accepted in *Mind and Language*
THIS IS A PREPRINT DIFFERENT FROM THE FINAL VERSION

Extended Cognition, The New Mechanists' Mutual Manipulability Criterion, and The Challenge of Trivial Extendedness

Beate Krickel (beate.krickel@rub.de)

ABSTRACT

The dispute between defenders and opponents of extended cognition (EC) has come to a dead end as no agreement on what the mark of the cognitive is could be found. Recently, many authors, therefore, have pursued a different strategy: they focus on the notion of *constitution* rather than the notion of cognition to determine whether constituents of cognitive phenomena can be external to the brain. One common strategy is to apply the new mechanists' *mutual manipulability account* (MM). In this paper, I will analyze whether this strategy can be successful. Thereby, I will focus on David Kaplan's (2012) version of this strategy. It will turn out that MM alone is insufficient for answering the question whether EC is true or not. What I call the *Challenge of Trivial Extendedness* arises due to the fact that mechanisms for cognitive behaviors are extended in a way that nobody would want to count as cases of EC. I will argue that this challenge can be met by adding a further necessary condition: cognitive constituents of mechanisms satisfy MM and they are what I call *behavior unspecific*.

Keywords: extended mind, situated cognition, constitution, processes, mechanistic explanation, mutual manipulability

1 Introduction

The different theories of what I will subsume under the label "Extended Cognition" or "EC" essentially make claims about what *constitutes* cognition. According to defenders of EC, cognition is not only constituted by features of the brain but by perception-action loops between an organism and its environment (*Enactivism*), parts of the extracranial body (*Embodiment*), or tools and other aspects of the external physical and social environment (*Extendedness*).

Examples that are taken to support EC are the role of saccadic eye-movements in the performance of memory tasks (Kaplan, 2012, p. 563; Wilson, 2004, p. 194), the use of a notebook in spatial navigation (Clark & Chalmers, 1998), an artist using a sketch pad (Hutto & Myin, 2017), Tetris players rotating blocks on a screen instead of mentally rotating them (Kirsh & Maglio, 1994), the role of bodily action in vision (Gibson, 1966; O'Regan & Noë, 2001), and the role of gestures in problem-solving, language production and processing (Chu & Kita, 2011; Clark, 2010; Goldin-Meadow & Alibali, 2013; Hostetter & Alibali, 2007), to name just a few. The differences between these examples and the differences between the claims they are supposed to support will not be relevant here. The question that is addressed here and that is answered affirmatively by all theories of EC: Is cognition (at least sometimes partly) constituted by processes external to the brain?

Opponents of EC object that all examples that defenders of EC invoke to support their claims can be explained in terms of mere causal interactions between the brain and its environment (this claim is called *Embeddedness*, see Rupert (2009)). The brain remains the only intelligible bearer of cognition. Still, it is crucially embedded in a body, perception-action loops, and an external environment that contribute inputs to the brain that, then, generates outputs into the body and the external environment.

The debate between defenders and opponents of EC has come to a dead end. David Kaplan (2012) argues that this is the case because, so far, no agreement on what determines cognition could been reached. According to Kaplan, defenders as well as contester of EC have put forward what he calls *proprietary demarcation criteria* that provide intensional characterizations of what defines cognition—or in other words: what the mark of cognition is. The debate is stagnating because these proprietary criteria all rest on assumptions that are rejected by the respective opposing camp (Kaplan, 2012, p. 549) (see also Hutto and Myin (2013, Chapter 7)). For example, Fred Adams and Ken Aizawa, who are among the most prominent opponents of EC, characterize cognition in terms of non-derived intentional content.

This characterization is rejected by many defenders of EC due to the notorious difficulty to naturalize intentionality. One characterization that is put forward by defenders of EC, such as Andy Clark and David Chalmers, is the *parity principle* (Clark & Chalmers, 1998), according to which a process is cognitive if we would not hesitate calling it 'cognitive' if it occurred in the head. This, again, is criticized by opponents of EC because it rests on a rather course grained individuation of causal roles (Kaplan, 2012, p. 550; Rupert, 2010). Since both parties work with characterizations of the mark of cognition that fits their own purposes but is rejected by the respective other camp, they are talking past each other rather than being engaged in a fruitful philosophical debate.

To revive the debate, some authors (Abramova & Slors, 2018; Gallagher, 2018; Hewitson, Kaplan, & Sutton, 2018; Kaplan, 2012; Kirchhoff, 2015, 2017; Pöyhönen, 2014; Theiner, Allen, & Goldstone, 2010; van Eck & Looren de Jong, 2016; Zednik, 2011) leave the search for the mark of cognition behind and instead focus on the "mark of constitution." (Baumgartner & Wilutzky, 2017, p. 1105). One popular strategy is to apply tools from the new mechanistic literature to make sense of constitution in the context of EC. The most promising tool is the new mechanists' *Mutual Manipulability* account (MM) of constitutive relevance (Craver, 2007b, 2007a) (it has even been adopted by working biologists (Japyassú & Laland, 2017)). Roughly, MM characterizes constitution in terms of two necessary conditions: (i) the constituents are *spatiotemporal parts* of the constituted phenomenon, and (ii) the constituents and the phenomenon are *mutually manipulable*. Applying MM to EC is prima facie promising because it provides a mark of constitution that is independent of any mark of cognition. Furthermore, as such, MM should be acceptable by defenders and opponents of EC alike given its success in other areas of (philosophy of) science.

In this paper, I analyze whether MM can be put to use for the purpose of defending EC. The most prominent and elaborated version of this strategy is David Kaplan's (2012) approach, which will be the focus of this paper. I discuss two objections against Kaplan's account that

have been raised by Michael Baumgartner and Wendy Wilutzky (2017)—what I call the *Inconsistency Problem* and the *Parthood Problem*—and show how they can be rejected. Although the two problems can be solved, the first goal of this paper is to show that MM alone cannot be sufficient to argue for EC. The reason is that many mechanisms for cognitive behaviors involve the body or the external surrounding in a way that should not be counted as making the case for EC – I call this the *Challenge of Trivial Extendedness* (CTE). For example, leg movements will come out as satisfying MM relative to the performance of spatial memory tasks such as finding the exit in a maze; arm movements will satisfy MM relative to the behavior of creating a copy of patterns of blocks and relative to the behavior of using a sketch pad in creative thinking. Though these examples suggest that our limbs are often components of the mechanisms for our cognitive behaviors, they do not show that cognition is extended in any relevant way – trivially, we need our body to move around or to move external objects.

In order to save the strategy pursued by Kaplan and others, one way to meet the CTE is to admit that MM alone is not sufficient to determine what the *cognitive constituents* of a cognitive behavior are. In order to identify those, MM has to be combined with a further criterion in line with the new mechanistic thinking that excludes the trivial cases of extended or embodied cognition. The second goal of this paper is to develop such a criterion in terms of what I call *behavior specificity* and show how extracranial elements might satisfy it.

More specifically, the paper proceeds as follows: in Section 2, I will present MM (Section 2.1) and Kaplan's approach (Section 2.2) in more detail. In Section 3, I will discuss three objections against MM's application to EC: I present and solve the Inconsistency Problem (3.1) and the Parthood Problem (3.2) and introduce the Challenge of Trivial Extendedness (3.3). In Section 4, I will discuss different proposals for how to meet the challenge. In Section 4.1, I will show why the notion of *causal specificity* will not be of much help (contra Hewitson, Kaplan, and Sutton (2018)). In Section 4.2, I will introduce a novel suggestion and show how it can be used to provide a new definition of what it is to be a cognitive constituent. Roughly, I will

argue, that cognitive constituents are components of mechanisms that are *behavior unspecific*. I will present an example that suggests that even extra-cranial elements can sometimes be behavior unspecific in this sense. Hence, EC may indeed be true: (at least sometimes) cognition seems to be (partly) constituted by processes external to the brain.

2 Constitution in Terms of Mutual Manipulability

2.1 The Mutual Manipulability Approach

The central claim of the so-called new mechanists is that explanations in the special sciences, especially biology, are *mechanistic*. Mechanistic explanations refer to the causal structures (i.e., mechanisms) underlying or causing the phenomena-to-be-explained. 'Underlying', thereby, is understood in terms of *mechanistic constitution*, or *constitutive relevance*.[1] One prominent example of such a constitutive mechanistic explanation is the explanation of spatial memory. Spatial memory is often investigated by observing mice navigating the Morris water maze (a pool filled with an opaque liquid; the mouse is supposed to find a platform that is hidden under the surface of the liquid). Spatial memory is instantiated in the mouse's navigation behavior (the phenomenon), and the mouse's hippocampus generating spatial maps is supposed to be a component of the mechanism responsible for that navigation behavior (Bechtel, 2008; Bechtel & Richardson, 2010; Craver, 2007b). Other examples of phenomena that are constitutive-mechanistically explained are the action potential (Craver, 2007b, pp. 114–122), the human heart pumping blood (Bechtel, 2006; Bechtel & Abrahamsen, 2005; Craver & Darden, 2013; Glennan, 2010), a cell synthesizing proteins (Craver & Darden, 2013; Darden, 2002; Machamer, Darden, & Craver, 2000), and long-term potentiation at synapses of neurons (Craver, 2007b, pp. 65–72; Craver & Darden, 2001, pp. 115–17, 2013, pp. 167–72; Machamer et al., 2000, pp. 8–11).

[1] Mechanistic constitution is the relation between the mechanism as a whole and the phenomenon; constitutive relevance relates single components of a mechanism and the phenomenon.

According to Craver (2007b, 2007a), in constitutive mechanistic explanations one refers to components of mechanisms that are *constitutively relevant* for the phenomenon. He specifies this notion in his *Mutual Manipulability account (MM)*:

(*MM*) X's Φ-ing is constitutively relevant for S's Ψ-ing iff:

(i) X's Φ-ing is a part of S's Ψ-ing, and

(ii) X's Φ-ing and S's Ψ-ing are mutually manipulable. (Craver, 2007b, p. 153)

X represents an entity (e.g., a hippocampus, a stomach, a neuron); the Φ-ing stands for the activity performed by X (e.g., generating spatial maps, digesting, firing), S represents a system (e.g., a mouse, a human, a car), where the Ψ-ing is a behavior of S (e.g., navigating through a maze, walking, accelerating). MM states that in order for an X's Φ-ing to be constitutively relevant for a Ψ-ing of a particular S two conditions have to be satisfied: first, constituents are *parts* of what they constitute. Following Leuridan (2012, p. 410), parthood, in this context, can be understood as follows:

(*Parthood*) X's Φ-ing is a part of an S's Ψ-ing iff the spatiotemporal region occupied by X's Φ-ing is contained in the spatiotemporal region of S's Ψ-ing.

For example, the hippocampus's generating spatial maps is a part of the mouse's navigation behavior because its spatiotemporal region is contained in the spatiotemporal region occupied by the mouse during its navigating. In the same sense, the mouse's stomach digesting or a virus infecting a cell can be part of the mouse navigating as well. In contrast to the latter two examples, constituents additionally satisfy the mutual manipulability condition. Mutual manipulability is taken to consist of two manipulations:

(*Manipulations*)

(i) Bottom-up intervention: there is an ideal intervention on X's Φ-ing with respect to S's Ψ-ing that changes S's Ψ-ing (or its probability distribution);

> (ii) Top-down intervention: there is an ideal intervention on S's Ψ-ing with respect to X's Φ-ing that changes X's Φ-ing (or its probability distribution). (Craver, 2007b, p. 153)

Ideal interventions, thereby, are spelled out in terms of Woodwardian interventionism (Woodward, 2003) (see Leuridan (2012)). Roughly, an intervention *I* on X's Φ-ing with respect to S's Ψ-ing is ideal only if X's Φ-ing changes only due to the influence of *I* and S's Ψ-ing is not changed by *I* directly.

Importantly, constitutive relevance is a non-causal relation. If an X's Φ-ing is constitutively relevant for S's Ψ-ing, then it is impossible that X's Φ-ing causes S's Ψ-ing (and vice versa) (Craver, 2007b; Craver & Bechtel, 2007). Most arguments in favor of this claim are based on the assumption that X's Φ-ing and S's Ψ-ing are not wholly distinct events because X's Φ-ing occupies a sub-region of the spatiotemporal region of S's Ψ-ing (condition (i)) and both are mutually dependent (condition (ii)) (Craver & Bechtel, 2007; Romero, 2015).[2] However, one platitude about causation is that its relata have to be wholly distinct (Craver and Bechtel 2007; Lewis 1973; 1986). Thus, it follows that causal relevance and constitutive relevance are mutually exclusive relations.

[2] A further argument in favor of the claim that constitutive relevance is a non-causal relation is that otherwise we are confronted with causal loops (Baumgartner & Gebharter, 2016; Craver & Bechtel, 2007; Romero, 2015). Causal loops are held to be problematic because, roughly, their existence would imply that something could be the cause of its own cause (Kim, 1999; Romero, 2015). Causal loops are unproblematic if they are taken to be feedback loops where an effect is a cause of an effect that is of the same type as the cause of the first effect. The problematic element comes in when assuming that the putative cause and effect occur at the same time. A further argument Romero provides is based on exclusion worries arguing that if constitutive relevance would be a causal notion that would imply that there is downward causation, which is not compatible with the causal closure of the physical. This argument does not depend on the assumption that the relata of constitutive relevance are parts and wholes. I will not discuss this objection in the present paper. For a discussion of exclusion worries in the context of interventionism see Baumgartner (2009), Woodward (2015), and Gebharter (2015). A discussion of exclusion worries concerning causation between mechanistic levels see Craver (2007b, Chapter 6), and between parts and wholes see Kim (1998, p. 84).

2.2 Applying Mutual Manipulability to Extended Cognition

In a recent paper, David Kaplan (2012) suggests to use MM to make sense of constitution in the context of EC. Using the mutual manipulability account is promising for several reasons. First, it does not depend on any specification of what defines cognition (Kaplan, 2012, p. 557). According to Kaplan, the mutual manipulability account provides a *generic demarcation criterion* that specifies the extension of cognition without depending on any intensional characterization of 'cognition.' Instead of starting with a definition of 'cognition', the strategy is to focus on examples that everyone accepts as cognitive (for an application of this strategy see also Newen (2017)). If these examples turn out to be cases of extended, embodied, or enacted cognition, EC is true. Since the mutual manipulability account deals with constituents of the *behaviors* of *specific objects* (Baumgartner & Casini, 2017; Baumgartner & Gebharter, 2016; Kaiser & Krickel, 2017), Kaplan focusses on behaviors of specific objects as well—in this case *cognitive behaviors of human agents* (Kaplan does not make this explicit but I take it to be a valid description of what he is doing).

Note that this strategy does not *identify* cognition with (a type of) behavior (for a criticism of the claim that cognition is (a type of) behavior see Aizawa (2017)). Kaplan only starts from the assumption that, intuitively, some behaviors involve cognition in a relevant way. Behaviors like walking, moving your arms, and scratching your nose are usually not taken to involve cognition in a way that would validate EC (while these behaviors might still be *caused* by a (purely internal) cognitive process). But everyone agrees that behaviors performed by human subjects such as copying a pattern of colored blocks, solving a math problem, thinking about what you had for lunch yesterday, learning how to play the piano, or typing a text on a keyboard are cognitive in a more substantial way. In this sense, Kaplan starts with an *extensional characterization* of cognition, i.e., with a list of examples that count as cognitive phenomena and those that do not. Kaplan's strategy now is *not* to ask the question "What makes these

behaviors cognitive?" Rather, he asks "What constitutes these behaviors?" to determine whether the constituents can be found outside the brain.

A second reason why Kaplan's suggestion is promising is that, as specified in the previous sub-section, constitution in terms of MM is taken to be a *non-causal* relation (Bechtel, 2008; Craver, 2007b; Craver & Bechtel, 2007). Hence applying MM to EC is promising with respect to avoiding the coupling-constitution (CC) fallacy which poses a challenge for arguments in favor of EC (Adams & Aizawa, 2001, 2008, 2010; Aizawa, 2010). According to Fred Adams and Ken Aizawa, many arguments for EC rely on an inference from a causal relation between a cognitive element and an external element to a constitution-claim about the external element constituting the cognitive element or about the cognitive and the external element together constituting a further cognitive element. This inference, according to Adams and Aizawa, is fallacious as there are many cases where elements are causally connected in the way identified by defenders of EC but where nobody would want to infer to a constitution-relation. Hence, causation cannot be sufficient for constitution. In the light of MM, according to which constitution and causation are mutually exclusive relations, it is clear that such a fallacious inference is blocked (I will discuss the CC fallacy in the context of MM in more detail in Section 3.2).

Third, MM is motivated by and is supposed to account for actual scientific practice. The account builds on Woodward's (2003) interventionism which uses the idea of an ideal experiment to spell out what causation is. The underlying notion of an ideal experiment (or an ideal intervention) also provides the driving ideal for designing experiments and reasoning about experiments in the actual empirical sciences. Craver uses Woodward's account to provide an operational characterization of mechanistic constitution. Hence, using this account in the context of EC promises to base the debate on an empirically adequate fundament and it promises a framework for evaluating empirical tests of constitution claims in the context of EC.

To illustrate his application of MM to EC, Kaplan uses the example of a memory study in which subjects have to copy a pattern of colored blocks. In this study it was found that instead of memorizing the whole pattern in order to copy it, subjects performed frequent saccadic eye movements between the original pattern and the copy they were creating (Ballard et al., 1995; Kaplan, 2012, p. 563). According to Kaplan, this shows that the saccadic eye movements are constituents of the memory behavior because they satisfy MM: first, by confronting subjects with such a memory task, their eye movements changed (top-down intervention) (Kaplan, 2012, p. 564). Second, when gaze was fixated to one spot, it took subjects much longer to finish the copying task (bottom-up interventions) (Kaplan, 2012, p. 564).

Indeed, this application seems to provide the right results with regard to non-constituents: gravity, although a necessary condition for any memory task, is not constitutive for the copying behavior because gravity cannot be changed by intervening into the subject's behavior during the task (while it is possible to change their behavior by changing the influence of gravity on the brain). Also, with help of the criterion one can exclude implausible extendedness claims such as the claim that perception is constituted by the objects of perception (Rupert, 2009). Again, although it is possible to change perception by changing the perceived object, it is not possible to change the perceived object by changing perception (e.g., by intervening into V1).

To summarize: the idea of using MM to defend EC is promising because it is independent of any specific definition of what cognition is, it does not rely on a flawed inference from causation to constitution, and it promises to provide an empirically testable criterion to solve the EC dispute. Still, as I will argue in the next section, the suggestion is confronted with several challenges.

3 Problems, Solutions, and a Challenge for MM's Application to EC

3.1 The Inconsistency Problem

The first problem is a problem for MM's application to EC in so far as it poses a general challenge for MM. Different authors (Baumgartner & Casini, 2017; Baumgartner & Gebharter, 2016; Eronen & Brooks, 2014; Kästner, 2017; Leuridan, 2012; Romero, 2015) have argued that the combination of interventionism and non-causal dependency relations, as for example mechanistic constitution, is problematic (these problems have been discussed in the context of EC by Baumgartner and Wilutzky (2017), and van Eck and de Jong (2016)).

Roughly, the problem is that ideal interventions into phenomena (S's Ψ-ing) with respect to their constituents (a set of X_i's Φ_i-ings) (top-down interventions; see Section 2.1) are impossible. Interventions into phenomena are necessarily *fat-handed*, i.e., they are necessarily common causes of the phenomenon and a constituent via two independent paths. Ideal interventions, as originally defined by Woodward (2003), must not be common causes in this way (see Section 2.1). As a consequence, MM fails as it makes use of a theoretical tool (interventionism) in a context for which the tool was not made. The Inconsistency Problem thus stems from the incompatibility of the criteria defining ideal interventions and constitution.

In order to address this problem, different authors (Baumgartner & Casini, 2017; Baumgartner & Gebharter, 2016; Baumgartner & Wilutzky, 2017) make use of Woodward's modified definition of an ideal intervention (Woodward, 2015) (let's call this ideal* intervention). Ideal* intervention can be fat-handed as long as the two variables relative to which it is fat-handed are related by supervenience or any other non-causal dependency relation. In order to be able to tell apart causal from constitutive components based on ideal* interventions, Baumgartner and Gebharter (2016, p. 748) introduce time into the definition of interventionist causation and constitution. Causation takes time, whereas constitution does not.

Hence, an ideal* intervention uncovers causal relations only if the two variables change at different times, whereas constitution implies simultaneous changes.

Although by presupposing this modified notion of an ideal intervention one solves the Inconsistency Problem, it results in a new problem: it leads to an inconclusive test for constitution based on which one can only abductively infer to constitution (Baumgartner & Casini, 2017; Baumgartner & Gebharter, 2016; Romero, 2015). This is the case because in order to empirically establish that X's Φ-ing and S's Ψ-ing are related by constitutive relevance it does not suffice to show that there is *one* ideal* intervention that leads to simultaneous changes in X's Φ-ing and S's Ψ-ing (thought it might suffice based on a further modified notion of an intervention; see Baumgartner, Casini, and Krickel (2018)). The reason is that the intervention might not have been ideal* at all but rather just an ordinary common cause of the changes in X's Φ-ing and S's Ψ-ing.

In order to establish a constitutive relation, one has to show that *all* interventions into S's Ψ-ing are fat-handed with respect to S's Ψ-ing and at least one of its constituents. As this is empirically impossible, the best one can do is to show that *all interventions actually performed* are fat-handed in the required sense and argue that the best explanation for this is that the two effects are constitutively related. A further downside of this account is that it does not specify what it means for an *individual* part to be a constituent. It only specifies under which conditions a *set of parts* is a set of constituents (Baumgartner et al., 2018; Krickel, 2018). Hence, it remains unclear what renders individual parts constitutively relevant for a given phenomenon.

According to a recent suggestion by Krickel (2018), these problems can be avoided if one takes the temporal extendedness of the phenomenon seriously (for further applications of Krickel's account to EC see Gallagher (2018) and Abramova and Slors (2018)). Mice navigating water mazes, beating hearts, or action potentials being transmitted are temporally extended processes (Krickel takes phenomena to be what she calls 'entity-involving occurrents' or 'EIOs'; see also Kaiser and Krickel (2017)). Due to this, they have what Krickel calls

'temporal EIO-parts.' An example will suffice to get the idea of what temporal EIO-parts are: the navigation behavior of the mouse has different temporal EIO-parts such as the moment when the mouse is put into the water maze, the different episodes of the swimming such as the mouse's turning left, swimming in circles etc., and finally it's reaching the platform. In other words, temporal EIO-parts could be described as the temporal phases of a given behavior of a specific object.

Based on this analysis of the temporal dimension of mechanistic phenomena, Krickel (2018) interprets MM in terms of two causal steps (she calls the resulting approach *Causation-based CR*; for an illustration see Figure 1):

(*Causation-based CR*) X's Φ-ing is constitutively relevant for S's Ψ*-ing iff:

(i) X's Φ-ing is a spatial EIO-part of S's Ψ*-ing,

(ii) there is a temporal EIO-part of S's Ψ*-ing that is a cause of X's Φ-ing, and

(iii) there is a temporal EIO-part of S's Ψ*-ing that it is an effect of X's Φ-ing.

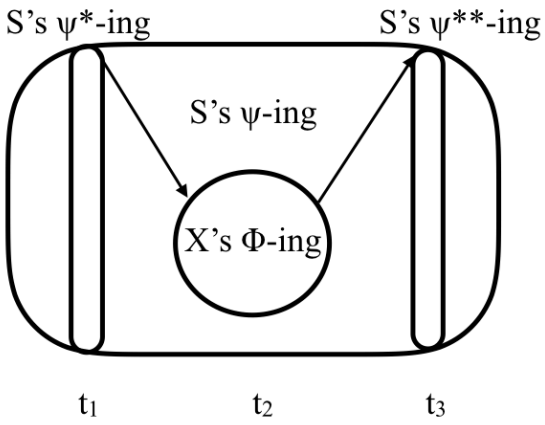

Figure 1 Krickel's interpretation of MM: mutual manipulability holds between temporal EIO-parts (S's Ψ*-ing and S's Ψ**-ing) of the phenomenon (S's Ψ-ing) and a constituent (X's Φ-ing), whereas constitutive relevance is a relation between X's Φ-ing and S's Ψ-ing (adopted from Krickel (2018), Fig. 6).

If we think of MM in this way, interventionism can be straightforwardly applied because constitution is analyzed in terms of two causal connections—which fall into the original and non-problematic area of application of interventionism. Given that S's Ψ*-ing, S's Ψ**-ing,

and also X's Φ-ing themselves will have lower-level constituents, mutual manipulability will turn out to consist of the following two *ideal** interventions:

1) *Bottom-up intervention*: an ideal* intervention on X's Φ-ing and at least one of its constituents with respect to S's Ψ**-ing, which is a temporal EIO-part of S's Ψ-ing, that causes changes in S's Ψ**-ing (or its probability distribution);
2) *Top-down intervention*: an ideal* intervention on S's Ψ*-ing, which is a temporal EIO-part of S's Ψ-ing, and at least one of S's Ψ*-ing's constituents with respect to X's Φ-ing that causes changes in X's Φ-ing (or its probability distribution).

Even though MM is analyzed in terms of two interventions identifying causal relationships, Krickel's analysis can maintain the distinction between causation and constitution: it still holds that if something is a cause of something else, it cannot be a constituent thereof, and *vice versa*. The reason is that constitution holds between S's ψ-ing and (some of) its parts, whereas the relevant causal connections hold between X's Φ-ing and temporal EIO-parts of S's ψ-ing of which X's Φ-ing is not a part. Furthermore, this analysis is applicable to individual components independently of a set of constituents. One only has to establish that there is a temporal EIO-part of the phenomenon that is a cause of the individual part and that there is a temporal EIO-part of the phenomenon that is an effect of that part.

There is one potential objection against the use of Krickel's account in the context of EC. Her account is *non-reductive* in the sense that the constitutive relation between X's Φ-ing and S's Ψ-ing is analyzed in terms of ideal* interventions that are common causes of temporal EIO-parts of S's Ψ-ing and at least one of their *constituents*. Given that interventionism is non-reductive in general as it spells out what causation is in causal vocabulary, this should not be worrisome in general. Krickel's account is non-viciously circular in the same way in which Woodward's interventionism is non-viciously circular (Woodward, 2003, pp. 20–22). But in the context of EC this feature might create a problem: if the constituents of the temporal EIO-parts of S's Ψ-ing (on which we intervene in a top-down manner) are external elements, we

seem to beg the question against the opponents of EC as we are *presupposing* that S's Ψ-ing has external constituents (for a related worry see Wilutzky and Baumgartner (2017, pp. 1116–1117)). The solution to this problem has to wait until the end of Section 3.3 because it requires a better understanding of the nature of the cognitive phenomena at issue, and the part-whole relation involved.

3.2 The Parthood Problem

In its original formulation, MM states that the part-whole relation between the phenomenon and the constituents is a necessary condition for constitutive relevance. Kaplan seems to drop this condition without argument. He summarizes MM as follows:

> (M1) When φ is set to the value φ_1 in an (ideal) intervention, then ψ takes on the value $f(\varphi_1)$ [or some probability distribution of values $f(\varphi_1)$].
>
> (M2) When ψ is set to the value ψ_1 in an (ideal) intervention, then φ takes on the value $f(\psi_1)$ [or some probability distribution of values $f(\psi_1)$] (Kaplan, 2012, p. 558).

(Here, φ stands for a lower-level mechanistic component, whereas ψ stands for a higher-level phenomenon.) Apparently, according to Kaplan's reading of MM, mutual manipulability between the putative constituent and the phenomenon is sufficient for constitution. However, the parthood condition, especially in the context of EC, is crucial. The parthood condition is the distinctive difference between constitutive and causal relevance (see Section 2.1). If the parthood condition is dropped, constitutive relevance just is causal relevance. In the present context, this is especially problematic because it leads Kaplan to commit the famous *coupling-constitution (CC) fallacy*.

The CC fallacy is one of the most prominent objections against constitution-based arguments for EC. It is presented and defended in various papers by Adams and Aizawa (Adams & Aizawa, 2001, 2008, 2010; Aizawa, 2010). The CC fallacy describes the inference from an assumption about a causal connection between two elements to a constitutive connection

between these elements or between these elements and a third element as fallacious. For example, only because the expansion of a bimetallic strip in a thermostat is causally connected to the motion of atoms of the air in the room the thermostat is in does not imply that the expansion is constituted by the air atoms as well (Adams & Aizawa, 2001, p. 56). Neither does this establish that the air molecules are constituents of whatever is constituted by the expansion of the bimetallic strip.

By dropping the parthood condition, Kaplan's reconstruction of MM traps into the CC fallacy. To see that consider Figure 2 that illustrates the mutual manipulability of two event-variables X and Y. Assume X represents the event type of being tired and Y stands for the event type of sleeping. Being tired causes people to sleep. Sleeping causes people to be less tired. Hence, X is a cause of Y, and Y is a cause of X. The two are mutually manipulable. Still, clearly, being tired is not constitutive for sleep, and sleep is not constitutive for being tired. Hence, inferring from mere mutual manipulability to constitution is to commit the CC fallacy.

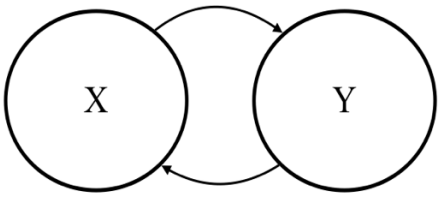

Figure 2 Mutual manipulability between variables representing event types.

This shows that we have to put the parthood condition back into MM. In other words, we have to require that in order for an external element to be a constituent of a cognitive behavior, the former has to be a part of it. However, according to Baumgartner and Wilutzky, making such an assumption is question-begging:

> [I]f 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. (Baumgartner & Wilutzky, 2017, pp. 1110–1111)

Is this objection justified? Despite the initial plausibility of Baumgartner and Wilutzky's worry, assuming parthood does *not* beg the question against the opponent of EC. According to MM, parthood is not sufficient for constitution. It is only a necessary condition. But the search for the bounds of cognition is a search for *constituents* of cognitive processes, not for *parts* in the sense used here (see Section 2.1). Rather, *denying* parthood would be question-begging as denying parthood is already sufficient for denying constitution. Hence, we should leave room for the possibility of a part-whole relation between a cognitive phenomenon and an extracranial element.

A further argument for why assuming parthood is more plausible than its denial is that denying parthood leads to an odd picture. To see this, let us take the example of the memory demanding copying task Kaplan uses as an example of EC. Here, the phenomenon is a temporally extended cognitive behavior of a subject copying a pattern of colored blocks. It starts with task onset and ends when the pattern of blocks is completely copied. What would it mean to deny that the saccadic eye movements are parts of the subject's behavior?

Two pictures are possible. First, one might deny *temporal containment*. One could argue that the behavior of copying is scattered in the sense that it stops when the eyes start their saccadic movements and begins only after the eyes' movements have started (see Figure 3a)). According to this picture, the saccadic eye movements would not occupy sub-regions of the spatiotemporal region of the copying behavior, and hence would not be part of it in the sense defined in Section 2.1. This is a very odd picture. First, we will not find the necessary clear-cut stating and end-points of the eye movements that would be necessary to draw this picture. Second, the eye movements will occur simultaneously with many other processes, such as neural activity, that opponents of EC would want to count as constituents of the cognitive behavior. Hence, they should not assume that the copying process is scattered in this way.

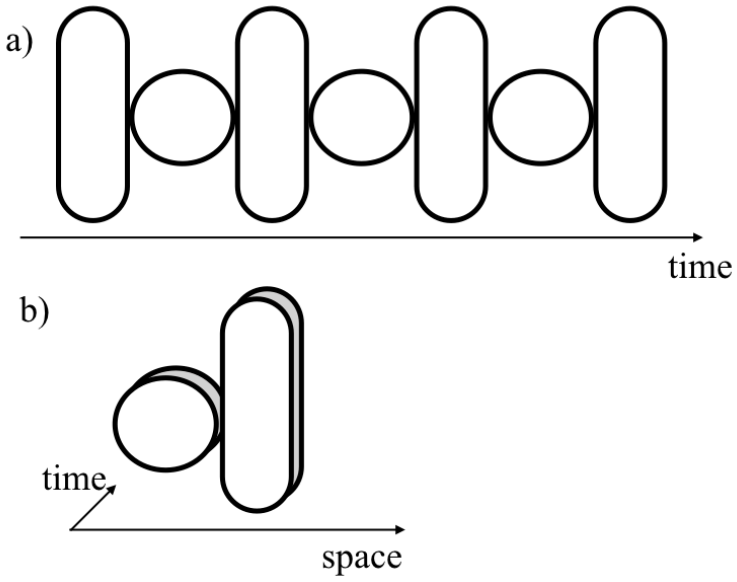

Figure 3 Possible strategies to deny parthood. The small circles stand for the putative constituent (e.g., saccadic eye-movements), whereas the large slices stand for the phenomenon (e.g., the copying behavior).

A second strategy to deny parthood is to deny *spatial containment*. One could argue that the eye movements, though temporally contained in the copying behavior, do not occur in the same spatial location. For example, the eye-movements might occur right next to the copying process (see Figure 3b)). How could one argue for such as claim? The only argument that comes to my mind is to insist that cognition only happens in the brain. But this would, obviously, beg the question against the defender of EC.

We can conclude that both conditions of MM—parthood and mutual manipulability—can in principle be satisfied by external elements relative to cognitive phenomena. Not only the Inconsistency Problem can be avoided. Also, the Parthood Problem can be solved. But does that suffice to make the case for EC?

3.3 The Challenge of Trivial Extendedness

Based on the interpretation of MM along the lines of Krickel's account as presented in Section 3.1, and based on the clarification of the parthood condition presented in Section 3.2—can MM's application to EC be successful? Can EC be defended based on MM?

At least, prima facie, the application of MM to EC delivers the right results. Saccadic eye-movements are parts of the memory demanding copying behavior (given the reasoning from Section 3.2) and the former and the latter are mutually manipulable (as specified in Section 3.1). Plausibly, if we change the copying task used by Ballard et al. by, for example, changing the pattern of colored blocks that is to be copied at t_1 this will lead to a change in the eye-movements at t_{n+1}. If we intervene on the eye-movements by fixing the gaze at t_2, the copying behavior will change afterwards (as is indicated by the fact that subjects need much more time). Hence, all conditions of the mutual manipulability account are satisfied.

Still, the application of MM to EC runs into a problem. Consider the following argument:

1. The subject's arm movements are part of the subject's copying behavior.
2. The subject's arm movements and her copying behavior are mutually manipulable.
3. Hence, the subject's arm movements are constituents of the copying behavior.

Based on MM, this is a sound argument. Given the reasoning from Section 3.2, we should take the arm movements to be parts of the subject's copying behavior. Also, given Krickel's interpretation of MM introduced in Section 3.1, the arm movements and the copying behavior are plausibly mutually manipulable. It is easy to imagine that if you, for example, fixate the subject's arms so that she cannot move them anymore, the copying behavior will be affected. Also, if you change the pattern of the colored blocks, the arm movements will differ. Thus, the two conditions of MM are satisfied, and the arm movements turn out to be constituents of the copying behavior. But should this lead us to accepting EC?

Kaplan himself argues that arm movements "naturally seem like background conditions and not working parts of the memory mechanism underlying task performance" (Kaplan, 2012, p. 564). Thus, even defenders of EC are not willing to accept arm movements in copying tasks as verifying EC. However, given that arm movements satisfy MM it remains unclear what distinguishes them from cases that are counted as verifying EC. Note that the problem generalized as there are many analogous cases: moving your lips will turn out to be a

constitutive part of speech production, eye movements will be constitutive for reading, and finding your way to the MOMA will involve your legs as constitutive parts. None of these cases will be counted as verifying EC. Still, in all of these cases, MM seems to be satisfied.

The examples indicate that the mechanisms that constitute cognitive behaviors will be extended purely due to the fact that we are dealing with cognitive *behaviors* (for a similar worry see Aizawa (2017, p. 4276)). Thus, as such MM cannot be sufficient to make the case for EC. What we need is a criterion that helps us to distinguish between those constituents of a cognitive behavior that are constituents of it qua the latter being a *behavior*, and those that are constituents qua the latter being a *cognitive* behavior. Finding such a criterion is what I call the *Challenge of Trivial Extendedness* (CTE).

Before I come to discussing different strategies for meeting this challenge, I want to highlight that the trivial extendedness of cognitive behaviors is advantageous in one respect. Remember the problem mentioned at the end of Section 3.1. The problem was that by using Krickel's account of MM, the whole endeavor seems to become question-begging. The reason was that Krickel's account is non-reductive: X's Φ-ing is a constituent of S's Ψ-ing only if there is an ideal* top-down intervention into a temporal EIO-part of S's Ψ-ing with respect to X's Φ-ing. This intervention will be ideal* because it affects not only the temporal EIO-part of S's Ψ-ing but also at least one of the constituents of the temporal EIO-part. But if the constituents of the temporal EIO-part are external too, we have to already presuppose that S's Ψ-ing has external constituents. Hence, (it seems) we have to presuppose the truth of EC.

Based on the insight that cognitive behaviors are extended in a trivial sense, one can solve this problem. Take again the example of the copying task used by Ballard et al. According to Kaplan, the top-down intervention on the copying behavior with respect to saccadic eye-movements consists in "engaging subjects in this cognitively demanding task" (Kaplan, 2012, p. 564). Let us assume that subjects are engaged in this task by receiving verbal instructions by an experimenter. In order for this intervention to be ideal* in the way required by Krickel it has

to be an intervention into a temporal EIO-part of the phenomenon, say, the subject's looking at the pattern-to-be-copied, and one of the constituents of this temporal EIO-part, which is, for example, the activity of the retina. Does this mean that we already have to presuppose the truth of EC because the retina has to be assumed to be a constituent of the copying behavior?

We do not. The reason is that not every constituent of a cognitive behavior will validate EC as argued above. Nobody would deny that behaviors such as the one described in the study by Ballard et al. involve the body: subjects have to look at the pattern, move their arms to move the blocks, etc. Hence, for every cognitive behavior the identification of some constituents can be presupposed in our analysis as long as these are constituents of the cognitive behavior *qua behavior*. Hence, Krickel's approach can be applied without begging the question.

4 Meeting the CTE: What are Cognitive Constituents?

In the previous section I argued that MM alone cannot be sufficient to argue in favor of EC. MM identifies also purely behavioral elements such as the arm movements in Ballard et al.'s copying task, as constituents of the cognitive behavior at hand. As a consequence, elements would come out as validating EC that nobody, not even defenders of EC, would want to count as cognitive elements. Hence, as a sufficient criterion, MM renders EC absurd. We need a way to distinguish between constituents of a cognitive behavior qua being cognitive and those that are constituents qua being behaviors. But can we meet the CTE?

Clearly, any strategy for addressing this challenge must not refer to any mark of the cognitive. The starting point of the present discussion was to accept that there is no account of the mark of the cognitive that defenders as well as opponents of EC accept. Hence, in order to avoid begging the question one basic working assumption of the present endeavor was: there is no mark of the cognitive. What we are looking for is a criterion to determine the extension of 'cognitive constituent' rather than its intension.[3]

[3] It is crucial to keep in mind that the criterion I am going to develop is not meant to identify cognitive constituents in any pre-defined sense of 'cognitive.' Rather, it will allow to draw a distinction between trivial cases of

Furthermore, in order to save the general strategy of using resources from the new mechanistic account to argue for EC, the challenge should be answered with the help of tools that are not foreign to the new mechanistic thinking and the explanatory practice of the cognitive and life sciences. Indeed, in a recent paper, Hewitson, Kaplan, and Sutton (2018) address a worry that looks similar to the one at issue that is compatible with the general strategy. They attempt to solve the problem by invoking the notion of *causal specificity* introduced by James Woodward (2010) that is in line with Craver's original ideas regarding MM. Hence, Hewitson, Kaplan, and Sutton's suggestion seems promising with regard to answering the CTE. In what follows, I will present their idea. I will argue that their suggestion fails as an answer to the CTE.

4.1 Causal Specificity

In a recent paper, Hewitson, Kaplan, and Sutton (2018) address a worry that looks similar to the one at issue: they argue that although the application of MM in the context of EC is prima facie fruitful, "there is a residual worry that MM is still not restrictive enough" (p. 28). Indeed, they argue that Craver himself was aware of this worry in the original formulation of MM in his (2007b)-book. They cite an example used by Craver (2007b, p. 158) to illustrate the problem: a subject's performing a word stem completion task and her heart's pumping blood might satisfy MM because if you lesion the heart, the subject will not be able to perform the word stem completion task anymore; and if you let the subject perform "torturous word-stem completion tasks outside of the context of the request for explanation" her heart beat will be affected (2007b, p. 158). Hence, the heart's pumping blood turns out to be a component of the mechanism for word stem completion. However, nobody would take the heart's activity to be explanatory relevant for how word stem completion works. Hence, MM is insufficient.

extendedness and more interesting ones. As I will argue below, the interesting cases deserve the label 'cognitive' in light of the assumptions of the present strategy as they are components of mechanisms for general *cognitive capacities* rather than just of the mechanisms that are counted as manifestations of these cognitive capacities.

Inspired by Craver's own suggested solution, Hewitson et al. propose to use the notion of *causal specificity* introduced by Woodward (2010) to distinguish between constituents and what they and Craver call *background conditions* (2018, pp. 28–29): intervening into background conditions will have rather nonspecific effects on the phenomenon at hand. For example, lesioning the heart will completely extinguish the performance of word stem completion. But it is impossible to intervene on the heart's activity in a way that leads to more specific changes in the performance of word stem completion. Hewitson et al. frame this suggestion as a new strategy to investigate potential cases of EC. They suggest that a comparison of the 'specificity profiles' of internal and external elements may provide useful: if it turned out that the profiles are similar, this would provide a new argument in favor of EC (2018, p. 29).

Although Hewitson et al.'s idea of combining MM with the causal specificity seems promising with regard to distinguishing between constituents and causal background conditions, it will not suffice to meet the CTE—for the main reason that trivial cases of extendedness are *not* background conditions of mechanisms. This is already indicated by the fact that distinguishing between constituents and causal background conditions poses a *general* challenge for MM independently of its application to EC. If the heart's activity turns out to be a component of the mechanism for word stem completion on the basis of MM, this shows that MM fails as an account of constitutive relevance *in general*. In order for MM to be successful, one has to implement a criterion that excludes cases like the heartbeat case in relation to word stem completion. But this has nothing to do with the CTE. This challenge arises even if the problem of causal background conditions is solved.

CTE is specific to MM's application to EC. The crucial point to note is that, given the basic motivation of MM, arm movements are plausibly counted as *components* of the relevant mechanism rather than background condition to it. Arm movements *are* explanatory relevant for how subjects perform the copying behavior in the setup used in this study—in contrast to,

for example, the subject's heart-beat which is not specific to the behavior at hand. In other words: arm movements are constituents rather than mere causal background conditions relative to the copying behavior. Causal background conditions in this example would be, for example, the heartbeat, blood circulation, gravity, or the oxygen concentration in the surrounding the subjects. With regard to *these* elements the test for causal specificity seems promising.

But this will not work for arm movements. Indeed, it is possible to intervene on arm movements in a way that leads to rather subtle changes in task behavior. For example, if one were to move the subjects' arms right before they place the blocks, the copied pattern would look slightly different. There are interventions into arm movements that do seem to have rather fine-grained causal influence over the copying behavior. But, as already explained above, this is a good result as arm movements are plausibly constituents of the mechanism for the copying behavior investigated by Ballard et al. The CTE requires us to find a criterion that distinguishes between *different kinds of mechanistic constituents*, i.e., elements that rightfully satisfy MM (and causal specificity). The question is not whether external elements can be components of mechanisms for cognitive behaviors (this seems unproblematic). Rather, the question is under which conditions system-external components of mechanisms for cognitive behaviors should count as validating EC.

In order to be able to express this difference, I will use the term *constitutive background* to refer to those constituents that never count as validating EC (even if they are system-external). These components provide the constitutive background for the *cognitive constituents*. Only if it can be shown that the cognitive components can be external to the brain, this should be counted as validating EC. The CTE, thus, can be rephrased: What is the difference between the constitutive background and cognitive constituents?

4.2 Behavior Specificity

The first step towards answering the question posed at the end of the previous sub-section is to highlight that cognitive behaviors investigated in experiments, such as Ballard et al.'s, are usually just a means to an end for the actual aim of investigating a more general *cognitive capacity*. For example, the study conducted by Ballard et al. does not primarily aim at finding out how subjects copy patterns of colored blocks. Rather, the aim is to get general insights about short-term or working memory, i.e., the capacity to recall a certain amount of information after a short period of time.[4] In other words, cognitive scientists are primarily interested in cognitive capacities which are investigated by conducting experiments testing their *manifestations*, i.e., cognitive behaviors.

Based on this consideration, the idea that I want to put forward here is that the cognitive constituents of a behavior are those elements that constitute the behavior *because the behavior is a manifestation of a general cognitive capacity* and not due to the specificities of the behavior chosen for the purposes of the experiment. In contrast to that, the constitutive background is composed of those constituents that are *specific to the behavior that was chosen to be tested in the experiment*. In other words: the constitutive background is what I call *behavior specific*, whereas cognitive constituents are *behavior unspecific*. In this section, I spell this idea out in more detail.

Following Cummins (1983, p. 53), cognitive capacities can be characterized in terms of input-output conditions. For example, working memory can be characterized as the capacity to recall a certain amount of information (output) a short time after receiving the information (input). A behavior counts as a manifestation of this cognitive capacity if and only if it

[4] Indeed, Ballard et al.'s study does not aim at finding the *mechanism* for working memory. Rather, the goal is formulated as "delimiting the upper bounds of [working] memory in (…) the performance of everyday tasks" (Ballard et al., 1995, p. 66). Thus, Ballard et al. are investigating to which degree a specific mechanism, namely working memory, is used in natural situations. Here, we are interested in the question of whether mechanisms for cognitive behaviors can be extended. Hence, in order to be able to use Ballard et al.'s study as an example, we have to frame it in terms of an investigation aiming to find the mechanism for working memory.

constitutes a process that connects the input with the output that characterizes the capacity at issue. For example, in Ballard et al.'s study, the behavior connecting the input ('receiving information') and the output ('recall information') is the copying behavior in which subjects look at a pattern of colored blocks and use their hands to move the blocks from the store. Note, that in that sense, only *successful* behaviors, i.e., those that actually produce the output, count as manifestations of the respective capacity.

For each cognitive capacity there will be different behaviors whose performance counts as a manifestation of it. For example, the inputs and outputs characterizing working memory could also be connected by a change-detection behavior (Rouder, Morey, Morey, & Cowan, 2011). Furthermore, for any given type of behavior (e.g., copying behavior, change-detection behavior) there are different ways the inputs and outputs can be operationalized in order to conduct experiments on the behaviors. For example, Ballard et al. use an operationalization of the input-output characterization of working memory according to which the input is 'subjects visually perceive a pattern of colored blocks' and the output is 'manually produce an instance of the same pattern.' They could have investigated the same type of behavior by operationalizing the relevant input and output type differently. For example, they could have asked the subjects to verbally report which block they would put where instead of actually doing it; they could have chosen different stimuli; and so on.

The core idea now is as follows: In order to experimentally investigate the mechanism underlying a *cognitive capacity* (such as working memory) one can only investigate *behavioral manifestations* of that capacity (such as the copying behaviors investigated by Ballard et al. where subjects used their hands to copy colored blocks). The mechanism that one will find by conducting experiments on these behavioral manifestations will involve elements that are not relevant for the execution of the cognitive capacity as such but only for the execution of that particular behavioral instance given the particular task setup chosen for the concrete experiment. Thus, in order to find out which elements are constitutive of the cognitive capacity,

we have to be able to distinguish between those components of the mechanism for the behavior that are components *qua the behavior's being a manifestation of a general cognitive capacities* and those elements that are components of the mechanism only *qua the behavior's being that particular behavior*. This is what the criterion of 'behavior specificity' is supposed to do. Based on these considerations, the criterion can be spelled out as follows:

(*behavior specificity*)

a) A constituent of a behavior B is *behavior specific* if and only if it is a component of the mechanism for B under *some but not all* operationalizations of the inputs and outputs that characterize the cognitive capacity of which B is a manifestation.

b) A constituent of a behavior B is *behavior unspecific* if and only if it is a component of the mechanism for B under *all* operationalizations of the inputs and outputs that characterize the cognitive capacity of which B is a manifestation.

Figure 4 illustrates the experimental procedure to test for behavior specificity.

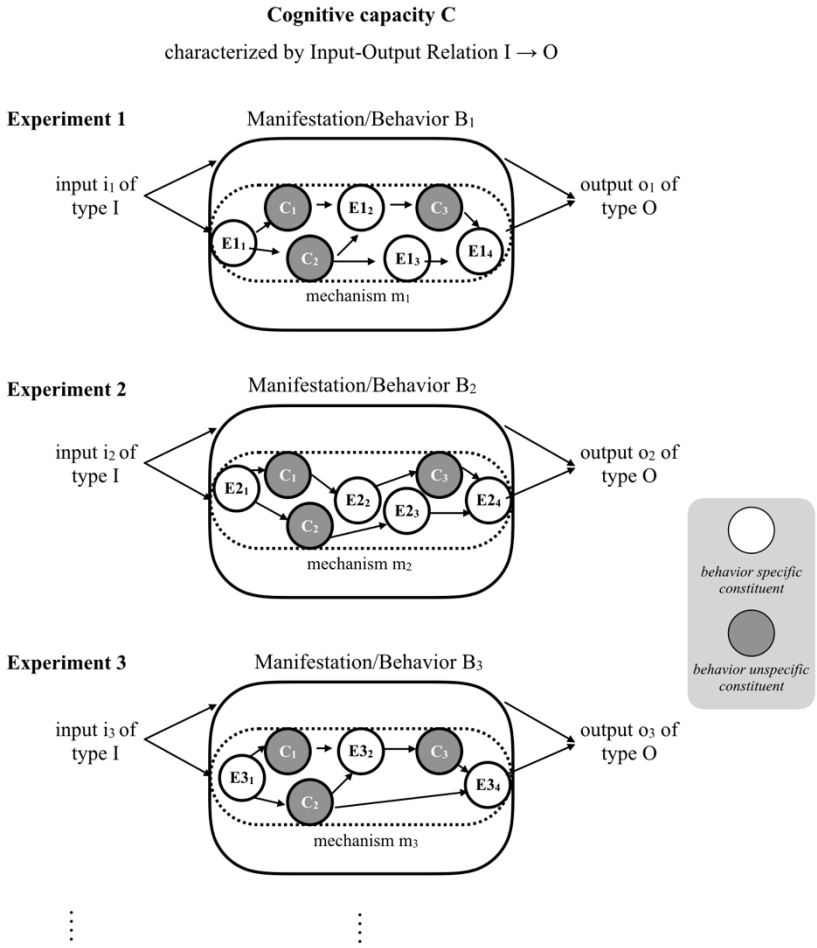

Figure 4 Illustration of *behavior specificity*: In all experiments the same cognitive capacity C characterized by input I and output O is tested by investigating the mechanisms for the different manifestations of C, i.e., behaviors B_1, B_2, B_3, … B_n (the circles and arrows surrounded by the dotted lines). The behaviors differ in how I and O are operationalized (indicated by small letters i_n and o_n). Some constituents will occur in all mechanisms (grey circles), which are thus *behavior unspecific*. Some constituents will occur in only some (white circles), which are thus *behavior specific*.

Based on the notion of behavior specificity we can answer the CTE: the constitutive background consists of components that are behavior specific whereas the cognitive constituents are behavior unspecific—they are *cognitive* constituents because they are constitutive for a specific *cognitive capacity*. This provides us with the following account of cognitive constituents:

(*cognitive constituent*) X's φ-ing is a *cognitive constituent* of a cognitive behavior ψ-ing of a system S iff:

(i) X's φ-ing is a *spatiotemporal part* of S's ψ-ing,

(ii) X's φ-ing and S's ψ-ing are *mutually manipulable* (and they satisfy *causal specificity*),

(iii) X's φ-ing is *behavior unspecific*.

Based on this account, it can be shown that the arm movements performed by the subjects in Ballard et al.'s study are not cognitive constituents but rather part of the constitutive background. As explained above, in the original study, working memory (i.e., the capacity to recall a certain amount of information a short time after receiving the information) was operationalized as 'subjects visually perceive a pattern of colored blocks and manually produce an instance of the same pattern.' In order to evaluate whether the arm movements belong to the constitutive background or not, we have to test other operationalizations of the capacity to recall a certain amount of information a short time after receiving the information and see whether the arm movements do not occur. One way to change the operationalization is to have subjects verbally report what they would do (and have a speech detecting computer program create the copy) instead of asking them to produce the pattern themselves. Clearly, this behavior would still count as copying behavior which would still be a manifestation of working memory. Also, clearly, in this version of the task, subjects will not use their arms. Hence, the arm movements are behavior specific and thus part of the constitutive background rather than cognitive constituents of the behavior.

What about the saccadic eye movements? Is there an operationalization of working memory that does not involve saccadic eye movements? Assume the copying task is changed such that subjects create copies based on tactile input while being blindfolded. I am not aware of any study testing this, but it is plausible to assume that there will not be any saccadic eye movements. However, it is not clear whether blindfolded subjects would indeed be able to succeed in the copying task. It may turn out that in versions of the task that use non-visual inputs subjects are indeed unable to succeed. Hence, whether the saccadic eye movements are constitutive of the cognitive capacity at hand (working memory) or not, and thus whether they are part of the constitutive background in the mechanism for the specific version of copying

behavior used by Ballard et al. rather than cognitive constituents remains an open empirical question.

Can we find any examples that clearly speak in favor of EC based on this notion of behavior specificity? Consider an example well-known in the EC literature: gestures (Chu & Kita, 2011; Freksa, Olteţeanu, Barkowsky, van de Ven, & Schultheis, 2017; Goldin-Meadow & Alibali, 2013; Wesp, Hesse, Keutmann, & Wheaton, 2001). In his (2010)-book, Andy Clark indeed uses a reasoning in line with the present approach to argue for the claim that gestures can constitute thought. He bases his claim on the following empirical evidence (Clark, 2010, p. 123):

a) We gesture when talking on the phone.
b) We gesture when talking to ourselves.
c) We gesture in the dark when nobody can see.
d) Gesturing increases with task difficulty.
e) Gesturing increases when speakers must choose between options.
f) Gesturing increases when reasoning about a problem rather than merely describing the problem or a known solution.

We only have to add:

g) When gesturing is not allowed, success in problem-solving tasks decreases (Chu & Kita, 2011).

On the assumption that hand movements are spatiotemporal parts of the thinking behavior (see Section 3.2), the evidence provided in statements d), e), and f) corresponds to evidence about top-down manipulability of gesturing via interventions into the thinking behavior. The evidence referred to in statement g) corresponds to evidence about bottom-up manipulability. Statements a), b), and c) might indicate that gesturing is indeed behavior unspecific: independently of the operationalization of the input and output (e.g., via phone vs. via written instructions) gestures occur.

The problem for Clark's reasoning only is that 'thinking' is a much too broad cognitive capacity – cognitive scientists usually deal with much more specific descriptions of cognitive capacities. For example, 'thinking' may include capacities such as problem-solving, decision-

making, spatial navigation, speech processing, and so on. It is very unlikely to find any mechanism that is shared by all these capacities. Indeed, the role of gestures is usually discussed in the context of the investigation of *spatial problem-solving* (Chu & Kita, 2011; Freksa et al., 2017; Hostetter & Alibali, 2007), i.e., the cognitive capacity of solving spatial problems (the input is 'being presented with a spatial problem', whereas the output is 'presenting a solution'). As manifestations of this capacity, researchers investigate behaviors in, for example, mental rotation tasks, route learning tasks, or rotating gear tasks. Indeed, for all of these types of behaviors, the occurrence of co-thought gestures has been documented for different operationalizations of the inputs and outputs (Eielts et al., 2018). This suggests that gestures are indeed cognitive constituents of spatial problem-solving. However, to prove this claim, one has to show that under *all* operationalizations of the inputs and outputs that characterize the capacity of spatial problem-solving gestures occur.

This brings us to one consequence of the present approach: it is much more difficult to prove that something is a cognitive constituent than to show that it is part of the constitutive background. In order to find constitutive background elements, one only has to show that there is *one* operationalization of the capacity that does not involve the constituent at issue. In order to show that a constituent is a cognitive constituent, one has to show that it occurs in *every* operationalization of the cognitive capacity[5]. Since the list of potential operationalizations of a given cognitive capacity might be rather long (though in principle finite), one has to conduct rather many experiments (and then inductively infer that a given constituent is cognitive). It is important to note that this problem arises for defenders as well as opponents of EC. In order to establish that, say, the activity of a certain brain region is a cognitive constituent of a given

[5] This consequence is similar to the implication of Baumgartner and Wilutzky's account according to which constituents can be empirically identified only abductively. However, note that their account is one of mechanistic constituents (vs. non-constituents) in general. Here, we are dealing with an account of cognitive constituents (vs. non-cognitive constituents).

behavior, one has to show that the brain region is active under every operationalization of the relevant cognitive capacity.

A further potential complication for the suggested account of behavior specificity is that many cognitive behaviors are plausibly multiply realized by different mechanisms. Even given the *same* operationalization of the inputs and outputs, there may be different mechanisms that constitute the cognitive behavior connecting them. This should not count as indicating behavior specificity of the mechanistic components—otherwise even components that are uncontroversial cognitive constituents would fail the test. For example, memory behavior may involve brain areas or micro ships implanted into the brain—and both should intuitively come out as cognitive constituents even if they do not occur in all operationalizations of memory behavior. This problem can be solved by highlighting that mechanisms are individuated relative to the behavior of a *specific object (type)* (see Section 2.2). It will not be surprising that, for example, memory mechanisms differ between lay people and memory experts, healthy subjects and patients, children and adults, and humans and non-human animals. Hence, behavior specificity has to be relativized to a specific type of objects.

Finally, a consequence of the present proposal is that whether something is a cognitive constituent or not can only be determined locally. Claims such as "Gesturing is a cognitive process" or "Activity in the pre-frontal cortex is a cognitive process" are false as long as they are not relativized to a specific cognitive capacity. One and the same element can be cognitive relative to one capacity but non-cognitive relative to another. For example, gesturing may be a cognitive constituent in spatial problem-solving but not in social communication.

To summarize: based on the present proposal, we can provide a definition of what cognitive constituents are that incorporates MM, is in line with the general strategy of using the new mechanists' tools to defend EC, is independent of any putative mark of the cognitive, and can be accepted by defenders and opponents of EC alike. Finally, based on the present proposal

we can identify a promising example of embodied cognition: gestures in spatial problem-solving.

5 Conclusion

Three problems for the strategy of applying MM to EC were discussed in this paper: The Inconsistency Problem, the Parthood Problem, and the CTE. I argued that all of these problems can be solved if MM is interpreted along the lines of Krickel (2018), the cognitive phenomena are taken to be cognitive behaviors that trivially involve the body, and MM is combined with a test for what I call *behavior specificity*. Cognitive constituents, apart from satisfying MM, are behavior unspecific in that they occur under every operationalization of the cognitive capacity of which the behavior is a manifestation. Based on the present proposal, a promising case speaking in favor of EC could be found: gestures as cognitive constituents of spatial problem-solving. Based on the account of 'cognitive constituent' developed in this paper, the debate on EC can be revived. The criteria can be accepted by defenders and opponents of EC and allow for a reevaluation of the empirical evidence put forward to argue for or against EC based on a non-question-begging criterion.

Acknowledgments: I thank Ken Aizawa, Albert Newen, Matej Kohár, Julia Wolf, Sabrina Coninx, Pascale Willemsen, Francesco Marchi, Alfredo Vernazzani, Sam Cosper, the members of the DFG funded research training group "Situated Cognition," and two anonymous reviewers for comments on earlier drafts of this paper. Also, I presented the paper at different workshops and conferences where I received helpful feedback. These workshops and conferences included: the symposium *Causation and Computation in Neuroscience* (Jerusalem 2017) organized by Oron Shagrir, Jens Harbecke, and Vera Hoffmann-Kolss, *ECAP* (München 2017), *Evolving Minds – A Workshop with Dan Hutto and Eric Myin* (Bochum 2017) organized by Tobias Schlicht, and the GAP-Satellite Workshop organized by the RTG Situated Cognition

(Cologne 2018). This publication is funded by the DFG-Graduiertenkolleg "Situated Cognition", GRK-2185/1.

References

Abramova, K., & Slors, M. (2018). Mechanistic explanation for enactive sociality. *Phenomenology and the Cognitive Sciences*, 1–24.

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

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Blackwell Pub.

Adams, F., & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Ed.), *The extended mind* (pp. 67–80). MIT Press.

Aizawa, K. (2010). The coupling-constitution fallacy revisited. *Cognitive Systems Research*, *11*(4), 332–342. http://doi.org/10.1016/j.cogsys.2010.07.001

Aizawa, K. (2017). Cognition and behavior. *Synthese*, *194*, 4269–4288. http://doi.org/10.1007/s11229-014-0645-5

Ballard, D. H., Hayhoe, M. M., Pelz, J. B., Ballard, D., Hayhoe, M., Pelz, J., … Yarbus, A. (1995). Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. http://doi.org/10.1162/jocn.1995.7.1.66

Baumgartner, M. (2009). Interventionist Causal Exclusion and Non-Reductive Physicalism. *International Studies in the Philosophy of Science*, *23*(2), 161–178.

Baumgartner, M., & Casini, L. (2017). An Abductive Theory of Constitution. *Philosophy of Science*, *84*(2), 214–233. http://doi.org/10.1086/690716

Baumgartner, M., Casini, L., & Krickel, B. (2018). Horizontal Surgicality and Mechanistic Constitution. *Erkenntnis*. http://doi.org/10.1007/s10670-018-0033-5

Baumgartner, M., & Gebharter, A. (2016). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756. http://doi.org/10.1093/bjps/axv003

Baumgartner, M., & Wilutzky, W. (2017). Is it possible to experimentally determine the extension of cognition? *Philosophical Psychology*, *30*(8), 1104–1125. http://doi.org/10.1080/09515089.2017.1355453

Bechtel, W. (2006). *Discovering Cell Mechanisms*. Cambridge: Cambrdige University Press.

Bechtel, W. (2008). *Mental Mechanisms*. *Philosophical Perspectives on Cognitive Neuroscience*. New York/London: Routledge.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 421–441. http://doi.org/10.1016/j.shpsc.2005.03.010

Bechtel, W., & Richardson, R. C. (2010). *Discovering Complexity. Decomposition and Localization as Strategies in Scientific Research*. Cambridge: MIT Press.

Chu, M., & Kita, S. (2011). The Nature of Gestures' Beneficial Role in Spatial Problem Solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. http://doi.org/10.1037/a0021790

Clark, A. (2010). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press. Retrieved from https://books.google.de/books?id=I-tSEhdEorUC

Clark, A., & Chalmers, D. J. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.
Craver, C. F. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research*, *32*(Section II), 1–20. http://doi.org/10.5840/jpr_2007_4
Craver, C. F. (2007b). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.
Craver, C. F., & Bechtel, W. (2007). Top-down Causation Without Top-down Causes. *Biology & Philosophy*, *22*(4), 547–563. http://doi.org/10.1007/s10539-006-9028-8
Craver, C. F., & Darden, L. (2001). Discovering Mechanisms in Neurobiology: The Case of Spatial Memory. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and Method in Neuroscience* (pp. 112–137). Pittsburgh: University of Pitt Press.
Craver, C. F., & Darden, L. (2013). *In Search of Mechanisms. Discoveries across the Life Sciences*. Chicago/London: University of Chicago Press.
Cummins, R. (1983). *The nature of psychological explanation*. MIT Press. Retrieved from https://mitpress.mit.edu/books/nature-psychological-explanation
Darden, L. (2002). Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science*, *69*(3), S354–S365. http://doi.org/10.1086/341858
Eielts, C., Pouw, W., Ouwehand, K., van Gog, T., Zwaan, R. A., & Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychological Research*, 1–12. http://doi.org/10.1007/s00426-018-1065-9
Eronen, M. I., & Brooks, D. S. (2014). Interventionism and Supervenience: A New Problem and Provisional Solution. *International Studies in the Philosophy of Science*, *28*(2), 185–202. http://doi.org/10.1080/02698595.2014.932529
Freksa, C., Olteţeanu, A.-M., Barkowsky, T., van de Ven, J., & Schultheis, H. (2017). Spatial Problem Solving in Spatial Structures (pp. 18–29). Springer, Cham. http://doi.org/10.1007/978-3-319-69456-6_2
Gallagher, S. (2018). New mechanisms and the enactivist concept of constitution. In M. P. Guta (Ed.), *The Metaphysics of Consciousness* (pp. 207–220). London: Routledge.
Gebharter, A. (2015). Causal Exclusion and Causal Bayes Nets. *Philosophy and Phenomenological Research*, 1–23. http://doi.org/10.1111/phpr.12247
Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Prospect Geights: Waveland Press, Inc.
Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research*, *81*(2), 362–381. http://doi.org/10.1111/j.1933-1592.2010.00375.x
Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's Role in Speaking, Learning, and Creating Language. *Annual Review of Psychology*, *64*(1), 257–283. http://doi.org/10.1146/annurev-psych-113011-143802
Hewitson, C. L., Kaplan, D. M., & Sutton, J. (2018). Yesterday the earwig, today man, tomorrow the earwig? *Comparative Cognition & Behavior Reviews*, *13*, 25–30. http://doi.org/10.3819/CCBR.2018.130003
Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73–95. http://doi.org/10.1075/gest.7.1.05hos
Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge: MIT Press. http://doi.org/10.1017/CBO9781107415324.004
Hutto, D. D., & Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. *Book*. MIT Press. Retrieved from https://mitpress.mit.edu/books/evolving-enactivism
Japyassú, H. F., & Laland, K. N. (2017). Extended spider cognition. *Animal Cognition*, *20*(3), 375–395. http://doi.org/10.1007/s10071-017-1069-7

Kaiser, M. I., & Krickel, B. (2017). The Metaphysics of Constitutive Mechanistic Phenomena. *The British Journal for the Philosophy of Science*, *68*, 745–779. http://doi.org/10.1093/bjps/axv058

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology and Philosophy*, *27*(4), 545–570. http://doi.org/10.1007/s10539-012-9308-4

Kästner, L. (2017). *Philosophy of Cognitive Neuroscience, Causal Explanations, Mechanisms and Experimental Manipulations*. Berlin, Boston: De Gruyter. http://doi.org/10.1515/9783110530940

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, *95*(1), 3–36.

Kirchhoff, M. D. (2015). Extended Cognition & the Causal-Constitutive Fallacy: In Search for a Diachronic and Dynamical Conception of Constitution. *Philosophy and Phenomenological Research*, *90*(2), 320–360. http://doi.org/10.1111/phpr.12039

Kirchhoff, M. D. (2017). From mutual manipulation to cognitive extension: challenges and implications. *Phenomenology and the Cognitive Sciences*, *16*(5), 863–878. http://doi.org/10.1007/s11097-016-9483-x

Kirsh, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*(4), 513–549. http://doi.org/10.1207/s15516709cog1804_1

Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A*, *68*, 58–67. http://doi.org/10.1016/j.shpsa.2018.01.003

Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science*, *63*(2), 399–427. http://doi.org/10.1093/bjps/axr036

Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*(17), 556–567. http://doi.org/10.2307/2025310

Lewis, D. (1986). Events. In D. Lewis (Ed.), *Philosophical Papers Vol. II* (pp. 241–269). Oxford University Press.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking About Mechanisms. *Philosophy of Science*, *67*(1), 1–25.

Newen, A. (2017). What are cognitive processes? An example-based approach. *Synthese*, *194*, 4251–4268. http://doi.org/10.1007/s11229-015-0812-3

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, *24*(5), 939-73; discussion 973-1031. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12239892

Pöyhönen, S. (2014). Explanatory power of extended cognition. *Philosophical Psychology*, *27*(5), 735–759. http://doi.org/10.1080/09515089.2013.766789

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, *192*(11), 3731–3755. http://doi.org/10.1007/s11229-015-0718-0

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–30. http://doi.org/10.3758/s13423-011-0055-3

Rupert, R. D. (2009). *Cognitive systems and the extended mind*. Oxford University Press.

Rupert, R. D. (2010). Systems, functions, and intrinsic natures: On Adams and Aizawa's The Bounds of Cognition. *Philosophical Psychology*, *23*(1), 113–123. http://doi.org/10.1080/09515080903538867

Theiner, G., Allen, C., & Goldstone, R. L. (2010). Recognizing group cognition. *Cognitive Systems Research*, *11*(4), 378–395. http://doi.org/10.1016/J.COGSYS.2010.07.002

van Eck, D., & Looren de Jong, H. (2016). Mechanistic explanation, cognitive systems demarcation, and extended cognition. *Studies in History and Philosophy of Science Part*

A, *59*(Supplement C), 11–21. http://doi.org/10.1016/j.shpsa.2016.05.002

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, *114*(4), 591–600. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11789342

Wilson, R. A. (2004). *Boundaries of the mind: the individual in the fragile sciences*. Cambridge University Press. Retrieved from https://books.google.de/books/about/Boundaries_of_the_Mind.html?id=WXBmuko2CqIC&source=kp_cover&redir_esc=y

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

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318. http://doi.org/10.1007/s10539-010-9200-z

Woodward, J. (2015). Interventionism and Causal Exclusion. *Philosophy and Phenomenological Research*, *91*(2), 303–347. http://doi.org/10.1111/phpr.12095

Zednik, C. (2011). The Nature of Dynamical Explanation*. *Philosophy of Science*, *78*(2), 238–263. http://doi.org/10.1086/659221