Toward a Mechanistic Account of Extended Cognition

Paul R. Smart[a]

[a]Electronics & Computer Science, University of Southampton, Southampton, SO17 1BJ, UK.

ABSTRACT

There have been a number of attempts to apply mechanism-related concepts to the notion of extended cognition. Such accounts appeal to the idea that extended cognitive routines are realized by mechanisms that transcend some salient border or boundary. The present paper describes some of the challenges confronting the effort to develop a mechanistic account of extended cognition. In particular, it describes five problems that must be resolved if we are to make sense of the idea that extended cognition can be understood via an appeal to extended mechanisms. These problems are intended to guide the philosophical effort to develop a mechanistic account of extended cognition.

KEYWORDS

Extended Cognition; Constitutive Relevance; Mechanism; Active Externalism; Mechanistic Explanation

1. Introduction

There have been a number of attempts to apply mechanism-related concepts to the notion of extended cognition (Fazekas, 2013; Kaplan, 2012; Miłkowski et al., 2018; van Eck & de Jong, 2016; Zednik, 2011). Such efforts are typically driven by the idea that theories of mechanistic explanation can be used to resolve disputes and disagreements that have emerged in the active externalist literature. Kaplan (2012), for example, advances the idea that a theory of mechanistic constitution (or constitutive relevance) can be used to demarcate the borders and boundaries of cognitive systems. This, he suggests, promises to progress the argumentative stalemate that has arisen between the supporters of the Hypothesis of Extended Cognition (HEC) (e.g., Clark, 2008) and the supporters of the Hypothesis of Embedded Cognition (HEMC) (e.g., Rupert, 2004).[1]

The appeal to mechanism-related concepts in the active externalist literature establishes a point of contact with work in so-called mechanical philosophy (a specialist area of the philosophy of science) (e.g., Glennan & Illari, 2018). This work highlights the importance of mechanistically-oriented explanatory accounts to a variety of scientific disciplines, including cognitive science (e.g., Craver & Darden, 2013). Of particular importance is the idea that scientists explain phenomena by formulating so-called *mechanistic explanations*, which explain phenomena by describing the *mechanisms* responsible for those phenomena. Mechanisms, themselves, are deemed to consist of *components*, which are the working parts of mechanisms (e.g., Craver, 2015), and one of the goals of mechanistic explanation is to identify these components and describe the way they work together so as to constitute or realize an explanandum phenomenon.[2]

The theoretical resources of mechanical philosophy seem particularly appropriate

CONTACT Paul R. Smart. Email: ps02v@ecs.soton.ac.uk

given the appeal to what are variously referred to as extended (Clark, 2011; Hurley, 2010; Kaplan, 2012; Zednik, 2011), wide (Miłkowski et al., 2018), or supersized (Clark, 2008) mechanisms within the active externalist literature. Relative to current theories of mechanistic explanation, such locutions suggest that extended cognitive phenomena (e.g., extended cognitive processes) are to be understood as phenomena that are constituted (or realized) by mechanisms that qualify as extended mechanisms. These are what we might call *extended cognitive mechanisms*—i.e., mechanisms that realize extended cognitive phenomena. Quite plausibly, an extended cognitive mechanism is a mechanism that qualifies as both a cognitive mechanism and an extended mechanism—that is to say, the class of extended cognitive mechanisms lies at the intersection of two classes: the class of cognitive mechanisms and the class of extended mechanisms (see Figure 1). In the present paper, a cognitive mechanism is conceptualized as a mechanism that constitutes or realizes a cognitive phenomenon, while an extended mechanism is conceptualized as a mechanism whose components are distributed across some kind of border or boundary. In short, an extended mechanism is a boundary-transcending mechanism, where the components of the mechanism are situated either side of this boundary. In the case of human cognition, the nature of this boundary is relatively obvious: it is the biological boundary of the human individual—the borders of skin and skull. In the case of human cognition, then, we encounter a case of extended cognition when we confront (e.g.) a cognitive process that is realized by a mechanism whose components lie beyond the biological borders of the human individual.[3] This form of extended or wide realization typically violates our expectations as to where we expect to find the components of the mechanism responsible for a cognitive phenomenon. Perhaps when we ascribe a cognitive capacity to an entity (such as a human subject) we expect to find the mechanisms responsible for the exercise of that capacity confined to the borders of whatever entity is deemed to possess or 'own' the capacity. In cases of extended cognition, however, these expectations are violated: mechanistic explanations thus reveal the presence of mechanisms whose components lie beyond the borders of the entity (the human individual) that is the focus of our explanatory efforts. The upshot is an appeal to the notion of an extended mechanism and (assuming the explanandum phenomenon qualifies as a cognitive phenomenon) extended cognition.

Such an account provides us with a means of understanding how theories of mechanistic explanation might be applied to extended cognition. In short, it seems perfectly plausible that putative cases of extended cognition could be revealed as part of the attempt to formulate what are called constitutive mechanistic explanations.[4] These explanations seek to explain some target phenomenon P that is ascribed to an entity E by describing the mechanism M that is responsible for P. When P qualifies as a cognitive phenomenon, then M will qualify as a cognitive mechanism. In addition, when at least some of the components of M are revealed to lie beyond the physical borders of E, then M will also qualify as an extended mechanism (see Figure 1). The upshot is that M will qualify as both a cognitive mechanism and an extended mechanism, and it will therefore qualify as an extended cognitive mechanism. Accordingly, we can conclude that P qualifies as an extended cognitive phenomenon.

The present paper seeks to clarify some of the issues raised by this attempt to apply theories of mechanistic explanation to extended cognition. In particular, it identifies five problems confronting the effort to formulate a mechanistic account of extended cognition. These problems are as follows:

(1) **The Problem of Cognitive Status:** What makes an explanandum phenomenon a cognitive phenomenon?

(2) **The Problem of Constitutive Relevance:** How do we know when some material object (a putative component) is constitutively relevant to the explanandum phenomenon?
(3) **The Problem of Cognitive Ownership:** Why do we see an explanandum phenomenon as being associated with (or 'owned by') a particular human individual (or, more generally, a cognitive agent)?
(4) **The Problem of Explanatory Focus:** What phenomenon is the focus of explanatory efforts in putative cases of extended cognizing? In formulating a mechanistic explanation, are we trying to explain the behavior of the human individual (cognitive agent) or the behavior of the larger (extended) organization of which the individual is a part?
(5) **The Problem of Extended Status:** What is it, exactly, that distinguishes an extended mechanism from a non-extended mechanism?

The primary aim of the present paper is to discuss these problems, specify why they are important, and link them to issues that have been raised in the active externalist literature. A consideration of these problems may help to provide a framework that coordinates the philosophical effort to apply mechanism-related concepts to the notion of extended cognition.

For the sake of clarity, it should be noted that the scope of the present paper is limited to the notion of extended cognition, *not the extended mind*. As noted by Pöyhönen (2014), there is a distinction between the Hypothesis of Extended Cognition (HEC) and the Hypothesis of Extended Mind (HEM), with the former featuring an appeal to explanatory kinds of interest to cognitive science (e.g., cognitive processes) and the latter featuring an appeal to explanatory kinds of interest to commonsense or folk psychology (e.g., dispositional belief). Given that theories of mechanistic explanation have their origins in the philosophy of science, it is likely that a mechanistic account will be most appropriate for phenomena (i.e., explanatory kinds) that are the focus of cognitive scientific investigations. This does not mean that a mechanistic account cannot be applied to the extended mind, but, for present purposes, I will limit my attention to the notion of extended cognition.

2. The Problem of Cognitive Status

The *problem of cognitive status* is the problem of determining whether some phenomenon (or mechanism) ought to be counted as a cognitive phenomenon (or a cognitive mechanism). A solution to this problem seems important, since whatever else we might say about the term "extended cognition," it is plausible to assume that the term is being applied to a specific class of phenomena, namely those phenomena (e.g., processes) that qualify as cognitive phenomena.

While some have questioned the need for a precise definition of what it means for a process (or, more generally, a phenomenon) to count as a cognitive process (or cognitive phenomenon) (see Allen, 2017), it is not clear that a mechanistic account of extended cognition can be developed in the absence of some sort of agreement as to what it is that makes a process a cognitive process. Inasmuch as we accept the idea that cognitive mechanisms are to be individuated relative to the phenomena they realize—that is to say, a cognitive mechanism is a mechanism that realizes a cognitive phenomenon—then we will need to have some agreement as to when a phenomenon ought to be regarded as a cognitive phenomenon. If we are unable to do this, then the cognitive status of the

underlying mechanism (the mechanism responsible for the phenomenon) will always be in doubt. In short: If we are unable to specify what it means for something to count as a cognitive phenomenon, then we will have no means to adjudicate the cognitive status of a mechanism that is responsible for the phenomenon. Given the claim that extended cognitive mechanisms are a proper subset of the class of cognitive mechanisms (see Section 1), if we cannot resolve the cognitive status of a mechanism, then we will be left in the dark as to whether or not we confront an extended *cognitive* mechanism.

As a means of reinforcing this point, it is worth noting that one could accept the idea that some phenomena are realized by extended mechanisms, whilst denying that any of those phenomena ought to be regarded as cognitive phenomena. Consider, for example, that both Wilson (2014) and Adams and Aizawa (2001) appear to accept the possibility of extended digestion (broadly understood as digestive processes that are not bounded by the physical body of the organism). Despite this point of agreement, however, they remain on opposite sides of the argument as regards the notion of extended cognition: while Wilson is one of the HEC's most vocal champions, Adams and Aizawa are two of its most ardent critics. The difference, it seems, is not related to the mere idea that some processes can be regarded as extended processes; it is more the issue of whether or not some extended processes ought to be regarded as *bona fide* members of the class of cognitive processes.

The problem of cognitive status is thus not one that can be easily overlooked or bypassed. In particular, it will do no good to simply cite examples of extended phenomena/mechanisms if the proponent of extended cognition cannot also state why such phenomena/mechanisms ought to be regarded as cognitive phenomena/mechanisms. The notion of extended cognition applies to the specific realm of cognitive phenomena, so it is incumbent on the proponent of extended cognition to specify why a given explanandum phenomenon ought to be accepted as a *bona fide* cognitive phenomenon.

Having said all this, it should be noted that the problem of cognitive status is *not* a problem that is specific to the notion of extended cognition. In fact, the problem of cognitive status is a problem that pertains to *all* cognitive phenomena, not just cognitive phenomena of the extended variety. Although the problem of cognitive status is a prominent topic of debate in active externalist circles, it is, in fact, a more general problem in the philosophy of mind and the philosophy of cognitive science.

Within philosophical circles, the attempt to identify the characteristic features of cognitive phenomena (e.g., the things that make a particular process a cognitive process) is typically referred to as the search for a "mark of the cognitive" (Adams, 2010; Adams & Garrison, 2013). While this mark of the cognitive problem is clearly related to the problem of cognitive status, it is important to appreciate that they are not the same. A solution to the mark of the cognitive would undoubtedly serve as a solution to the problem of cognitive status, but it is possible that we could have a solution to the problem of cognitive status without resolving the mark of the cognitive problem. To help us see this, it is worth bearing in mind that the problem of cognitive status is the problem of determining whether a given phenomenon (e.g., a process) counts as a cognitive phenomenon (e.g., a cognitive process); it is not the problem of providing a general philosophical account of what distinguishes cognitive from non-cognitive phenomena. Given that some phenomena (e.g., learning, memory, perception, reasoning) seem to be regarded as paradigmatically cognitive, even by those who are opposed to extended cognition (Adams & Aizawa, 2001, p. 48), it is possible that the problem of cognitive status could be resolved by directing attention to phenomena whose cognitive status is not in doubt. Consider, for example, that if all mnemonic phenomena are regarded as cognitive phenomena, and we are also able to state what it is that makes

a given phenomenon a member of the class of mnemonic phenomena, then we may be able to adjudicate the cognitive status of a given extended phenomenon simply by recognizing that it is a *bona fide* member of the class of mnemonic phenomena. This, of course, requires consensus on what it is that makes a phenomenon a mnemonic phenomenon, but understanding the properties of this class (the mark of the mnemonic) may be somewhat easier than the effort to understand what it is that unites disparate kinds of cognitive phenomena under a common conceptual umbrella.

A popular response to the problem of cognitive status comes in the form of the *parity principle*. According to the parity principle:

> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing [it] as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. (Clark & Chalmers, 1998, p. 8, original emphasis)

As noted by Wheeler (2011), the parity principle is subject to a persistent misreading, which tends to tie issues of parity to issues of functional equivalence. It is thus a mistake to construe the parity principle as requiring the extra-cerebral portions of an extended cognitive mechanism to somehow function in a way that resembles the internally-situated portions. Nor is there any reason to think that the extra-cerebral portions need to work in a way that is different from (or complementary to) the inner portions (see Sutton, 2010). This may or may not be true in specific cases, but issues of complementarity have no bearing on the problem of cognitive status. (Nor, as far as I can tell, do they have any bearing on the other problems discussed below.) The upshot is that the notion of a *complementarity principle* for extended cognitive systems (e.g., Sutton, 2010), while a popular feature of philosophical debates, is unlikely to be a central feature of a mechanistic account of extended cognition.

As noted by Clark (2011), the intended purpose of the parity principle is to guide our intuitions about cognitive status by asking us to imagine what we would say about a phenomenon if such a phenomenon was to be found inside the head of a given agent:

> [The idea] was not that external stuff must work in much the same way as inner stuff if cognition is to depend on extended mechanisms. Rather it was to probe how we would treat the functional analogues of certain external contributions were they (appropriately) internally relocated. (Clark, 2011, p. 451)

Thus construed, the parity principle is best seen as an appeal to *equality of opportunity* rather than an appeal to functional equivalence. The intended purpose of the parity principle is thus to avoid the sort of bias that might be incurred if we were to judge all phenomena relative to our neurocentric intuitions about the way in which (human) cognitive phenomena are typically realized. Cast in this light, the goal of the parity principle is to encourage us to treat biological or metabolic factors as something akin to a "protected characteristic" in equality legislation (Malleson, 2018). In deciding whether or not a given phenomenon ought to be regarded as a cognitive phenomenon we are being asked to, in effect, blind ourselves to the material nature of the resources that are involved in the mechanistic realization of the phenomenon. As noted by Clark (2011, p. 449), the idea behind the parity principle "was simply to invite the reader to judge various potential cognitive extensions behind a kind of 'veil of metabolic ignorance'." One way to do this (as the parity principle suggests) is to imagine that some candidate cognitive process P is occurring inside the head of a human agent. If, in this situation, we are content to accept P as a cognitive process, then the claim is that we should accept the cognitive status of P even if the mechanism

that realizes P is one that includes extra-neural (or extra-organismic) constituents.

Despite these clarifications, the parity principle has proved to be a highly controversial approach to resolving the problem of cognitive status. In general, there is no widespread agreement on what it is that makes a given phenomenon a specifically cognitive phenomenon, and the problem of cognitive status thus remains unresolved.

3. The Problem of Constitutive Relevance

As we have seen, a mechanistic account of extended cognition appeals to the idea that extended cognitive routines are realized by extended mechanisms, which, in the case of human cognition, are mechanisms that transcend the borders of skin and skull. This means that we need to know something about the structural organization of a mechanism. In particular, we need to know what the components of a mechanism are and where these components are located relative to some border or boundary. Within mechanical philosophy, the problem of determining what parts of the material world ought to be included in a mechanism is referred to as the *problem of constitutive relevance* (Craver, 2007a, 2007b). In short, the problem of constitutive relevance is the problem of picking out the components of a mechanism (i.e., the entities/activities that should be included in a mechanistic explanation of a phenomenon). By identifying these components, we have a means to determine whether or not some extra-organismic resource ought to be included in the mechanism that realizes a cognitive phenomenon. This, it should be clear, is a central issue in debates about extended cognition. Indeed, the problem of constitutive relevance is often the primary point of contention in debates about the respective merits of extended or embedded approaches to cognition (see Kaplan, 2012).

Before going further, it is worth noting that the problem of constitutive relevance is a problem that pertains to *all* mechanisms; it is not a problem that is specific to the class of extended cognitive mechanisms, or even the more generic class of cognitive mechanisms. Instead, the problem of constitutive relevance is one that applies to mechanisms of any stripe, whether these be social mechanisms, astrophysical mechanisms, computational mechanisms, and so on. This point is important, for the active externalist literature is littered with attempts to address the problem of constitutive relevance in a manner that is specific to the realm of cognitive systems (and human cognitive systems, in particular). To my mind, a solution to the problem of constitutive relevance will need to be applicable to *all* mechanisms, regardless of whether or not they qualify as cognitive mechanisms. We thus ought to embrace something along the lines of a regulative ideal for the problem of constitutive relevance. For the sake of convenience, let us refer to this as the *ideal of general applicability*. According to this ideal, for any proposed solution to the problem of constitutive relevance, we ought to ask ourselves, "Would this account make any sense if it were to be applied to mechanisms beyond the disciplinary borders of cognitive science?" If the answer to this question is "no," then, in all likelihood, we are not tackling the problem of constitutive relevance. Either that, or we are not providing an acceptable solution to the problem of constitutive relevance.

Appreciating the generic nature of the problem of constitutive relevance is important if we are to avoid confusion and misunderstanding in future philosophical work. Consider, for example, the criteria proposed by Clark and Chalmers (1998) as part of their seminal treatment of the extended mind. These criteria are what are now known as the trust+glue criteria. They involve appeals to the availability, accessibility, and trusted status of a bio-external resource (or the informational deliverances of such a resource).

According to Clark (2010, p. 46), these criteria were offered as "a rough-and-ready set of additional criteria to be met by nonbiological candidates for inclusion into an individual's cognitive system."

Let us assume that these criteria are intended to resolve the problem of constitutive relevance—the criteria are, in short, intended to guide our intuitions as to whether or not some extra-organismic resource ought to be included as a component in an extended cognitive mechanism. Now, inasmuch as we adhere to the regulative ideal outlined above, it should be clear that something has gone awry, for the criteria make no sense if we apply them to mechanisms that lie outside the disciplinary borders of cognitive science. Suppose we apply these criteria to a conventional automobile engine. Clearly, it makes no sense to insist that the piston must trust the crankshaft in order for the piston to count as a *bona fide* component of the automobile's propulsion mechanism. Nor is it particularly clear how we are to interpret the appeal to issues of accessibility and availability in the context of such a mechanism. This is not to say that the trust+glue criteria are incorrect or inappropriate; it is simply to note that they cannot be applied to the problem of constitutive relevance in the specific context of a mechanistic account of *extended cognition*. This is important, for the trust+glue criteria were originally developed in the context of arguments for the extended mind (see Clark & Chalmers, 1998). As noted in Section 1, however, the scope of the present paper is limited to an ostensibly different notion, namely, the notion of extended cognition. Accordingly, it is perfectly possible that the trust+glue criteria are playing an important and valid role in philosophical arguments for the extended mind, but it is at best unclear whether these criteria should be imported into a mechanistic account of extended cognition.

A second point to note about the problem of constitutive relevance is that it is largely unrelated to the problem of cognitive status. There is, in particular, no reason to think that a solution to the problem of constitutive relevance will tell us anything useful about the problem of cognitive status. We could, for example, have a complete account of constitutive relevance and thus be able to determine the borders and boundaries of a mechanism. But this is unlikely to be of much use when it comes to evaluating the cognitive status of a mechanism (i.e., whether or not a mechanism ought to be regarded as a cognitive mechanism). A complete account of constitutive relevance might tell us whether or not an extra-organismic resource is part of some discernible mechanism that is centered on an individual cognitive agent. By itself, however, this will not tell us whether the mechanism is responsible for a specifically cognitive phenomenon (as opposed to a phenomenon of some other type). In short, there is little reason to think that the search for a "mark of the cognitive" (see Adams & Garrison, 2013) can be substituted with a search for the "mark of the constitutive." The problem of cognitive status and the problem of constitutive relevance are likely to be independent problems.

One of the proposed solutions to the problem of constitutive relevance comes in the form of an intervention-based criterion dubbed the mutual manipulability criterion (see Kaplan, 2012). This criterion draws on what is perhaps the most popular account of constitutive relevance in mechanical philosophy, namely, the mutual manipulability account (Craver, 2007a, 2007b). Unfortunately, the viability of the mutual manipulability account (and thus the associated appeal to a mutual manipulability criterion) has been challenged by the results of recent philosophical work (Baumgartner & Gebharter, 2016). These results have also been applied to the case of extended cognition, leading to doubts about the extent to which a mutual manipulability criterion can be used to resolve disputes relating to extended cognition (Baumgartner & Wilutzky, 2017). In contrast to Kaplan's upbeat assessment of the prospects for resolving the problem of

constitutive relevance, the conclusions of Baumgartner and Wilutzky (2017) strike a more pessimistic note. In particular, they suggest that it is, in principle, "impossible to experimentally determine whether cognitive processes have extracerebral constituents" (p. 1104).

At the present time, there is no consensus on a philosophical account of constitutive relevance, and the problem of constitutive relevance thus remains unresolved. It is unclear whether more recent accounts of constitutive relevance, such as the causal situationist account (Prychitko, 2021), might be able to resolve the problems that afflict the mutual manipulability account. Also unclear is the extent to which the problem of constitutive relevance requires an interventionist (or intervention-based) solution. Many extant accounts of constitutive relevance appeal to the role of experimental interventions in resolving issues of constitutive relevance. In general, however, scientists have a variety of means of studying mechanisms, and not all of these rely on the use of experimental interventions. Astrophysicists, for example, do not apply experimental interventions to astrophysical phenomena, and yet such scientists still seem able to describe the mechanisms responsible for such phenomena (see Illari & Williamson, 2012). In other contexts, the problem of constitutive relevance looks to be insignificant. In engineering, for example, mechanisms are designed and built so as to realize certain kinds of functionality. It is hard to see why the problem of constitutive relevance would arise in such contexts, for engineers arguably have a good understanding of the componential structure of the mechanisms that they themselves create. At the very least, it is difficult to see how the results of experimental interventions would inform the engineer's understanding of what is to count as a component within a given mechanism. An automobile engineer already knows that the piston is a component in a propulsive mechanism, and no amount of experimental intervention is likely to change this. Accordingly, it seems that there might be multiple routes to resolving the problem of constitutive relevance. This is important given the conclusions reached by Baumgartner and Wilutzky (2017, p. 1104) regarding the role of experimental techniques in determining the extended status of cognitive routines. While Baumgartner and Wilutzky (2017) may be correct in their critique of the mutual manipulability account, we should not assume that the failure of one specific account of constitutive relevance means that it is impossible to determine the componential ingredients of a putatively extended cognitive mechanism. Nor should we assume that the use of interventionist-style experimental techniques is a prerequisite for resolving the problem of constitutive relevance.

4. The Problem of Cognitive Ownership

Many of those who are familiar with the active externalist literature are likely to be aware of the problem of cognitive status and the problem of constitutive relevance. The next problem, however, has been the subject of somewhat less discussion. This is the *problem of cognitive ownership*. The problem concerns the way in which we see a cognitive phenomenon (e.g., a cognitive state or process) as, in some sense, belonging to a particular (cognitive) agent. This looks to be largely unproblematic in the case of non-extended cognition. If, for example, a cognitive process is performed inside the head of a human individual, then it is relatively clear that the process belongs to the human individual. It is, after all, the human individual who is performing the relevant process. Matters are not so clear-cut in the case of extended cognition. This is because some of the components of the mechanism that realizes the relevant cognitive

routine lie outwith the biological borders of the individual. Nevertheless, in cases of extended cognition, it is common to ascribe ownership of an extended cognitive routine to a particular individual. Consider, for example, the use of pen and paper resources to solve long multiplication problems. This world-involving variant of the long multiplication routine is sometimes presented as a form of extended cognizing due to the fact that extra-organismic resources are being used to perform a recognizably cognitive process (see Wilson & Clark, 2009). The world-involving variant of the long multiplication routine thus differs from the non-extended (in-the-head) version, in the sense that the former relies on the use of extra-organismic props, aids, and artifacts, while the latter relies on resources (e.g., the brain) that are internal to the biological borders of the human individual. Despite these differences, however, we tend to see the relevant routine as being tied to a particular individual in both cases. Thus, in both the extended and non-extended cases, we see a distinct cognitive agent as, in some sense, 'owning' the relevant cognitive process, at or least being responsible for the outcome (e.g., the success or failure) of the process.

The problem of cognitive ownership is an important problem for the proponents of extended cognition. As noted by Clark (2011), a great deal of work in the area of extended cognition (and the extended mind):

> ... is best seen as an investigation of... conditions which must be met so as to ensure the *proper ownership* of some candidate extended process by a distinct cognitive agent.... (Clark, 2011, p. 454; original emphasis)

Despite its importance, however, the problem of cognitive ownership is seldom recognized as a distinct problem, i.e., one that is to be tackled independently of the problem of cognitive status and constitutive relevance. Some insight into the problem is revealed by a consideration of mechanisms that involve multiple cognitive agents, such as social mechanisms (see Ylikoski, 2018) or distributed cognitive mechanisms (Hutchins, 1995). Such mechanisms resemble an extended cognitive mechanism in the sense that the mechanism includes resources (i.e., components) that lie external to the biological borders of a given individual. Nevertheless, it doesn't seem appropriate to regard such mechanisms as in any way extended. We do not, for example, talk of social mechanisms as being extended simply because they are constituted by multiple human individuals. Similarly, I suspect it is a mistake to confuse the notion of distributed cognition with the notion of extended cognition. To help us see this, let us direct our attention to a paradigmatic example of distributed cognition: the case of ship navigation. According to Hutchins (1995), the processes supporting navigational efforts aboard a large maritime vessel are ones that exploit a distributed nexus of biological and non-biological resources. Such resources include multiple human individuals and a rich array of material artifacts. From a mechanistic standpoint, we might say that these resources work together to form a distributed cognitive mechanism that contributes to the navigational performances of the ship. But does this distributed cognitive mechanism also qualify as an extended cognitive mechanism? In my view, the answer is "no," and the reason for this relates to the difficulty in ascribing ownership of the larger navigational process to one of the components (e.g., a human individual) of the mechanism that realizes the navigational process. In particular, it does not make sense to say that the larger routine 'belongs' to one of the components of the distributed cognitive mechanism, or that the routine is somehow owned by that component. Instead, the routine appears to be a property of the larger systemic organization in which the distributed cognitive mechanism is situated (i.e., the ship). The problem here is that there seems little reason to think that the mechanism responsible for the navigational

process (the explanandum phenomenon) ought to be regarded as a *bona fide* extended mechanism, for there is no sense in which the boundaries of the mechanism transcend the borders of the system (the entity) to which the routine is ascribed (i.e., the ship). In this case, we could very well have a solution to the problem of cognitive status (we can be sure that the explanandum phenomenon—the navigational process—qualifies as a cognitive process) and a solution to the problem of constitutive relevance (we know what all the components of the distributed cognitive mechanism are), but unless we are able to somehow tie the explanandum phenomenon to a particular human individual (or, more generally, a particular component/entity within the mechanism), then it is unclear why we ought to regard the mechanism as a mechanism of the extended variety.

As a means of pressing this particular point, let us consider the notion of human-extended machine cognition (Smart, 2018). Human-extended machine cognition is a form of extended cognition in which one or more human agents are incorporated into the cognitive routines of a computational–cognitive system, such as an Artificial Intelligence (AI) system. In this case, we encounter a form of role reversal, such that the entity that we would ordinarily see as the target of cognitive incorporation (e.g., a computational device or system) is, instead, the entity that does the incorporating. In conventional cases of extended cognition, it is the computer (*qua* extra-organismic resource) that is incorporated into the cognitive routines of a human agent (and the routines thus belong to the human agent) (see Figure 2a). This contrasts with the state-of-affairs in human-extended machine cognition (see Figure 2b). In this case, it is the human agent that is incorporated into the cognitive routines of the AI system (presumably, then, the routines belong to the AI system or machine agent).

The thing that is important to note here is the *similarity* of the mechanisms in Figure 2. Both mechanisms feature the same components, and they could very well be identical with regard to their causal structure. In the extreme case, the two mechanisms could be indistinguishable with respect to their structural connectivity and patterns of information flow. And yet despite this mechanistic similarity, the two cases are evidently not the same: one is a case of conventional (human-centered) extended cognition (Figure 2a); the other is a case of human-extended machine cognition (Figure 2b). How, then, are we to discriminate between the two cases? This, in a nutshell, is the problem of cognitive ownership.[5]

How this problem is to be resolved remains unclear, although previous efforts have focused on issues of creation (who or what created an extended mechanism) (see Clark, 2008), control (who or what controls the flow of information within an extended cognitive mechanism) (see Wilson, 2004), and responsibility (who or what ought to be credited with the outcome of an extended cognitive process) (see Roberts, 2012). A solution to the problem is important, for it should be clear that the problem of cognitive ownership has a bearing on the extent to which it makes any sense to talk of extended cognitive mechanisms. If, for example, cognitive processes are always owned by the system that performs a cognitive process, then it seems that extended cognitive processes will always belong to the larger (extended) cognitive system that is brought into existence courtesy of the human individual's interactions with extra-organismic resources. The problem here is that the mechanisms responsible for the ('extended') cognitive process will be contained within the borders of the ('extended') system to which the cognitive process belongs. Accordingly, it becomes unclear why we need to talk of extended mechanisms. Inasmuch as extended cognitive processes belong to extended cognitive systems, then the mechanisms that realize such processes are not really extended, for such mechanisms do not extend beyond the borders of the thing (the 'extended' system) to which an explanandum phenomenon is ascribed.

5. The Problem of Explanatory Focus

It might be thought that if we were able to provide a satisfactory answer to the problems of cognitive status, constitutive relevance, and cognitive ownership, then we would be in a good position to judge the extended status of a particular cognitive routine. If, for example, the world-involving variant of the long multiplication process qualifies as a *bona fide* cognitive process, the extra-organismic resources are constitutively relevant to the process, and the process is seen to belong to a given human individual, then perhaps we ought to regard the long multiplication process as a *bona fide* case of extended (mathematical) cognizing.

Our problems, however, are not quite over, for I suspect that a mechanistic account of extended cognition entails a further problem, one that is distinct from (although not entirely unrelated to) the problems discussed thus far. This particular problem relates to the phenomena that are targeted by specific explanatory efforts. It is what I will call *the problem of explanatory focus*.

To help us understand this particular problem, let us look at the way an extended theorist might approach the world-involving variant of the long multiplication process. The extended theorist, we can assume, is already sensitized to the possibility of wide or extended realization bases, and they will thus direct their attention to the way in which some ostensibly cognitive routine is realized courtesy of the use of extra-organismic resources. For the extended theorist, then, the thing that is to be explained in the long multiplication case is the long multiplication process as it occurs *in the world*. For the extended theorist, this world-involving multiplicative routine is what counts as the explanandum phenomenon, and the resources that are constitutively relevant to this phenomenon include the human individual and the extra-organismic resources. These are the components of the mechanism that are responsible (in a constitutive sense) for the world-involving variant of the long multiplication routine.

Now let us look at matters from the standpoint of the embedded theorist. The embedded theorist may very well accept whatever criterion of constitutive relevance is being adopted by the extended theorist, but there is no guarantee that their explanatory interests will converge with those of the extended theorist. Given the crucial role of the human agent in selecting, creating, and manipulating external resources, not to mention their role in coordinating the flow of information between these resources as part of the performance of the long multiplication task,[6] the embedded theorist may feel inclined to limit their explanatory focus to the behavioral performances of the human subject. Accordingly, in the attempt to provide a mechanistic explanation of the cognitive processes associated with long multiplication, the embedded theorist might insist that we ought to direct our attention to the mechanisms that support the expression of adaptive behavioral responses at various junctures in the long multiplication process. In this case, we witness a shift in explanatory focus: rather than our attention being directed to the larger process that involves the use of pen and paper resources, our attention instead zooms in to focus on the actions of the individual. Such responses are undoubtedly driven by cognitive routines—they are, after all, intelligent responses to intermediate problem states—but the routines in question are *not* ones that extend beyond the biological borders of the individual, and they do not, therefore, qualify as extended. The upshot is a non-extended account of mathematical cognizing: The phenomena to be explained in the long multiplication case are the cognitive processes that culminate in the generation of particular behavioral responses (e.g., task-relevant scribblings). These processes are realized by mechanisms that are internal to the biologically-bounded human individual, and we do not, therefore, confront a genuine

case of extended cognizing.

The problem here, it should be clear, is not so much one of constitutive relevance as it is one of determining the proper *focus* of explanatory attention. Mechanisms are typically assumed to be interest-relative, in the sense that the borders and boundaries of mechanisms vary according to the 'location' of our explanatory interests (Craver, 2007b; Glennan, 1996).[7] (This is what is sometimes referred to as *Glennan's Law*). In the long multiplication case, if our attention is directed to the larger process involving pen and paper, then we will see a mechanism that extends beyond the biological borders of the human individual. If, however, our attention is directed to the behavioral responses of the human individual, then we will observe a purely internal mechanism (i.e., a mechanism that fails to breach the biological borders of skin and skull).

The issue at the heart of the problem of explanatory focus is thus revealed: As noted by Craver (2007a) and others, there are no mechanisms simpliciter; mechanisms are always *for* something. The borders and boundaries of mechanisms will thus vary according to our explanatory interests. If what we want to explain in the long multiplication case is the agent's manifest ability to manipulate bio-external resources in such a way as to pave the route to a successful mathematical outcome, then we will never encounter an extended mechanism, and we will thus never observe an extended cognitive routine. This is likely to remain the case, I think, whatever else is true about the relationship between the human agent and bio-external artifacts. Within the active externalist literature, one sometimes finds an appeal to issues of continuous reciprocal causation and/or continuous mutual interaction (e.g., Palermos, 2014) as a means of motivating claims for cognitive extension. Unfortunately, however, I very much doubt that such appeals can be used as a safeguard against the problem of explanatory focus. The reason for this is that appeals to continuous reciprocal causation are typically intended to resolve the problem of constitutive relevance; in particular, they are intended to motivate the idea that two or more entities are wired up in such a way as to legitimate claims about their inclusion in a common mechanism. Issues of constitutive relevance are, however, not really the problem here. The problem is more one of how the causal structure of the world is carved up according to our explanatory interests. If the phenomena we seek to explain are tied to the operation of one of the components of an otherwise integrated mechanism, then the contributions from the other components within that mechanism will be deemed to be of causal significance (or causal relevance) only; that is to say, they will *not* be deemed to be constitutively relevant to the phenomenon that is now at the center of our explanatory concerns.

It is important to note that the problem of explanatory focus is not one that hinges on an intellectual affiliation to extended or embedded cognition. Suppose, for example, that we are fully paid up members of the active externalist club, and we thus accept that long multiplication is an extended cognitive process. Now suppose that we are asked to explain some phenomenon that is specific to the human individual (e.g., the human's perceptuo-motor routines). In this case, our explanatory attention will shift to focus on mechanisms that are (in all likelihood) contained within the biological borders of the human subject. Such mechanisms, it should be clear, will not count as extended. Relative to such mechanisms, the bio-external environment will be seen to be of mere causal relevance to the *inner* cognitive goings-on, and this is despite the fact that *nothing about the causal structure of the world needs to have changed during the shift from one explanandum phenomenon to another*. We could thus still confront a situation where the human agent is engaged in a form of reciprocal causal exchange with elements of the extra-organismic environment, but the mere presence of this reciprocal causal loop does not mean that extra-organismic resources ought to

be seen as the components of *every* cognitive phenomenon that is associated with (or 'owned by') a given human individual.

What is important here is the idea that mechanisms are individuated according to our explanatory interests, and thus the same region of the world can be carved up in different ways according to the kinds of phenomena we seek to explain. The embedded theorist could thus accept the general idea that (in the long multiplication case) there is a form of reciprocal influence between the human agent and the resources of the extra-organismic environment. They can also accept the idea that these material objects probably constitute some larger bio-technological or bio-artifactual system. This does not mean, however, that they are also obliged to regard the activity of that larger bio-technological or bio-artifactual system as a fitting target for *cognitive scientific* explanation. If what the embedded theorist seeks to explain is the adaptive behavioral responses of the human subject at particular junctures in the world-involving variant of the long multiplication task, then, in all likelihood, they will end up with a (constitutive) mechanistic explanation that makes no reference to the extra-organismic elements. This will be true, regardless of whether or not the human agent is reciprocally coupled to some part of the extra-organismic environment.

All of this highlights the importance of clarity when it comes to describing the phenomena that are to be explained by constitutive mechanistic explanations. This is particularly important for a mechanistic approach to extended cognition, and it ought to be regarded as a regulative ideal for future philosophical work in this area. (For the sake of convenience, let us refer to it as the *ideal of explanatory clarity*.) Adherence to this ideal is important, for I suspect that much of the tension between the proponents of extended and embedded cognition can be traced to a difference in explanatory focus. Inasmuch as this is the case, then there is nothing particularly remarkable about the division between extended and embedded theorists, for the same disagreements will arise in any context where the target phenomenon is ill-defined. If scientists (or engineers) are trying to discover (or build) the mechanisms for a given phenomenon, then the mechanisms they discover (or build) will vary according to the phenomenon they are trying to explain (or implement).

Suppose we accept the ideal of explanatory clarity. The question that arises is what counts as the proper focus of explanatory efforts in putative cases of extended cognizing? In the long multiplication case, for example, we can choose to limit our explanatory attention to the individual human agent, or we can opt to broaden our explanatory focus to encompass the non-biological elements of the task environment. But which of these phenomena ought to be the target of our mechanistically-oriented explanatory efforts?

The choice, I suspect, will be governed by what we regard as the proper focus of attention for *cognitive science*. At this point, however, we begin to see the links between the problem of explanatory focus and some of the other problems that confront a mechanistic account of extended cognition. The proponent of extended cognition needs to tread carefully here, for some of the 'extension-friendly' solutions adopted for these earlier problems (most notably the problem of cognitive status and the problem of cognitive ownership) are apt to influence the sort of things we deem worthy of cognitive scientific scrutiny. This tension is perhaps most easily demonstrated by the problem of cognitive status. Suppose, for example, that we accept the idea that the realm of the cognitive ought to be individuated with respect to whatever it is that gives rise to behavior that is deemed "appropriate, adaptive, flexible and coordinated with respect to environmental and organismic circumstances" (Smart, Clowes, & Heersmink, 2017, p. 11). If this is what we mean by a cognitive process (i.e., a process that

produces intelligent behavior), then our attention is perhaps already leaning towards a non-extended take on the long multiplication case. This is because the 'source of the smarts' for the system that performs the long multiplication process seems to relate to the performances of the human subject. In fact, it is not particularly clear that it makes much sense to talk of the behavior of the larger systemic organization (i.e., the system comprising the human + pen + paper) as particularly flexible or adaptive, and inasmuch as these sorts of behaviors are observed, then they can probably be traced to the actions of the human individual. This does not mean that we are obliged to deny the existence of the larger systemic organization (or the mechanisms that explain its behavior), but whether that larger system is the proper target of *cognitive scientific* attention (as opposed to, let's say, the cognitively-driven responses of the human individual) is another matter.

The problem of explanatory focus may also be informed by solutions to the problem of constitutive relevance.[8] As noted by a number of philosophers, scientists sometimes revise their understanding of phenomena as a result of their efforts to delineate the mechanisms responsible for those phenomena (Bechtel & Richardson, 1993/2010, chap. 8). This reveals a means by which the problem of explanatory focus could be linked to the problem of constitutive relevance: A solution to the problem of constitutive relevance may help us to delineate the mechanism responsible for a given phenomenon, but, in delineating this mechanism, we may be obliged to revise (or, as Bechtel and Richardson call it, "reconstitute") the phenomenon that was the initial focus of explanatory attention.[9]

6. The Problem of Extended Status

We come, at last, to our final problem: *the problem of extended status*. This problem relates to the distinction between extended and non-extended mechanisms. This distinction is important because we need some way of discriminating between cognitive mechanisms of the extended and non-extended variety. Recall the conceptual scheme illustrated in Figure 1. This figure depicts extended cognitive mechanisms as a proper subset of the class of cognitive mechanisms. The class of cognitive mechanisms is, in turn, defined with respect to the phenomenon that a mechanism realizes or constitutes. That is to say, if a mechanism M is deemed to be responsible (in a constitutive sense) for a phenomenon P, and P qualifies as a cognitive phenomenon (according to functional criteria), then M will be categorized as a *bona fide* cognitive mechanism. A cognitive mechanism is thus individuated according to functional criteria, which looks to be important if we are to allow for the possibility that functionally similar cognitive phenomena (e.g., the extended and non-extended versions of the long multiplication process) might be realized by structurally and/or materially distinct physical mechanisms.[10]

This way of thinking about cognitive mechanisms suggests that there is nothing special about extended cognitive mechanisms in respect of their *cognitive status*. That is to say, we cannot discriminate between extended and non-extended cognitive mechanisms by focusing our attention on the functional characteristics of the phenomena they realize; instead, we need to direct our attention to the mechanisms themselves. In the long multiplication case, for example, the world-involving variant of the long multiplication routine is functionally similar to the in-the-head version. Despite this similarity, however, the two phenomena are not the same: the world-involving variant of the routine is a candidate case of extended cognizing, while the in-the-head version of the

routine is not. The reason for this difference does not relate to the abstract functional characteristics of the routine itself; rather, it relates to the nature of the mechanism that constitutes/realizes the routine. Thus, in the extended (or world-involving) version of the long multiplication process we observe that some of the components of the mechanism responsible for this particular phenomenon lie external to the biological borders of the individual.

At this point, it should be clear that the distinction between extended and non-extended cognitive phenomena has to be made at the level of the mechanisms that realize the phenomena. In short, we require some means of discriminating extended from non-extended mechanisms. Extended cognitive phenomena, we may suppose, are realized by extended mechanisms, and these mechanisms possess features that distinguish them from non-extended mechanisms. But what exactly are those features? What is it, exactly, that makes some mechanism (cognitive or otherwise) a member of the class of extended mechanisms? In Section 1, I suggested that we can think of an extended cognitive mechanism as a form of boundary-transcending mechanism. That is to say, what makes something an extended cognitive mechanism (or, more generally, an extended mechanism) is the fact that there is some sort of boundary by which we judge the mechanism to be extended. We thus encounter an extended mechanism when the components of a mechanism are located either side of a particular boundary. This boundary, let us call it B, is what makes an extended mechanism a boundary-transcending mechanism. The key issue, then, for the problem of extended status is to identify the boundary by which we judge a mechanism to be extended. The problem of extended status is, in short, the problem of identifying B.

This issue will no doubt sound trivial to many of those who are familiar with the active externalist literature. The answer to the question "what is the relevant boundary by which we judge a mechanism to be an extended mechanism?" looks to be straightforward: It is, of course, the traditional biological boundary of the human individual—the borders of skin and skull.

Things, however, are not quite so straightforward. Perhaps such a response could be made to work if we were to restrict our attention to the realm of *human* cognizing. But what if we want a more general account of extended cognition, one that applies to multiple kinds of intelligent system? This looks to be important given that the notion of extended cognition is sometimes invoked in situations that do not involve human agents. These include cases involving non-human animals (e.g., Japyassú & Laland, 2017), plants (Parise, Gagliano, & Souza, 2020), and AI systems (e.g., Smart, 2018). The latter cases—the ones involving AI systems—are particularly problematic for the idea that extended mechanisms ought to be individuated according to the presence of a biological (e.g., skin/skull) boundary. An intelligent robot could exist as a purely technological entity with no discernible biological boundary. Nevertheless, it seems odd to think that such a system would be unable to benefit from any form of extended cognizing.

To help us see this, suppose we were to engineer a humanoid robot that solved long multiplication problems in precisely the same way as its human counterpart (i.e., via the use of pen and paper resources). Suppose, also, that we accept that the world-involving variant of the long multiplication task qualifies as a form of extended cognitive processing. Wouldn't this suggest that the robot was engaged in a form of extended cognizing despite the fact that it lacked a biological boundary? More generally, if we are willing to accept the existence of extended cognitive mechanisms in the case of human cognition, then why should we demur from the existence of such mechanisms in the case of machine cognition? After all, according to the proponents of

extended cognition, we ought to judge putative cases of extended cognition behind a 'veil of metabolic ignorance' (see Section 2). If, however, we don this veil every time we encounter an intelligent system, then there seems little reason to think that we will always be able to discern a metabolic/biological boundary once the veil is lifted and the biological/non-biological nature of the target system is revealed.

Here is another reason to doubt that a simple appeal to a biological boundary will suffice for the problem of extended status. Within the active externalist literature, one sometimes encounters talk of *socially-extended cognition* (e.g., Lyre, 2018), which is a form of extended cognition involving other human individuals. The cognitive wherewithal of one human agent H is thus deemed to be extended courtesy of the fact that other human agents are incorporated in H's cognitive routines. The problem here should be obvious: Inasmuch as socially-extended cognition is realized by an extended mechanism, then it seems reasonable to conclude that the extended mechanism will transcend the biological borders of *multiple* human agents. There will thus be multiple biological (skin/skull) borders to choose from. But which one of these borders ought to be identified as B? It should be clear from this example that the boundary-transcending nature of an extended mechanism cannot be determined simply by appealing to B's status as a biological boundary, so we will obviously need some other means of resolving B.

The problem of extended status is hard, much harder I think than it first appears. The problem could perhaps be tackled via an appeal to the idea that a mechanism is extended relative to the borders of whatever entity (object, system, or agent) is deemed to 'own' a phenomenon. Perhaps, then, the problem of extended status could be reduced to the problem of cognitive ownership. Accordingly, once we understand what it is that ties a particular cognitive process P to a particular cognitive agent A, then we will be in a position to judge whether or not P is extended by determining whether or not the mechanism M that realizes P is one that extends beyond the borders of A. This will, of course, require us to keep the borders of A constant, at least for the duration of P. In particular, we cannot state that the presence of P transforms A into an *extended agent*, such that the (agential) borders of A now encompass the spatial borders of M. If we do this, then we will have lost sight of the means by which the boundary B can be discerned. This is, in fact, a particular instance of a more general problem—a problem we might refer to as the problem of boundary expansion (or boundary inflation). In general, we need to resist a state-of-affairs in which some potential B is shifted as a consequence of the instantiation of an extended mechanism. The problem here is that if B should expand to include all the components of M, then M will not qualify as an extended mechanism. It will not qualify because the defining feature of an extended mechanism is that it should be a boundary-transcending mechanism. But if the boundary by which a mechanism is judged to be extended is one that expands to include the components that comprise the mechanism, then the mechanism cannot be extended, for it will no longer be a mechanism whose components are distributed across that all-important boundary B.

7. Conclusion

A mechanistic approach to extended cognition seeks to advance our conceptual understanding of extended cognition by drawing on theories of mechanistic explanation. From a mechanistic standpoint, extended cognitive phenomena are distinguished from non-extended cognitive phenomena by the mechanisms that constitute such phenomena.

In particular, extended cognitive phenomena are deemed to be cognitive phenomena (e.g., cognitive processes) that are constituted/realized by *extended mechanisms*. Such an approach is attractive because it promises to transform long-standing philosophical debates about the extended/embedded nature of cognition into a scientific search for mechanisms of a particular kind (see Kaplan, 2012).

The present paper identified some of the problems confronting the attempt to develop a mechanistic account of extended cognition. These problems are summarized in Table 1. A consideration of these problems is important because they provide a framework for future philosophical work. In particular, the five problems described in the present paper are intended to help coordinate (and, in some cases, compartmentalize) the philosophical effort to understand extended cognition. Some of these problems (e.g., the problem of constitutive relevance) are orthogonal to other problems and can be tackled as part of more general efforts in mechanical philosophy (and the philosophy of science). Other problems, however, are not so independent. It may be, for example, that a solution to the problem of cognitive status influences what we deem to be the proper target of cognitive scientific explanation, which touches on the problem of explanatory focus. In addition, a solution to the problem of cognitive ownership may be required in order to make progress on the problem of extended status.

Historically, the active externalist literature has tended to overlook (or at least underplay) some of the distinctions that are important for a mechanistic account of extended cognition. One such distinction is revealed by the problem of cognitive status and the problem of constitutive relevance. According to the present analysis, these problems are independent of one another. It is thus a mistake to assume that a solution to the problem of cognitive status (the mark of the cognitive) can be used to resolve the problem of constitutive relevance (the mark of the constitutive). Conversely, there is no reason to think that a solution to the problem of constitutive relevance will leave us any the wiser about whether or not we confront an extended cognitive system—we may be able to identify the borders and boundaries of a mechanism that is responsible for a phenomenon, but this will not tell us whether the phenomenon itself is a *bona fide* cognitive phenomenon. Similarly, there is no reason to think that solutions to the problem of cognitive status and the problem of constitutive relevance will mark the end of the quest for a mechanistic account of extended cognition. We will, I suggest, still need to understand what it is that makes a mechanism (cognitive or otherwise) a *bona fide* member of the class of extended mechanisms (i.e., the problem of extended status). This is important, for I suspect that one of the biggest challenges confronting a mechanistic approach to extended cognition is to specify what is meant by the term "extended mechanism." This is an issue that has thus far received little attention in the philosophical literature. It is, however, a crucial issue. A mechanistic account of extended cognition assumes that, one way or another, we will be able to distinguish extended mechanisms from non-extended mechanisms. If we are unable to do this, then it is far from clear that a mechanistic approach to extended cognition can be made to work.

Notes

[1]To some extent, of course, mechanistic considerations have always been a feature of philosophical debates in this area. In introducing the notion of extended cognition, for example, Clark (2008, p. xxviii) suggests that: "...the actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feedforward, and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body, and world. The local mechanisms of mind, if this is correct, are not all in the head. Cognition leaks out

into body and world."

[2]Note that for the purposes of this paper I will assume that the terms "mechanistic constitution" and "mechanistic realization" are synonymous. These terms are used interchangeably throughout the paper, such that talk of phenomena being constituted by mechanisms is deemed to be semantically-equivalent to talk of phenomena being realized by mechanisms.

[3]Sometimes claims about extended cognition are formulated with respect to the incorporation of extra-neural bodily resources, as opposed to extra-organismic (extra-corporeal) resources (see Boem, Ferretti, & Caiani, 2021; Facchin, Viola, & Zanin, 2021). In such cases, the relevant border or boundary is identified with the anatomical borders of the biological brain. A mechanism is thus deemed to be extended on account of the fact that some non-neural entity is revealed to be a constituent of a mechanism that realizes the functionality typically ascribed to brain-based or neural mechanisms.

[4]It is typically assumed that mechanistic explanations come in two varieties: etiological and constitutive mechanistic explanations. In etiological mechanistic explanations, a phenomenon is explained by describing the mechanism that *causes* the phenomenon. In constitutive mechanistic explanations, by contrast, a phenomenon is explained by describing the mechanism that *realizes* or *constitutes* the phenomenon. Of these two forms of mechanistic explanation, the notion of constitutive mechanistic explanation is the more important one for a mechanistic account of extended cognition. This is because constitutive mechanistic explanations require a means by which the components of mechanisms can be identified. As noted by Kaplan (2012), identifying the components of mechanisms (e.g., determining whether an extra-organismic resource should be included as a component in a mechanism) is a central concern in debates about the extended or non-extended (i.e., embedded) nature of human cognition.

[5]If the appeal to AI systems should be problematic for any reason, then the case can be rerun using two human agents instead of one human agent and one machine agent. As far as I can tell, this does not affect the outcome of the case.

[6]Note that these features may be relevant to the problem of cognitive ownership. That is to say, it may be the role played by the human agent in selecting, creating, and manipulating external resources that guides our intuitions regarding the human agent's ownership of the long multiplication routine. At the same time, this role may accentuate the extent to which we see the behavioral performances of the human agent as being the proper focus of explanatory attention in the context of the long multiplication task.

[7]According to Craver (2007b, p. 123): "The boundaries of mechanisms—what is in the mechanism and what is not—are fixed by reference to the phenomenon that the mechanism explains." As noted by Kaiser (2018), this raises a question about the extent to which mechanism borders and boundaries are objective features of reality. In respect of this issue, Kaiser (2018, p. 127) notes that once the phenomenon of interest is sufficiently specified and fixed, the boundaries of the mechanism for this phenomenon are also fixed. In this sense, mechanism borders and boundaries are relative to explanatory interests, but this does not mean that such borders and boundaries cannot be determined once a consensus about the explanandum phenomenon has been established.

[8]Just to be clear, the problem of explanatory focus is not the same as the problem of constitutive relevance, and this is despite the fact that both problems have a bearing on what is included within a mechanism. One could thus have agreement regarding the choice of explanandum phenomenon, but disagreement regarding an account of constitutive relevance. Or one could have agreement on an account of constitutive relevance, but disagreement as to what phenomenon ought to be explained by a mechanistically-oriented account. If there is disagreement about the problem of constitutive relevance, then we are likely to arrive at different answers as regards the borders/boundaries of mechanisms, and this will be so even if the phenomenon we are trying to explain is held constant. Conversely, if there is disagreement about the explanandum phenomenon, then we are likely to end up with different answers as regards the borders/boundaries of mechanisms, even if the thing that is held constant is the account of constitutive relevance.

[9]Such shifts in explanatory focus may also occur in response to the availability of specific methods and techniques (see, for example, Bollhagen, 2021).

[10]The functional individuation of cognitive mechanisms is broadly consistent with the way in which cognitive phenomena are conceptualized in the philosophical literature. As noted by Illari and Williamson (2012, p. 131), "In psychology in particular, capacities like memory are often given a purely functional description. There are indefinitely many ways the human brain could divide up the task of remembering things."

References

Adams, F. (2010). Why we still need a mark of the cognitive. *Cognitive Systems Research*, *11*(4), 324–331.

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

Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines*, *23*(3), 339–352.

Allen, C. (2017). On (not) defining cognition. *Synthese*, *194*(11), 4233–4249.
Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756.
Baumgartner, M., & Wilutzky, W. (2017). Is it possible to experimentally determine the extension of cognition? *Philosophical Psychology*, *30*(8), 1104–1125.
Bechtel, W., & Richardson, R. C. (1993/2010). *Discovering complexity: Decomposition and localization as strategies in scientific research.* Cambridge, Massachusetts, USA: MIT Press.
Boem, F., Ferretti, G., & Caiani, S. Z. (2021). Out of our skull, in our skin: the Microbiota-Gut-Brain axis and the Extended Cognition Thesis. *Biology & Philosophy*, *36*(Article 14), 1–32.
Bollhagen, A. (2021). The inchworm episode: Reconstituting the phenomenon of kinesin motility. *European Journal for Philosophy of Science*, *11*(Article 50), 1–25.
Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* New York, New York, USA: Oxford University Press.
Clark, A. (2010). Memento's Revenge: The Extended Mind, Extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43–66). Cambridge, Massachusetts, USA: MIT Press.
Clark, A. (2011). Finding the mind. *Philosophical Studies*, *152*(3), 447–461.
Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.
Craver, C. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research*, *32*, 3–20.
Craver, C. (2007b). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience.* Oxford, UK: Clarendon Press.
Craver, C. (2015). Levels. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND: Philosophy and the Mind Sciences in the 21st Century* (pp. 1–26). Frankfurt am Main, Germany: MIND Group.
Craver, C., & Darden, L. (2013). *In Search of Mechanisms: Discoveries Across the Life Sciences.* Chicago, Illinois, USA: The University of Chicago Press.
Facchin, M., Viola, M., & Zanin, E. (2021). Retiring the "Cinderella view": the spinal cord as an intrabodily cognitive extension. *Biology & Philosophy*, *36*(Article 45), 1–25.
Fazekas, P. (2013). The Extended Mind Thesis and Mechanistic Explanations. In D. Moyal-Sharrock, V. A. Munz, & A. Coliva (Eds.), *Mind, Language, and Action* (pp. 125–127). Kirchberg am Wechsel, Austria: Austrian Ludwig Wittgenstein Society.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*(1), 49–71.
Glennan, S., & Illari, P. M. (Eds.). (2018). *The Routledge Handbook of Mechanisms and Mechanical Philosophy.* New York, New York, USA: Routledge.
Hurley, S. (2010). The Varieties of Externalism. In R. Menary (Ed.), *The Extended Mind* (pp. 101–153). Cambridge, Massachusetts, USA: MIT Press.
Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, Massachusetts, USA: MIT Press.
Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135.
Japyassú, H. F., & Laland, K. N. (2017). Extended spider cognition. *Animal Cognition*, *20*(3), 375–395.
Kaiser, M. I. (2018). The components and boundaries of mechanisms. In S. Glennan & P. M. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 116–130). New York, New York, USA: Routledge.
Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy*, *27*(4), 545–570.
Lyre, H. (2018). Socially Extended Cognition and Shared Intentionality. *Frontiers in Psychology*, *9*(Article 831), 1–9.
Malleson, K. (2018). Equality Law and the Protected Characteristics. *The Modern Law Review*, *81*(4), 598–621.
Miłkowski, M., Clowes, R., Rucińska, Z., Przegalińska, A., Zawidzki, T., Krueger, J., ... Hohol, M. (2018). From wide cognition to mechanisms. *Frontiers in Psychology*, *9*(Article 2393), 1–17.

Palermos, S. O. (2014). Loops, Constitution, and Cognitive Extension. *Cognitive Systems Research*, *27*, 25–41.

Parise, A. G., Gagliano, M., & Souza, G. M. (2020). Extended cognition in plants: is it possible? *Plant Signaling & Behavior*, *15*(2), 1710661.

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

Prychitko, E. (2021). The causal situationist account of constitutive relevance. *Synthese*, *198*, 1829–1843.

Roberts, T. (2012). You do the maths: Rules, extension, and cognitive responsibility. *Philosophical Explorations*, *15*(2), 133–145.

Rupert, R. D. (2004). Challenges to the Hypothesis of Extended Cognition. *Journal of Philosophy*, *101*(8), 389–428.

Smart, P. R. (2018). Human-Extended Machine Cognition. *Cognitive Systems Research*, *49*, 9–23.

Smart, P. R., Clowes, R. W., & Heersmink, R. (2017). Minds Online: The Interface between Web Science, Cognitive Science and the Philosophy of Mind. *Foundations and Trends in Web Science*, *6*(1–2), 1–232.

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process. In R. Menary (Ed.), *The Extended Mind* (pp. 189–225). Cambridge, Massachusetts, USA: MIT Press.

van Eck, D., & de Jong, H. L. (2016). Mechanistic explanation, cognitive systems demarcation, and extended cognition. *Studies in History and Philosophy of Science Part A*, *59*, 11–21.

Wheeler, M. (2011). In search of clarity about parity. *Philosophical Studies*, *152*(3), 417–425.

Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition.* New York, New York, USA: Cambridge University Press.

Wilson, R. A. (2014). Ten questions concerning extended cognition. *Philosophical Psychology*, *27*(1), 19–33.

Wilson, R. A., & Clark, A. (2009). Situated Cognition: Letting Nature Take its Course. In P. Robbins & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55–77). Cambridge, UK: Cambridge University Press.

Ylikoski, P. (2018). Social mechanisms. In S. Glennan & P. M. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 401–412). New York, New York, USA: Routledge.

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

Table 1. Summary of problems for a mechanistic account of extended cognition. (The "Scope" column indicates the scope of a potential solution. For example, a solution to the problem of constitutive relevance ought to be applicable to all mechanisms, not just mechanisms of the cognitive kind.)

Problem	Question	Scope
Cognitive Status	What makes an explanandum phenomenon a cognitive phenomenon?	All cognitive phenomena.
Constitutive Relevance	How do we know when some object is constitutively relevant to an explanandum phenomenon?	All mechanisms.
Cognitive Ownership	Why do we see the explanandum phenomenon as being 'owned' by a particular (cognitive) agent or entity?	All cognitive phenomena.
Explanatory Focus	What phenomenon is being explained? Is the explanandum phenomena a fitting target for cognitive scientific explanation?	All cognitive phenomena.
Extended Status	What distinguishes extended mechanisms from non-extended mechanisms?	All extended mechanisms.

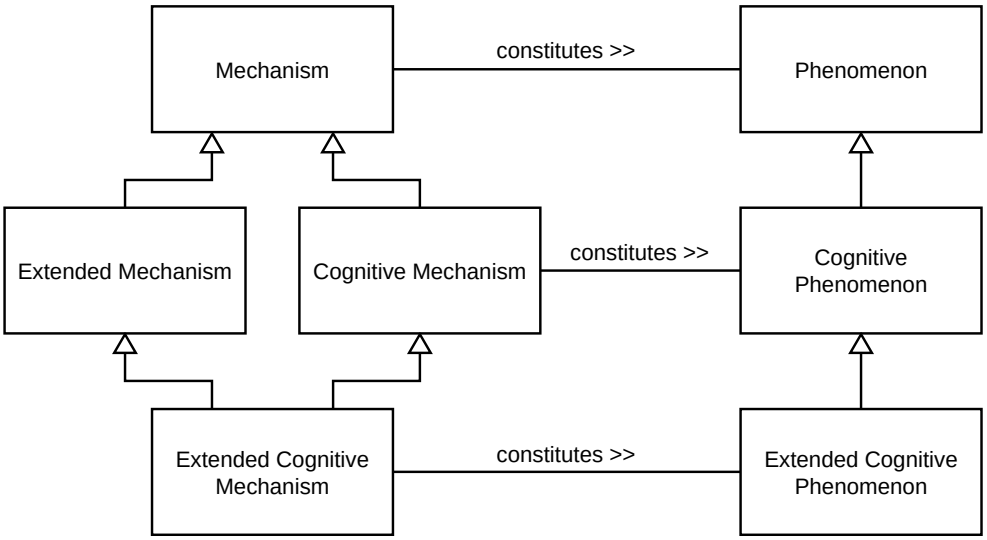
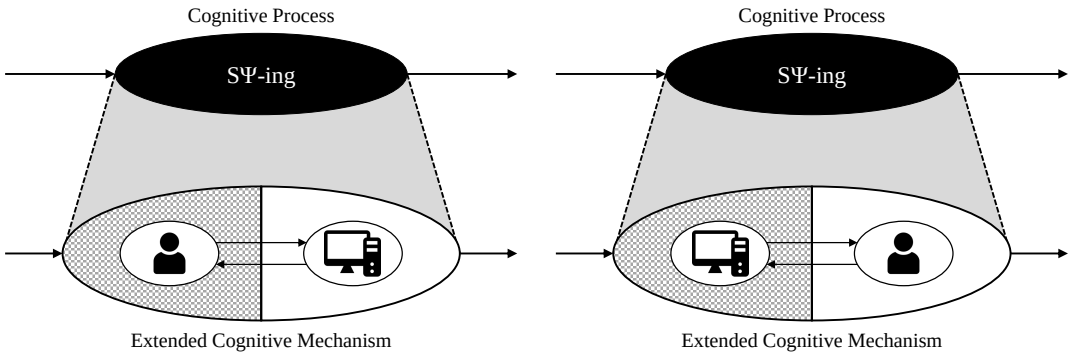


Figure 1. An extended cognitive mechanism is conceptualized as a mechanism that qualifies as both a cognitive mechanism and an extended mechanism. Cognitive mechanisms are defined by the nature of the phenomena they constitute; i.e., the defining feature of a cognitive mechanism (the thing that distinguishes it from other mechanisms) is the fact that it is responsible for phenomena of the cognitive variety. (Triangles symbolize taxonomic or subtype-of relationships. The double arrows to the right of the word "constitutes" indicate the directionality of a relationship. So, the figure should be read as Mechanism—constitutes—Phenomenon, not Phenomenon—constitutes—Mechanism.)



(a) Extended Cognition (b) Human-Extended Machine Cognition

Figure 2. Two kinds of extended cognition. (a) The conventional view of extended cognition, in which an individual human agent incorporates an extra-organismic resource (in this case, an AI system) into a cognitive routine. (b) A different view of extended cognition, in which the human agent is a component of the cognitive routines of an AI system. [The hatched region of the lower ellipse indicates the entity to whom ownership of the cognitive routine is assigned. Both diagrams draw on the conventions used by Craver (2007b, p. 7) to represent the relationship between phenomena and mechanisms. These are what have come to be known as "Craver Diagrams."]