade, the *hypothesis of (extracerebrally) extended cognition* (HEC) has structured numerous debates in the philosophy of cognitive science. In a nutshell, HEC states that not all cognitive processes are located wholly in the brain (Clark and Chalmers 1998, 9), or in other words, that there exist cognitive processes that extend outside the brain (Drayson 2010, 367). HEC has its roots in empirical results from (neuro-)psychology indicating that it is advantageous to model certain cognitive processes in such a way that some of their relevant parts are situated in the extracerebral body (Kirsh and Maglio 1994; Ballard et al. 1995; Romo et al. 1998). The challenging contrast of these externalist conjectures to traditional internalist (i.e. cerebral) accounts of the cognitive quite naturally induced their absorption in philosophy. Some philosophers emphatically endorse the hypothesis of (partly) embodied cognition and do not hesitate to moreover liberate cognition from the confines of the human body (Clark and Chalmers 1998; Clark 2008; Rowlands 1999, 2009; Wheeler 2010). Others—with no less emphasis—argue that the proponents of HEC mistake causal coupling for constitutional integration, that is, commit the infamous coupling-constitution fallacy (Adams and Aizawa 2008, ch. 6). According to HEC's critics, cognitive processes are merely causally intertwined with extracerebral systems, but their actual constituents are fully located within the brain (Adams and Aizawa 2001, 2008; Rupert 2004, 2009).

Even though both camps tend to back up their positions with empirical findings, the core issues of the debate are of metaphysical or conceptual nature. In essence, the debate can be seen to turn on the question as to what is the adequate mark of the cognitive (Rowlands 2009). Advocates of HEC assume a coarse-grained functionalist account of the cognitive, subject to which the nature of cognition consists entirely in performing input-output functions as implied by a folk-psychological understanding of the mental, independently of the latter's physical implementation (Clark 2008). Their opponents, by contrast, stipulate that cognitive processes must be identified on the basis of their fine-grained functional roles as are investigated in science (Rupert 2004) or based on the intrinsic properties of their physical realization, for instance, their involvement of non-derived content (Adams and Aizawa 2001). Although both sides make notable efforts to grant each other as much of their respective metaphysical assumptions as possible, the debate has, from the outset, been stuck in an argumentative stalemate, in which each camp simply presumes a mark of the cognitive that the other ultimately rejects.

In light of this gridlock, several authors have recently begun to import analytical tools developed in the context of theories of mechanistic explanation into the debate on HEC (e.g. Theiner et al. 2010; Kaplan 2012; Kirchhoff 2014; Pöyhönen 2014). Roughly put, a mechanistic explanation accounts for a macro phenomenon in terms of the activities of its micro constituents (e.g. Machamer et al. 2000; Bechtel and Abrahamsen 2005). The core dependence relation exploited in such explanations is the relation of constitution. Therefore, theories of mechanistic explanation require a theory of constitution specifying, among other things, where causal coupling ends and constitutional integration begins. As this is the very issue at the heart of the debate on HEC, it has been contended that the import of the dominant theory of mechanistic constitution, *viz.* Craver's (2007) *mutual manipulability* theory (MM), is particularly apt to move the debate beyond its current standstill by turning the argumentative focus away from the mark of the cognitive to the *mark of constitution*. Accordingly, MM figures prominently in recent arguments in favor of HEC, the most elaborate of which has been presented by Kaplan (2012), but Zednik (2011) has made a similar proposal.

Assessing the extension of cognition against the background of MM comes with two attractive prospects. First, as MM has been developed in complete independence from the debates on HEC, both proponents and opponents of HEC should be able to view it as unbiased arbiter and accept its verdict on the boundary between causal coupling and constitutional integration. Second, MM induces a straightforward experimental protocol of constitutional discovery and, thus, promises to move the debate on HEC away from the question as to the proper mark of the cognitive, and to settle it experimentally.

This paper explores whether it is possible to deliver on that promise. We first find that MM indeed has far-reaching implications for HEC. However, instead of grounding an experimental evaluation of HEC it turns out that MM is incompatible with both internalist and externalist accounts of cognition on purely conceptual grounds. Subject to MM, cognitive processes are realized neither inside nor outside of the brain. These sweeping consequences can either be interpreted to reduce all non-eliminativist and non-dualist views on cognition to absurdity, or they can be taken to show that the constraints MM imposes on the notion of constitution are too strong. As the first option runs counter to the purposes of all participants to the HEC debate, we then continue to explore the second option by investigating how the project of experimentally determining the bounds of cognition fares relative to a suitable weakening of MM, *viz.* MM*. We find that, while MM* is compatible with both cerebral and extracerebral accounts of cognition, it presupposes rather than produces clarity about the extension of cognition and fails to furnish evidence-based inferential leverage on HEC. The experimental designs induced by MM* systematically underdetermine the inference to the bounds of cognition. Finally, we generalize our results and conclude that, while it is possible to conclusively establish experimentally that a particular physical process *does not* constitute a cognitive process, it is impossible to establish experimentally that a particular physical process *does* constitute a cognitive process. Delineating the bounds of cognition is an inherently pragmatic matter for which virtues as explanatory power, predictive strength, simplicity, or coherence must be called upon.

The paper has two main parts. The first one, sections 2 and 3, critically discusses Kaplan's (2012) and Zednik's (2011) MM-based argument in favor of HEC. As our criticism indicates that a precise understanding of the definitional details of MM and its theoretical embedding in Woodward's (2003) interventionist theory of causation is lacking in the literature on HEC, we will lay out all the relevant definitions and extract their pertinent consequences—some of which are widely discussed in the literature on causation. The second part, sections 4 and 5, explores alternative ways of experimentally determining the bounds of cognition and presents our underdetermination argument.

2 Mutual Manipulability (MM)

The question whether there exist cognitive processes that extend beyond the brain can be understood as asking whether there exist extracerebral processes that constitute cognitive processes, rather than merely causally interacting with them. According to the standard view, causation and constitution are two metaphysically distinct dependence relations. Causation holds among mereologically independent entities such that causes temporally precede their effects, and causal dependence is unidirectional in the sense that effects depend on their causes but not vice versa. By contrast, constitution holds among wholes and their parts, that is, among spatiotemporally overlapping entities, and it amounts to a bidirectional form of dependence in the sense that the parts depend on the wholes and vice versa (Craver and Bechtel 2007). Constitution is the core dependence relation exploited in *mechanistic explanations*, which account for a system's upper level (or macro) behavior, that is, for a *phenomenon*, in terms of the lower level (or micro) behaviors and activities of its constitutively relevant parts, that is, in terms of its *constituents* (Machamer et al. 2000; Bechtel and Abrahamsen 2005).

Before reviewing Craver's (2007) mutual manipulability theory of constitution (MM), we must render transparent one crucial background assumption of our argument and introduce our notation. The background assumption is often made in mechanistic theorizing; it states that the relation between a mechanism's upper and lower level is to be analyzed in terms of *non-reductive supervenience* (Glennan 1996, 61-62; Eronen 2011, ch. 11). More specifically, relative to a given mechanistic organization of the constituents, phenomena supervene on their constituents, meaning that every change in a phenomenon is necessarily accompanied by a change in its constituents (Craver 2007, 153). Moreover, phenomena are not reducible—in particular, not identical—to their constituents. While non-reducibility may be an open issue in certain special sciences (e.g. chemistry), the non-reducibility of the mental to the physical seems exceedingly plausible, both in philosophy and in science. As we will only be concerned with cognitive (i.e. mental) phenomena in this paper, we will assume that phenomena are non-reducible to their constituents.

Phenomena and their constituents are types of behavior exhibited by specific entities on upper and lower levels, respectively. We represent such behaviors by *specific variables* as introduced by Spohn (2006). More concretely, we use Ψ for the behavior of an upper level entity s, and Φ_1, Φ_2, etc. for the behaviors of lower level entities x_1, x_2, etc. That is, $\Phi_1 = \phi_i$ stands for x_1 exhibiting behavior ϕ_i—in Craver's jargon, for x_1's ϕ_i-ing.

According to MM, constitution is a difference-making relation that can be accounted for by supplementing the resources of the currently most popular difference-making theory of causation: Woodward's (2003) interventionism. In a nutshell, interventionism stipulates that a variable X is a cause of another variable Y iff it is possible to ideally intervene on X in such a way that Y changes when all causes of Y not located on a path through X, i.e. all *off-path* causes of Y, are held fixed (cf. Woodward 2003, 59). An *ideal intervention* on X with respect to Y is a variable $\mathcal{I}_X$ taking one of its values, $\mathcal{I}_X = i_n$, and thereby *surgically* fixing the value of X without having an impact on Y that is not mediated via X and without being correlated with any off-path causes of Y (cf. Woodward 2003, 98). As constitution, contrary to causation, is a bidirectional dependence relation among parts and wholes, unidirectional manipulability as in interventionism does not suffice to establish constitutive relevance. Therefore, Craver (2007, 159) adds a parthood and a mutuality constraint: constituents are spatiotemporal parts of phenomena and both are mutual difference-makers of each other. More formally:[1]

[1] Note that Kaplan (2012, 560) misreads MM as only providing a sufficient condition for constitutive relevance. Textual evidence, however, contradicts that assessment. For instance, on p. 159 of (2007), Craver presents mutual manipulability as a sufficient condition for constitutive relevance and the absence of mutual manipulability as a sufficient condition for constitutive irrelevance, which entails that mutual manipulability is sufficient and necessary for constitutive relevance. If MM were to merely provide a sufficient condition, it would be strongly biased in favor of HEC's proponents.

(MM) Φ is constitutively relevant to Ψ iff (i) the instances of Φ are spatiotemporal parts of instances of Ψ; (ii) there exists a possible ideal intervention $\mathcal{I}_\Phi = i_m$ on Φ w.r.t. Ψ that is associated with a change in Ψ; and (iii) there exists a possible ideal intervention $\mathcal{I}_\Psi = i_n$ on Ψ w.r.t. Φ that is associated with a change in Φ.

One of the main selling points of MM is that it entails an experimental protocol for constitutional discovery. According to MM, constitutive relations can be established by performing interventions on a phenomenon w.r.t. its parts, so-called *top-down* interventions, and interventions on the parts w.r.t. the phenomenon, *bottom-up* interventions. If such tests reveal mutual difference-making, the parts are *experimentally proven* to be constituents of the phenomenon. By contrast, if mutual difference-making cannot be established along its interventionist protocol—in a representative series of experiments—, MM warrants an inductive inference to the absence of a constitutive relation. The relevance of such an account for the gridlocked debate on HEC is obvious: it yields the design for an *experimentum crucis* determining whether an extracerebral process is a constituent of a cognitive process or merely causally coupled with it. In that light, disputes over the proper mark of the cognitive, which currently paralyze the debate on HEC, appear otiose.

3 Applying MM to Cognition

Since HEC was first formulated, both its proponents and opponents have regularly backed up their arguments with suitably interpreted results of experimental studies. However, as a clear criterion determining whether experimental findings are indicative of constitution or mere causation has been missing from the debate, both camps could easily interpret relevant studies in ways favorable to their positions. The import of MM therefore promises to finally provide a criterion that regulates the proper interpretation of experimental results. Correspondingly, authors that have recently argued in favor of HEC by drawing on MM, have eagerly reconstructed pertinent studies against the background of MM. Kaplan (2012, 562-564), for instance, applies MM to the studies of Feldman and Levin (1995) and Ballard et al. (1995), or Zednik (2011, 259-260) reconstructs the study by Beer (2003) using MM.

This section explores whether such applications of MM to experimental results can indeed settle the debate on HEC. As our example we choose the study by Ballard et al. (1995), which is discussed both by proponents (e.g. Clark 2008, 11-13) and opponents of HEC (e.g. Rupert 2009, §5.3.1). The authors of that study investigate the question how humans can perform memory-demanding tasks despite their severe limitation in short-term memory capacity. To that end, subjects' eye

Based on a merely sufficient condition supporters of HEC could license an inference to constitution whenever extracerebral processes and cognitive phenomena happen to be mutually manipulable. Yet, based on a merely sufficient condition, critics of HEC could never build a case against cognition extending beyond the brain, for failures of mutual manipulability could not be taken to falsify HEC.

movements were recorded while they performed the task of copying a pattern of colored blocks depicted in a 'model' area of a computer screen to a 'workspace' area by drawing from a stack of colored blocks in a 'resource' area. In a first experiment, "[s]ubjects were instructed simply to copy the model pattern, 'as quickly and accurately as possible,' using the mouse to move the blocks" (Ballard et al. 1995, 68). According to a traditional information-processing account of cognition, the task of copying a single block in the pattern is expected to be solved by, first, memorizing both the position and color of the block in the model, second, selecting an appropriate block from the resource, and third, placing it at the proper location in the workspace. Thus, this account predicts a total of two saccades between the areas: one from the 'model' to the 'resource' area and one from the 'resource' to the 'workspace' area. However, Ballard et al. (1995) found that subjects, for the copying of most blocks, perform at least two saccades between the 'model' and 'resource' areas already, presumably to match the color of blocks and determine their location in the pattern sequentially instead of simultaneously. Numerous further experiments were conducted, in which the experimenters, for example, increased the pattern complexity, removed the model from view after variable durations, instructed the subjects to fixate their gaze at the center of the screen, or increased the distance between the 'model' and 'resource' areas. With each of these added degrees of difficulty subjects' success rate in copying the pattern accurately decreased and the time for task completion increased. From their findings, Ballard et al. (1995, 71) conclude "that eye movements are an integral part of the economical execution of the task", hence implying that the saccades are not causally conducive to but in fact constituents of the cognitive process of block matching.

On the face of it, this conclusion seems straightforwardly supported by MM. The experimental manipulations of Ballard et al. (1995) appear to amount to top-down and bottom-up interventions revealing mutual difference-making: engaging subjects in block matching to a top-down intervention associated with changes in the saccades, and fixating subjects' gaze to a bottom-up intervention associated with changes in block matching. By virtue of MM, these results entail that the saccades are constitutively relevant to block matching, which establishes the existence of extracerebral cognition on empirical grounds (Kaplan 2012, 564).

Before evaluating whether that argument holds up to scrutiny, a preliminary qualification is called for. Strictly speaking, performing one successful top-down and one bottom-up intervention is not sufficient for MM to entail that the saccades are constituents of block matching. The reason is that MM, in condition (i), additionally requires that the mutually manipulated behaviors stand in a spatiotemporal parthood relation. Obviously though, this is a controversial issue in the debate on HEC. Thus, if we were to stick to the wording of MM, it is clear that Kaplan's (2012) and Zednik's (2011) project of settling that debate on the basis of MM is a nonstarter because MM's application *presupposes* rather than produces clarity on the mereological relationship between cognitive and extracerebral processes.[2] However, both proponents and opponents of HEC agree that, in order to steer clear of question-begging argumentative circles, the bounds of cognition must not be drawn with recourse to spatiotemporal criteria (e.g. Clark and Chalmers 1998, 8) and, in particular, that cognition must not be *defined* as occurring within the brain or any other region (e.g. Adams and Aizawa 2001, 46). Both sides, hence, concur that the entities involved in cognitive processes *possibly* are extended systems, comprising cerebral as well as extracerebral elements. The disagreement concerns the question whether cognitive systems are not only possibly but *actually* extended: supporters of HEC endorse actual extension, whereas critics do not. In other words, a central driving force of the debate stems from dissent about adequate *non-spatiotemporal* criteria for delineating the boundaries of cognition. In this light, Kaplan (2012) and Zednik (2011) must be understood as proposing to evaluate HEC on the basis of MM's interventionist conditions (ii) and (iii) alone. Accordingly, we will subsequently bracket MM's parthood requirement and investigate whether mutual manipulability as expressed in conditions (ii) and (iii) of MM can serve as a non-spatiotemporal criterion identifying the constituents of cognitive processes.

To help evaluate whether an application of MM—more precisely, of conditions (ii) and (iii)—to the results of Ballard et al. (1995) establishes the existence of extracerebral cognition, figure 1a provides a schematic representation of the mechanism in question. The variable Ψ in the upper level ellipse exhibits the cognitive phenomenon of block matching, which is hypothesized to be constituted on the lower level by a causal feedback among, on the one hand, neural processes realizing various information storage and comparison functions in the brain, and, on the other, eye movements that feed their sensory input to the brain and are repeated until a match is established. To keep things as simple as possible, we represent the cerebral processes by the single variable Φ_1 within the grey shading and the extracerebral saccades by Φ_2 outside of the grey area.

Now, let us assume the proponents of HEC are right that Ψ is not only constituted by Φ_1, but also by Φ_2. Subject to MM, this entails, among other things, that there exists an intervention variable $\mathcal{I}_\Psi$ on Ψ w.r.t. Φ_2 as depicted in figure 1a, such that intervening on Ψ via $\mathcal{I}_\Psi$ is associated with changes in Φ_2. Woodward's interventionism, which constitutes the theoretical background of MM, implies that an intervention $\mathcal{I}_\Psi$ on Ψ w.r.t. Φ_2 that is associated with changes in both Ψ and Φ_2 (when all off-path causes of Φ_2 are fixed) is a cause of both Ψ and Φ_2 (Woodward 2003, 59). This can be structurally realized in one of two ways: either $\mathcal{I}_\Psi$ causes Ψ and Φ_2 along *one* causal path, e.g. $\mathcal{I}_\Psi \longrightarrow \Psi \longrightarrow \Phi_2$, or along *two*

[2] Kaplan (2012) interprets MM as a criterion for demarcating the boundaries of mechanisms. We believe that this is a misinterpretation due to the fact that he disregards MM's parthood condition. The application of MM presupposes that the spatiotemporal boundaries of analyzed mechanisms are given by the definition of the scrutinized phenomenon. Everything occurring within those boundaries then counts as a part of the phenomenon, and if such a part satisfies MM's interventionist criteria, it additionally counts as a constituent. That is, MM is not designed for mechanism demarcation but for identifying constituents, *given* that it is clear what the boundaries of the relevant mechanism are.

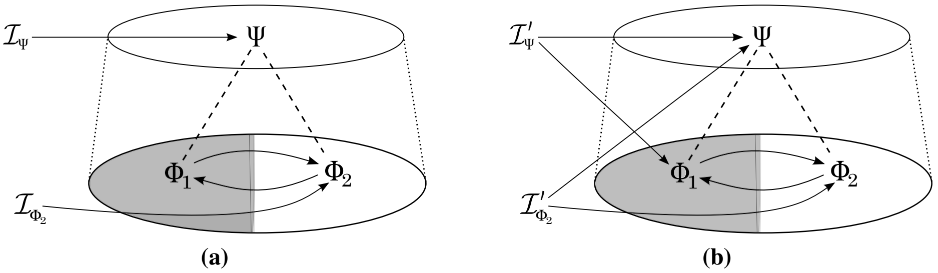


Figure 1: Feedback mechanism between neural processes (Φ_1) in the brain (grey shading) and saccades (Φ_2) outside the brain hypothesized to constitute the cognitive process of matching colored blocks (Ψ). Dashed lines represent constitution, directed edges symbolize causation, and the dotted lines stand for spatiotemporal overlap. Model (a) depicts the impossible surgical interventions required by MM; model (b) features the possible fat-handed interventions.

paths, $\Psi \longleftarrow \mathcal{I}_\Psi \longrightarrow \Phi_2$. The former option is excluded by the fact that Ψ and Φ_2 represent mereologically dependent behaviors and their relationship, hence, is non-causal (which is violated in the one-path option). In light of the non-identity of phenomena and their parts and the standard definition of (directed) causal paths in terms of ordered n-tuples of variables, it follows that $\mathcal{I}_\Psi$ causes Ψ and Φ_2 along two different paths, *viz.* $\langle \mathcal{I}_\Psi, \Psi \rangle$ and $\langle \mathcal{I}_\Psi, \Phi_2 \rangle$ with $\Psi \neq \Phi_2$, meaning that $\mathcal{I}_\Psi$ is a common cause of Ψ and Φ_2. However, that, in turn, entails that $\mathcal{I}_\Psi$ does not surgically cause Ψ and is, therefore, not an intervention variable for Ψ w.r.t. Φ_2 after all. Hence, contrary to first appearances, engaging subjects in the matching task does not amount to a top-down intervention; rather it causes the cognitive phenomenon on a first path and the neural activity with ensuing saccades on a second path. In other words, engaging subjects in matching blocks does not have the structural features of $\mathcal{I}_\Psi$ in figure 1a but those of $\mathcal{I}'_\Psi$ in figure 1b.

Likewise, fixating subjects' gaze is no bottom-up intervention in the vein of $\mathcal{I}_{\Phi_2}$ of figure 1a. As gaze fixating is associated with changes in saccades as well as block matching, it is a cause not only of Φ_2 but also of Ψ. Yet, if Φ_2 and Ψ are mereologically dependent, they are causally unrelated. Against the background of the non-identity of Φ_2 and Ψ and the standard definition of a causal path, it follows that fixating subjects' gaze is a common cause of the changes in saccades and block matching. That is, it has the structural features of $\mathcal{I}'_{\Phi_2}$ in figure 1b. In sum, if the saccades do in fact constitute the matching process, the experiments conducted by Ballard et al. (1995) do not amount to surgical interventions as required by MM; rather, they are so-called *fat-handed* manipulations, that is, manipulations influencing their effects along two different causal paths. A fortiori, subject to MM, these experiments do not establish that the saccades constitute the matching of colored blocks, i.e. that there exist cases of embodied cognition.

By contrast, advocates of a traditional cerebral account of cognition seem to have no trouble to model the relationship between the matching process Ψ and the saccades Φ_2 in entirely causal—i.e. non-constitutional—terms. They analyze Φ_2

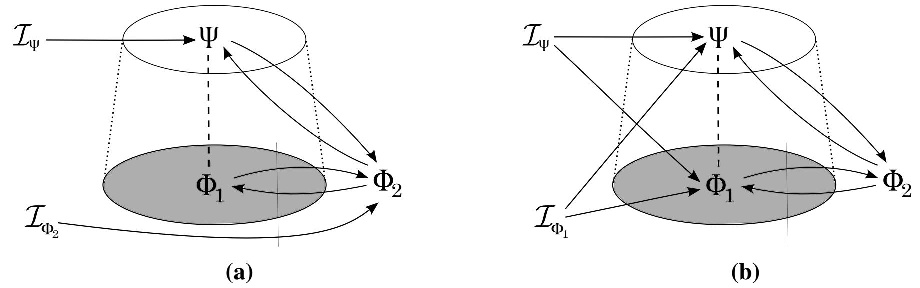


Figure 2: Model (a) accounts for the relationship between block matching (Ψ), cerebrally constituted by Φ_1, and the extracerebral saccades (Φ_2) in terms of causal feedbacks. Model (b) depicts the possible fat-handed interventions on Ψ and Φ_1.

as mereologically non-overlapping with Ψ, which is only constituted by Φ_1. Ψ is hence not excluded to be causally related to Φ_2. Therefore, the fact that engaging subjects in block matching as well as fixating their gaze cause changes in both Ψ and Φ_2 does not entail that these experimental manipulations are fat-handed. Instead, they can be modeled as surgical top-down and bottom-up interventions that cause Ψ and Φ_2 in a chainlike manner along *one single* causal path. The overall result is a causal coupling of Ψ and Φ_2 in terms of a feedback structure as depicted in figure 2a. Thus, it appears that the controversy over the extension of cognition can indeed be resolved on the basis of MM, albeit in the opposite direction of the one envisaged by Kaplan (2012) and Zednik (2011): cognitive processes are *not* constituted by extracerebral processes but merely interact with them causally.

Yet, drawing the conclusion that MM supports internalist accounts would be hasty, for the above argument can be repeated for the hypothesis of internalist cognition. Against the background of a purely cerebral model of block matching, as in figure 2a, engaging subjects in the task does not amount to an intervention $\mathcal{I}_\Psi$ on Ψ w.r.t. the neural processes Φ_1 either, for it is associated with changes in the two causally unrelated and non-identical variables Ψ and Φ_1 and, thus, is a common cause of Ψ and Φ_1 rather than a surgical top-down intervention. The same holds for manipulations $\mathcal{I}_{\Phi_1}$ of the neural processes Φ_1 that are associated with changes in block matching Ψ: they are common causes of Φ_1 and Ψ and not surgical bottom-up interventions. Figure 2b provides an illustration. In sum, the claim that block matching is entirely constituted by neural processes does not receive any support from MM either.

The argument as to the inapplicability of MM to the experiments of Ballard et al. (1995) has traction beyond that study. In fact, the argument can be generalized to the point where it not only reveals that MM does *not support* externalist or internalist accounts of cognition but where MM is shown to entail that cognitive processes are *not constituted* in the brain or outside thereof. To see this, note that, as the constituents of a cognitive phenomenon form the latter's supervenience base, every change in the phenomenon is *necessarily* accompanied by a change in at least one of its constituents. Subject to the non-causal nature of the relationship between cognitive phenomena and their constituents and subject to the non-identity of the former and the latter, it follows that every cause of a cognitive phenomenon necessarily is a common cause of that phenomenon and at least one of its constituents. Hence, (ideal) top-down interventions on cognitive mechanisms are downright *impossible* (cf. Baumgartner and Gebharter 2015). This yields that MM, according to which the *possibility* of such interventions is necessary for constitution, has the sweeping consequence that cognitive phenomena are not constituted by any of their parts. Applied to the context of the debate on HEC that means that MM determines both externalist and internalist accounts of cognition to be false.

Two aspects of this result deserve separate emphasis. First, it might be objected that these sweeping implications are a mere modeling artifact due to the fact that we applied MM to models—as the ones in figures 1 and 2—featuring both causally and constitutively related variables, which may be considered illegitimate (e.g. Eronen 2012; Yang 2013). However, that cognitive phenomena and their constituents can only be manipulated with a fat hand is not a consequence of their integration in the same models but stems from the fact that they are assumed to be non-identical *in the world*. Irrespective of whether a corresponding model contains both causally and constitutively related variables, interventions on a cognitive phenomenon are connected (in the world) to the latter's constituents on paths that do not go through the phenomenon itself. Mechanistic systems can only be manipulated on all of their levels at the same time, and, if these levels are seen as non-reductively supervening on one another, which seems exceedingly plausible in the case of the mental, the relevant interventions turn out to cause all levels on different routes in the world, not merely in some ill-defined model—in violation of the requirements MM imposes on constitution.

Second, these implications of MM do not hinge on experimental evidence, rather they are of purely conceptual (*a priori*) nature. MM—with its theoretical embedding in Woodward's (2003) interventionism—in combination with the commonly accepted principles that constitution is a non-causal form of dependence and that cognitive phenomena and their physical realizers are non-identical deductively reduces both cerebral and extracerebral accounts of cognition to absurdity (prior to all experiments). The only theories of cognition that receive support from MM are theories not conceiving of the relation between the cognitive and the physical in terms of constitution—for instance, dualism, in virtue of which the cognitive and the physical are different realms comprising ontologically independent phenomena, or eliminativism, subject to which only one of the two realms is real. It goes without saying that these are not the customers Kaplan and Zednik have in mind for their proposal of how to resolve the HEC debate.

In light of these far-reaching metaphysical consequences of MM, the argument presented in this section can also be seen to reduce MM itself to absurdity, rather than certain accounts of cognition. After all, a theory of constitution that supports dualism or eliminativism on mere conceptual grounds can justifiably be argued to build more into the notion of constitution than is really there. Therefore, before returning a definite verdict on the prospects of experimentally settling the debate on HEC by importing MM, the next section is going to propose a modification of MM, *viz.* MM*, which does not have any *a priori* ramifications for theories of cognition.

4 Modifying MM

MM ties constitutive relevance to the possibility of surgical top-down and bottom-up interventions that target *one* level of a mechanism and thereby change the other level. However, the previous section has shown that, as constitution is a non-causal form of dependence, it is only possible to induce changes in non-identical variables on upper and lower levels of a mechanism by targeting *both* levels at the same time, on separate causal paths. Contrapositively put, whenever surgical interventions that target a first variable and induce changes in a second one are possible, these variables are not linked in terms of constitution but in terms of causation—as is duly entailed by the interventionist theory of causation (Woodward 2003). An obvious conclusion to draw is that the definitional line taken by MM is misguided. Constitution should not be analyzed in terms of surgical (or ideal) mutual manipulability, rather it must be cashed out in terms of *non-surgical* interventions of some sort.

Indeed, prompted by problems of the original version of interventionism with macro-to-micro causation, Woodward (2015) has recently offered a modified variant of his theory, *viz. interventionism**, which weakens his original notion of an ideal intervention by introducing exemption clauses for supervenience relations. While he required that an intervention targets exactly one variable in Woodward (2003, 98), he now (Woodward 2015) allows for multiple targets, provided that these targets are related in terms of supervenience. More concretely, Woodward (2015, 333-334) newly defines an intervention on X w.r.t. Y to be a variable $\mathcal{I}_X$ taking one of its values, $\mathcal{I}_X = i_n$, and thereby fixing the value of X without having an impact on Y that is not mediated via X *or via a variable* Z*, which is related in terms of supervenience to* X *or* Y, and without being correlated with any off-path cause Z of Y *such that* Z *is not related in terms of supervenience to* X *or* Y. Against that background, $\mathcal{I}_\Psi$ can pass as an intervention variable for a macro variable Ψ w.r.t. to one of its constituents Φ even if $\mathcal{I}_\Psi$ causes Φ along a path that does not go through Ψ but through variables that are related in terms of supervenience to Ψ—which is (trivially) satisfied if Φ itself is part of the supervenience base of Ψ. Woodward contends "that an intervention on a macro variable Ψ also should be treated as automatically changing (indeed as also an intervention on) the supervenience base SB(Ψ) of Ψ" (2015, 333, adjusted to our symbolism). Thus, if phenomena and their constituents are related by supervenience, as is usually assumed in the debate on HEC, interventions that target both the former and the latter on different causal paths (because phenomena are not identical to their constituents) count as interventions by the standards of interventionism*, notwithstanding the fact that they are non-surgical but fat-handed. For brevity, we will speak of *permissibly fat-handed interventions*.[3]

It can easily be seen that the interventions $\mathcal{I}'_\Psi$ and $\mathcal{I}'_{\Phi_2}$ in figure 1b and $\mathcal{I}_\Psi$ and $\mathcal{I}_{\Phi_1}$ in figure 2b are permissibly fat-handed in that sense. In fact, our results of the previous section entail that all interventions that induce changes on multiple levels of a mechanism necessarily are of the permissibly fat-handed type, because different levels of mechanisms are related by supervenience. In this light, the following is a modification of MM that comes readily to mind:

(MM***)** Φ is constitutively relevant to Ψ iff (i) the instances of Φ are spatiotemporal parts of instances of Ψ; (ii) there exists a possible permissibly fat-handed intervention $\mathcal{I}_\Phi = i_m$ on Φ w.r.t. Ψ that is associated with a change in Ψ; and (iii) there exists a possible permissibly fat-handed intervention $\mathcal{I}_\Psi = i_n$ on Ψ w.r.t. Φ that is associated with a change in Φ.

Replacing surgicality by permissible fat-handedness along the lines of MM* allows for reconstructing the experimental manipulations of Ballard et al. (1995) as interventions that are conducive to the identification of constitutive relations. Contrary to MM, MM* is thus applicable to the phenomenon investigated by Ballard et al. (1995)—and in an analogous manner to cognitive processes in general. Accordingly, this modification of MM does not give rise to the reductio argument of the previous section. It has no *a priori* implications for the debate on HEC.

The next question then becomes whether MM* can help to resolve the debate on experimental grounds. On the face of it, this indeed seems to be the case. From the perspective of the advocates of HEC, telling subjects to match colored blocks or fixating subjects' gaze amount to MM*-conformant interventions, which are moreover associated with changes in both the upper and the lower level. Based on these results, MM* entails that the saccades are constituents of block matching. This arguably establishes the existence of a cognitive process that has at least one extracerebral constituent and, hence, validates HEC.

Nevertheless, opponents of HEC will hardly be convinced. First of all, it should be emphasized again that in addition to mutual difference-making, MM*—just as MM—requires spatiotemporal parthood in condition (i), which is controversial in the debate on HEC. Hence, as in the case of MM, applying MM* to resolve the debate without begging the question requires bracketing condition (i) and implementing conditions (ii) and (iii) as non-spatiotemporal criteria for identifying the constituents of cognitive processes. Yet contrary to the case of MM, conditions (ii) and (iii) of MM* are not without presuppositions as regards the bounds of cognition, for whether an intervention satisfies these conditions hinges on whether it is of the permissibly fat-handed type, which depends on whether its target variables are related in terms of supervenience, which, in turn, hinges on where exactly the

[3] While the weakened notion of an intervention discussed in this paragraph is Woodward's, not ours, referring to such interventions as *permissibly fat-handed* is our terminology, not Woodward's. Woodward (2015, 334), instead, speaks of *IV***-interventions.*

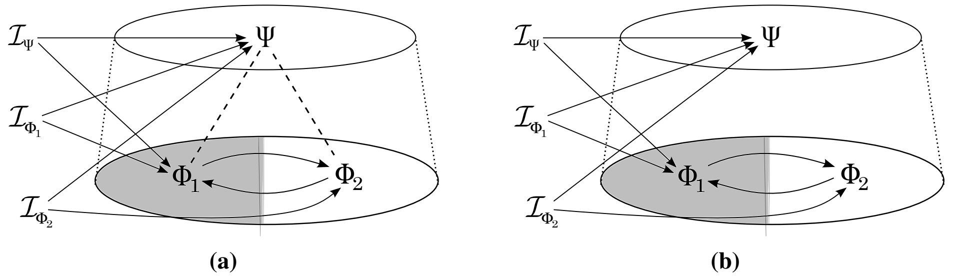
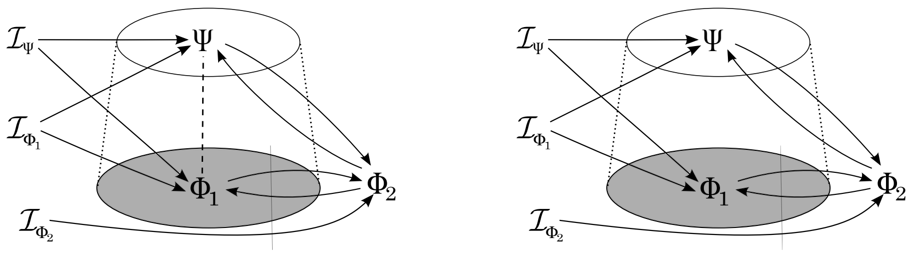


Figure 3: Two models that are experimentally indistinguishable from the perspective of a proponent of HEC. Model (a) features constitutive dependencies, model (b) does not.

boundary of the supervenience base of a scrutinized cognitive process is drawn. To see this, note that opponents of HEC deny that the saccades are contained in the supervenience base of block matching, because, for them, that cognitive process supervenes entirely on neural activation in the brain. In consequence, they deny that fixating subjects' gaze amounts to a permissibly fat-handed intervention; rather, they view it as surgical intervention that first induces a change in the saccades, which then causes a change in task performance. Moreover, they take this to establish that the saccades are causally and not constitutively linked to block matching. But obviously, this line of argument—just as the one of the proponents of HEC—presupposes a specific stance on where to draw the boundary of a cognitive process, which is exactly the matter at issue.

In sum, whether or not MM* entails that the saccades are constituents of block matching depends on whether interventions on the saccades are viewed as permissibly fat-handed or surgical, which, in turn, depends on whether the saccades are taken to belong to the supervenience base of block matching, and this, evidently, depends on where the boundaries of the mechanism in question are drawn. Yet, the latter is the very question at the heart of the HEC debate. Therefore, applying MM* to cognitive processes *presupposes* clarity on their boundaries and cannot be used to *produce* such clarity. In other words, implementing MM* for the purpose of resolving the debate on HEC is inherently question-begging.

There is yet another reason why MM* does not serve its intended purpose: MM* does not furnish an experimental method for identifying constitution. To see this, consider the extracerebral model of the block matching mechanism in figure 3a. According to MM*, Φ_1 and Φ_2 can be identified as constituents of Ψ by means of three permissibly fat-handed interventions as $\mathcal{I}_{\Psi}$, $\mathcal{I}_{\Phi_1}$, and $\mathcal{I}_{\Phi_2}$, which mutually make a difference on upper and lower levels. But data produced by common causes of two target variables are uninformative as regards the relationship between these variables. If $\mathcal{I}_{\Psi}$, $\mathcal{I}_{\Phi_1}$, or $\mathcal{I}_{\Phi_2}$ deliver correlations among their target variables, these correlations can be fully accounted for by the mere fact that $\mathcal{I}_{\Psi}$, $\mathcal{I}_{\Phi_1}$, and $\mathcal{I}_{\Phi_2}$ are common causes of the changes on upper and lower levels. It follows that there is no need to introduce constitutive dependencies in order to account for data generated by permissibly fat-handed interventions. Put differently, model 3a, which



(a) **(b)**

Figure 4: Two models that are experimentally indistinguishable from the perspective of an opponent of HEC. Model (a) features constitutive dependencies, model (b) does not.

features constitutive dependencies, and model 3b, which does not, imply the very same correlations under manipulations. They are experimentally indistinguishable. Nonetheless, the two models provide very different accounts of block matching. Subject to model 3a, the neural activation in the brain and the saccades physically realize block matching, whereas model 3b states that upper and lower levels are populated by ontologically independent entities that behave in a highly correlated manner simply because they are systematically coupled via common causes.

This problem does not only affect the proponents of HEC that would like to receive experimental support from MM* but also the hypothesis' opponents. If permissibly fat-handed interventions on block matching and neural activation give rise to correlations between these variables, it is indeterminate whether these correlations are due to constitution or mere common-cause coupling. Hence, as alternative to the constitutional model in figure 4a, which is the preferred model of advocates of cerebral cognition, there exists an experimentally indistinguishable but ontologically very different common-cause model, *viz.* the one in figure 4b. That is, MM* systematically underdetermines the inference to both cerebral and extracerebral constituents of cognitive processes.

As nothing in the above reasoning hinges on the details of the block matching mechanism, its conclusion can be generalized beyond that example: giving up surgicality along the lines of MM* entails that the mutual manipulability of upper and lower levels can be accounted for by the mere fat-handed nature of corresponding manipulations. Mutual manipulability via common-cause interventions provides no evidence in favor of the existence of constitutive dependencies. To every model featuring constitutive dependencies there exists an experimentally equivalent pure common-cause model. As to MM*, the inference to constitution is systematically underdetermined by experimental evidence.[4]

[4] Note that the claim here is not that causation and constitution can never be distinguished on evidence-based grounds. There are many causal structures that can be empirically distinguished from constitutive ones. For instance, in a causal chain $A \longrightarrow B \longrightarrow C$, the instances of A and C are entailed not to spatiotemporally overlap; A can therefore not be a constituent of C. Rather, the claim is that for every constitutive structure there exists one particular type of causal structure, *viz.* a common-cause structure, that is empirically indistinguishable from it. More specifically, whenever

Overall, the project of experimentally resolving the debate on HEC by implementing Craver's mutual manipulability framework as neutral arbiter leads into a dilemma: either (I) mutual manipulability is defined in terms of surgical interventions, which generate unconfounded data that can be unequivocally modeled but which cannot possibly be performed on two different levels of a cognitive mechanism, or (II) mutual manipulability is cashed out in terms of permissibly fat-handed interventions, which can be performed on two different levels of a cognitive mechanism but which beg the question HEC addresses and generate confounded data that cannot be modeled unequivocally. Horn (I) avoids question-begging and systematic empirical underdetermination but results in a theory of constitution that is incompatible with both externalist and internalist accounts of cognition and, hence, is not metaphysically neutral; horn (II), by contrast, has no metaphysical implications but presupposes rather than provides clarity on the bounds of cognition and comes with systematic empirical underdetermination.

Importing the mutual manipulability theory of mechanistic constitution into the debate on HEC requires that a choice be made. If horn (I) is chosen, the debate is successfully resolved. However, that resolution is of purely conceptual nature and favors neither side of the debate; instead, it supports dualist or eliminativist accounts of cognition. Thus, (I) clears the argumentative gridlock by conceptually strengthening alternative theoretical contenders. By contrast, if horn (II) is chosen the focus of the debate is successfully moved from the metaphysical question as to the proper mark of the cognitive to data-driven constitutional discovery, but the gridlock is not cleared. Rather, both proponents and opponents of HEC can implement the mutual manipulability framework, in its MM* variant, to interpret the experimental results of pertinent studies in ways that are favorable to their respective positions.

5 Outlook

Our findings so far suggest that the project of experimentally resolving the HEC debate by calling on Craver's mutual manipulability framework as neutral arbiter is bound to fail. Does that mean that the question as to the extension of cognitive processes generally cannot be answered experimentally, or do our results simply exhibit that the mutual manipulability framework is unsuited for that purpose—but maybe an alternative theoretical background could fill the bill?

We cannot exhaustively answer that question in the remainder of this paper, but still want to offer some tentative considerations. To get a sense of the direction in which to search for a viable alternative theory of constitution, it will be helpful to pinpoint the ultimate source of the deficiency of the mutual manipulability theory. Its underlying idea is that constitution is a difference-making relation that can be analyzed by supplementing the resources of the most popular difference-

A can be modeled as constituent of C, A and C can also be modeled as parallel effects of common causes, e.g. $A \longleftarrow B \longrightarrow C$.

making theory of causation, Woodward's (2003) interventionism, by a parthood and a mutuality tweak. At the same time, however, the theory is meant to render constitution as decidedly non-causal form of dependence. This creates a tension at its very heart, which becomes particularly virulent when MM or MM* are imported into the debate on HEC. The experimental protocol entailed by the mutual manipulability framework is not capable of exhibiting the difference between constitution and causation, but an evidence-based resolution of that debate demands a method that exhibits this very difference.

The ontological differences between causation and constitution—mereological independence vs. dependence, unidirectionality vs. bidirectionality—create an important difference as regards the experimental discoverability of these relations. Since causes and effects are mereologically independent and only unidirectionally related, it is possible to surgically intervene on causes w.r.t. effects, to break causal interactions via suitable interventions, and to isolate cause-effect pairs from confounding background influences. As a result, there exist ideal discovery circumstances in which crucial experiments can be conducted that produce unconfounded data affording conclusive evidence for causal dependencies. By contrast, such ideal discovery circumstances do not exist for constitutive relations among mutually non-reducible entities and activities on different levels. As the constituents of a phenomenon realize the latter on a lower level, manipulating the phenomenon is always tantamount to manipulating the constituents. It is impossible to surgically intervene on phenomena, to break constitutive dependencies, and to isolate single phenomenon-constituent pairs. If phenomena and their constituents are assumed to be non-identical—as is standard in the case of cognitive mechanisms—they can only be manipulated with a fat hand, meaning it is impossible to produce unconfounded data that could furnish conclusive evidence for constitution. To every model featuring constitutive dependencies there exists an empirically equivalent model without such dependencies. There cannot exist an *experimentum crucis* for constitution. Therefore, mutual difference-making—whether expressed along the lines of MM, MM* or of any other variant of the mutual manipulability scheme—is unsuited as identifying criterion for constitution.

Instead of mutual difference-making, we submit, based on the above considerations, that a characteristic feature of constitution—among possibly others—is that it relates upper and lower levels of mechanisms in such a way that they can only be manipulated with a fat hand. In other words, a mark indicating that a set $\mathbf{\Phi} = \{\Phi_1, \ldots, \Phi_n\}$ of spatiotemporal parts of a phenomenon Ψ—where Ψ is non-reducible to $\mathbf{\Phi}$—comprises the constituents of Ψ is that $\mathbf{\Phi}$ and Ψ are *systematically coupled via common causes* (cf. Baumgartner and Casini ming). That means, more concretely, (i) every cause of Ψ is a common cause of Ψ and at least one $\Phi_i \in \mathbf{\Phi}$, and (ii) every cause of $\Phi_i \in \mathbf{\Phi}$ that is associated with a change in Ψ is a common cause of Φ_i and Ψ. Or differently, it is impossible to surgically induce a change in Ψ without inducing a change in at least one $\Phi_i \in \mathbf{\Phi}$ on another causal path, and it is impossible to surgically induce a change in $\Phi_i \in \mathbf{\Phi}$ that is associated with a change in Ψ.[5]

Testing for systematic common-cause coupling requires considerably more intricate test designs than testing for mutual manipulability. The latter is an existentially defined criterion, meaning that it accounts for constitution in terms of the *existence* of a suitable top-down and bottom-up intervention each. If constitution could indeed be adequately accounted for along the lines of MM, one successful top-down and one bottom-up experiment would be sufficient for an inference to constitution. As shown in the previous section, however, a pair of successful mutual manipulations is far from warranting such an inference.[6] The criterion of common-cause coupling, by contrast, is of *universal* logical form. No (finite) test series can ever conclusively establish its satisfaction; rather, the criterion's universal quantifiers can only be inductively corroborated. To this end, a whole battery of experiments are required that explore the whole space of possible ways to manipulate a mechanism's upper and lower levels. Only if all of these experimental manipulations turn out to be fat-handed (i.e. non-surgical), an inductive inference to systematic common-cause coupling is warranted.

In this light, a handful of suitable experimental manipulations are not sufficient but only necessary for an inference to constitution. Thus, data generated by such manipulations can, at best, *empirically falsify* the hypothesis that a particular extracerebral process Φ_i is a constituent of a cognitive process Ψ. This can be accomplished by producing data showing that Φ_i and Ψ can be surgically manipulated independently of one another or by showing that there are causes of both Φ_i and Ψ that are not common causes of Φ_i and Ψ. But, of course, HEC does not claim that a specific extracerebral process constitutes a specific cognitive process, rather, HEC makes an unspecific existential claim: there exists at least one cognitive process that has an extracerebral constituent. Empirical falsifiability of specific constitutional claims is not directly conducive to the evaluation of such an existential claim.

What is more, if an extended test series inductively corroborates the systematic common-cause coupling of a phenomenon Ψ and a set $\mathbf{\Phi}$ of its spatiotemporal parts, it does not follow on evidence-based grounds that the elements of $\mathbf{\Phi}$ are constituents of Ψ. The reason is that—as shown above—the common-cause coupling of Ψ and $\mathbf{\Phi}$ can be equally accounted for by a constitutional and a pure common-cause model. To every constitutional model there exists an experimentally equivalent common-cause model.

As it is impossible, in principle, to generate unconfounded—i.e. conclusive—experimental evidence for constitution, the boundary between causes and con-

[5] As is common in scientific modeling, we assume here that modeled variable sets do not contain variables that are logically or conceptually dependent on one another.

[6] Correspondingly, in scientific practice, numerous top-down and bottom-up experiments, under many different circumstances, are typically conducted on mechanisms before constitutive relations are considered established—a case in point being the study by Ballard et al. (1995). In our view, this indicates that scientists do not think that MM provides a sufficient condition for constitution.

stituents of cognitive processes cannot be drawn experimentally. The inference to constitution is of *inherently pragmatic nature*, involving, for instance, measures of explanatory power. To illustrate, reconsider the two equivalent models in figure 4 and assume that standard procedures of causal discovery reveal that all causes of Ψ (in an extensive dataset) are common causes of Ψ and Φ_1. Both models 4a and 4b can reproduce that common-cause coupling, but model 4a moreover provides a reason for it, *viz.* the constitutive dependence between Ψ and Φ_1, and hence explains why there do not exist surgical causes of Ψ. The common-cause model 4b does not afford such an explanation because, if Ψ and Φ_1 merely are two parallel effects of common causes, as expressed in 4b, it should be possible to surgically change their values independently of one another; hence, 4b does not furnish a rationale for why such surgical manipulations do not exist. Therefore, whenever data reveal systematic common-cause coupling, models featuring constitutive relations are preferable over equivalent common-cause models because they outperform the latter with respect to *explanatory power*.

Other pragmatic virtues might be called upon as well. For example, reconsider the two models in figure 3. Any given dataset δ, in which Ψ and $\mathbf{\Phi} = \{\Phi_1, \Phi_2\}$ are systematically common-cause coupled, can be accurately modeled by either the constitutional model 3a or the pure common-cause model 3b. Yet, in addition to explaining the common-cause coupling of Ψ and $\mathbf{\Phi}$ in δ, model 3a makes a much stronger prediction than model 3b with respect to expansions of δ. 3a predicts that the common-cause coupling of Ψ and $\mathbf{\Phi}$ will remain unaltered across all follow-up studies that expand δ by further data-points or integrate further variables into δ. That is, 3a predicts that the common-cause coupling of Ψ and $\mathbf{\Phi}$ is unbreakable, meaning that it continues to hold in every expansion δ' of δ. Model 3b makes no such prediction. As to 3b, the common-cause coupling of Ψ and $\mathbf{\Phi}$ might well be broken in every expansion δ'. Hence, in accordance with the principle of maximizing explanatory power, the pragmatic virtue of *predictive strength* also gives preference to constitutional over pure causal models.

However, there likewise exist pragmatic virtues that prefer causal over constitutional models. For instance, causal models beat constitutional models with respect to the often invoked virtue of *simplicity*. Constitutional models feature all the causal dependencies of empirically equivalent common-cause models and introduce additional constitutive dependencies. That is, pure causal models account for the data by introducing fewer dependencies than constitutional ones. The former are simpler than the latter and, hence, preferable on grounds of simplicity. Another pragmatic virtue to be considered is *coherence* with, say, standard theoretical commitments in a scientific community. This virtue is of particular relevance in the context of the HEC debate, for, notwithstanding the fashionableness of theories of extended cognition, it is to be suspected that models featuring causal (and not constitutive) dependencies between cognitive and extracerebral processes will fare better with respect to this virtue than models stipulating extracerebral constituents of cognitive processes.

All of this shows, on the one hand, that resolving the debate on HEC by focusing on the mark of constitution—instead of the mark of cognition—presupposes that the participants in the debate first find an agreement on a set of pragmatic virtues warranting an inference to constitution. On the other hand, however, the above considerations suggest that such an agreement might be difficult to come by.[7] It is to be expected that proponents of HEC will promote pragmatic virtues that are favorable to the inference to constitution, while HEC's opponents will endorse virtues with an opposite leaning. As a result, turning from the mark of the cognitive to the mark of constitution is likely to simply replace an argumentative stalemate that is due to a disagreement over the metaphysics of cognition by a stalemate that stems from a disagreement over the pragmatic virtues based on which to select among empirically equivalent scientific theories or models.

6 Conclusion

The philosophical debate on the hypothesis of (extracerebrally) extended cognition (HEC) is stuck in a standstill, which is due to the fact that each side assumes a mark of the cognitive that the other side rejects. Such a standstill is unfruitful. The question as to the truth of HEC can only be profitably discussed on a ground of shared background assumptions—metaphysical, methodological and other. The idea of moving the debate forward by importing analytical means from theories of mechanistic explanation, which has recently been voiced by numerous writers, looks very attractive; first, because such theories have been developed independently of the debate on HEC and can serve as neutral background against which to profitably discuss the extension of cognition, and second, because the dominant theory of mechanistic constitution, *viz.* Craver's (2007) mutual manipulability theory (MM), induces a straightforward experimental protocol of constitutional discovery and, hence, promises to resolve the HEC debate on experimental grounds.

This paper, however, has shown that it is impossible to deliver on that attractive promise. MM is incompatible with both internalist and externalist accounts of cognition, and a metaphysically neutral modification of MM begs the question HEC addresses. Moreover, we have seen that it is impossible to produce unconfounded experimental data that could establish constitutive relations between cognitive phenomena and their spatiotemporal parts. The reason is that the latter non-reductively realize the former on a lower level, which yields that upper and lower levels are systematically coupled via common causes, which, in turn, entails that they can only be manipulated with a fat hand. We concluded that the inference to the extension of cognition is of inherently pragmatic nature. Before the debate on HEC can possibly be resolved on the basis of theories of constitution, an agreement would need to be reached on the details of the pragmatic criteria that regulate the inference to

[7] This finding echoes with Sprevak (2010), who argues that neither externalist nor internalist accounts of cognition have a clear edge over their rivals with respect to criteria of explanatory power or coherence with scientific practice.

constitution. It is up to future research on constitution and cognition to determine whether such an agreement is feasible.

References

Adams, F. and K. Aizawa (2001). The bounds of cognition. *Philosophical Psychology 14*(1), 43–64.

Adams, F. R. and K. Aizawa (2008). *The Bounds of Cognition*. Malden, MA: Blackwell Pub.

Ballard, D., M. Hayhoe, and J. Pelz (1995). Memory representations in natural tasks. *Cognitive Neuroscience 7*(1), 66–80.

Baumgartner, M. and L. Casini (forthcoming). An abductive theory of constitution. *Philosophy of Science*. http://www.journals.uchicago.edu/doi/abs/10.1086/690716.

Baumgartner, M. and A. Gebharter (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axv003.

Bechtel, W. and A. Abrahamsen (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences 36*(2), 421–441.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior 11*, 209–243.

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

Clark, A. and D. J. Chalmers (1998). The extended mind. *Analysis 58*(1), 7–19.

Craver, C. (2007). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Craver, C. and W. Bechtel (2007). Top-down causation without top-down causes. *Biology & Philosophy 22*, 547–563.

Drayson, Z. (2010). Extended cognition and the metaphysics of mind. *Cognitive Systems Research 11*(4), 367–377.

Eronen, M. (2011). *Reduction in Philosophy of Mind: A Pluralistic Account*. Frankfurt (Main): Ontos.

Eronen, M. (2012). Pluralistic physicalism and the causal exclusion argument. *European Journal for the Philosophy of Science 2*, 219–232.

Feldman, A. G. and M. F. Levin (1995). The origin and use of positional frames of reference in motor control. *Behavioral and Brain Sciences 18*, 723–806.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis 44*, 49–71.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology and Philosophy 27*(4), 545–570.

Kirchhoff, M. (2014). Extended cognition & constitution: Re-evaluating the constitutive claim of extended cognition. *Philosophical Psychology 27*(2), 258–283.

Kirsh, D. and P. Maglio (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science 18*(4), 513–549.

Machamer, P. K., L. Darden, and C. F. Craver (2000). Thinking about mechanisms. *Philosophy of Science 67*(1), 1–25.

Pöyhönen, S. (2014). Explanatory power of extended cognition. *Philosophical Psychology 27*(5), 735–759.

Romo, R., A. Hernández, A. Zainos, and E. Salinas (1998). Somatosensory discrimination based on cortical microstimulation. *Nature 392*, 387–390.

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

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology 22*(1), 1–19.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy 101*(8), 389–428.

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

Spohn, W. (2006). Causation: An alternative. *British Journal for the Philosophy of Science 57*, 93–119.

Sprevak, M. (2010). Inference to the hypothesis of extended cognition. *Studies in History and Philosophy of Science Part A 41*(4), 353–362.

Theiner, G., C. Allen, and R. L. Goldstone (2010). Recognizing group cognition. *Cognitive Systems Research 11*(4), 378–395.

Wheeler, M. (2010). In defence of extended functionalism. In R. Menary (Ed.), *The Extended Mind*. MIT Press.

Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research 91*(2), 303Ű347.

Yang, E. (2013). Eliminativism, interventionism and the overdetermination argument. *Philosophical Studies*, 321–340.

Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science 78*(2), 238–263.