

(Forthcoming in *Cognitive Systems Research*)

# Loops, Constitution, and Cognitive Extension

*University of Edinburgh*

S. Orestis Palermos

**Abstract:** The ‘causal-constitution’ fallacy, the ‘cognitive bloat’ worry, and the persisting theoretical confusion about the fundamental difference between the hypotheses of embedded (HEMC) and extended (HEC) cognition are three interrelated worries, whose common point—and the problem they accentuate—is the lack of a principled *criterion of constitution*. Attempting to address the ‘causal-constitution’ fallacy, mathematically oriented philosophers of mind have previously suggested that the presence of non-linear relations between the inner and the outer contributions is sufficient for cognitive extension. The abstract idea of non-linearity, however, can be easily misunderstood and has, in the past, led to incorrect and counterintuitive conclusions about what may count as part of one’s overall cognitive system. In order to prevent any further mistakes I revisit dynamical systems theory to study the nature of the *continuous mutual interactions* that give rise to the aforementioned non-linear relations. Moreover, focusing on these interactions will allow us to provide two distinct arguments in support of the ontological postulation of extended cognitive systems, as well as an objective criterion of constitution. Accordingly, I put forward a version of HEC that treats continuous mutual interactions (and the resultant non-linear relations) not just as sufficient but also as necessary for cognitive extension. Such a qualified version of HEC may exclude certain alleged cases of cognitive extension where the agent does not mutually interact with his artifacts (*e.g.*, shopping lists and directory services), but it is immune both to the ‘causal-constitution’ fallacy and the ‘cognitive bloat’ worry, and it can be sharply distinguished from HEMC.

**Key words:** Dynamical Systems Theory; Hypothesis of Extended Cognition; ‘Causal-Constitution’ Fallacy; ‘Cognitive Bloat’, Continuous Mutual Interactions.

## 1. Introduction

The Extended Cognition (HEC) and the Extended Mind hypotheses are two related formulations of active externalism within contemporary philosophy of mind. As a general approach to the nature of mind, active externalism (Clark and Chalmers 1998; Rowlands 1999; Wilson 2000; Wilson 2004; Menary 2007) is standardly contrasted with Putnam (1975) and Burge's (1986) meaning, or passive, externalism as it concentrates on the aspects of the environment that *drive* one's cognitive loops in an ongoing way.

The Extended Cognition and the Extended Mind hypotheses, however, constitute two different (especially in their degree of radicalism) interpretations of active externalism. Focusing on *cognitive processing*, the hypothesis of extended cognition is the claim 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" (Clark 2007, sec. 2); cognitive processing can and (under the appropriate conditions) literally extends to the agent's surrounding environment. Think about solving a mathematical problem by using pen and paper, or perceiving a chair through a tactile visual substitution system. According to HEC, the involved artifacts are proper parts of the ongoing cognitive processing.

However provocative this claim may sound, the Extended Mind hypothesis is usually thought to be more challenging still. Instead of concentrating on cognitive processes, the claim, in this case, is that it is *mental states*—experience, beliefs, desires, emotions, and so on—that get extended. The typical argument (Clark & Chalmers 1998) involves Otto—an Alzheimer's patient—whose dispositional beliefs are taken to be partly *constituted* by his well-organized notebook; his mind, therefore, extends to his notebook.

Barring for the most part the Extended Mind hypothesis, my primary focus here will be the debate over HEC. Specifically, I will focus on two of the most important objections facing the view, known as the 'causal-constitution' fallacy and the 'cognitive bloat' worry. Discussing these two objections will also lead us to the juxtaposition of HEC with its most promising alternative, the hypothesis of Embedded Cognition (HEMC). This alternative view recognizes the dependence of cognition on its environment but denies its extension beyond the organismic boundaries. Actually, as we shall see in the following section, this is a series of interrelated worries whose common point—and the problem they accentuate—is the lack of a principled *criterion of constitution*; were we able to explain what constitutes a cognitive process and thence a

cognitive system, we would be able to provide a clear answer to all three of the above criticisms.

To approach the above problems, in section 3, I will focus on certain mathematical insights from dynamical systems theory. Eventually, these insights will give rise to an immunized version of HEC, which relies for the most part on Clark and Chalmers' (1998) classical proposal, but differs in one central aspect: It treats continuous mutual interactions between the organismic agent and the artifact as both necessary and sufficient for cognitive extension.

Even though not entirely clear in (Clark & Chalmers 1998), Clark (2007; 2008) intends continuous mutual interactions as only sufficient for cognitive extension—probably to allow space for the Extended Mind hypothesis as motivated by the Otto and similar cases, where (possibly) no such interactions are involved. Regardless of the potential stringency of the move (I will return to this issue in section 5), it is tempting to accentuate the importance of this criterion by treating it as both a sufficient and necessary condition on cognitive extension. Doing so, as I will argue in section 4, will allow us to safeguard HEC from both the 'causal-constitution' fallacy and the 'cognitive bloat' worry *and* draw a clear distinction between HEC and its rival hypothesis, HEMC.

Notably, mathematically oriented cognitive scientists and philosophers of mind have previously commented on the importance of dynamical systems theory when individuating cognitive systems. Clark (2008) and Wheeler (2005) have pointed towards this direction and some critiques of HEC are aware of the force of this argumentative line (Rupert 2009). Moreover, Chemero (2009) has suggested in passing an efficient response to the 'causal-constitution' fallacy, which is technically equivalent to the one we will be here considering. Specifically, he holds that the existence of non-linear relations between two systems is the criterion by which we can judge they constitute an overall system comprising of both of them.<sup>1</sup>

As we shall see in section 4, however, there is a serious problem with Chemero's analysis. As it becomes apparent from the exposition of his view, *viz.*, Radical Embodied Cognitive Science—and in particular from his defense of direct realism/anti-representationalism—he employs an incorrect understanding of what non-linear relations are supposed to be. Ironically, this misunderstanding seems to lead him back to the fallacy he has previously offered a solution out of.

Although I agree with Chemero that non-linear relations are crucial for constitution, the difference of the present approach consists in revisiting dynamical

---

<sup>1</sup> I should note that a previous version of this paper made no mention of Chemero's work, which, until recently, I was unaware of. I am thankful to Andrew Wilson for drawing my attention to it.

systems theory to provide two arguments in support of this claim (section 3). These two arguments will help us grasp the metaphysical importance of, and clarify how we should understand, the non-linear relations Chemero is appealing to. Specifically, they will make us focus on the nature of the *continuous mutual interactions that give rise to those non-linear relations*. In return, we will be able to pin down when two elements are non-linearly related and see how, in his attempt to defend direct realism, Chemero was misled back to the ‘causal-constitution’ fallacy. In addition, this focus on continuous mutual interactions will help us project DST in real case scenarios, thereby allowing us to directly address the ‘cognitive bloat’ worry, and clearly distinguish between HEC and HEMC.

Nevertheless, as noted above, considering continuous mutual interactions (on the basis of which non-linear relations arise) as both necessary and sufficient for cognitive extension is likely to provide a more restricted (and thus more plausible?) access to the list of genuine cases of cognitive extension. In an attempt to test the stringency of this criterion, in section 5, we will finally revisit (but not necessarily revise) our intuitions regarding the limits of cognitive extension, as well as the underlying concept of mind that motivates them.

## **2. The Hypothesis of Extended Cognition, the ‘Causal-Constitution’ Fallacy, ‘Cognitive Bloat’, and the Hypothesis of Embedded Cognition**

As HEC is usually put forward, *when* parts of the environment become ‘properly coupled’ to an agent’s brain, they can be considered as constitutive parts of the overall cognitive mechanism—*i.e.*, cognition potentially extends to the world surrounding the agent. “In these cases, the human organism is linked with an external entity in a two-way interaction creating a *coupled system* that can be seen as a cognitive system in its own right” (Clark & Chalmers 1998, 8).

Clark has also termed this two-way interaction as ‘continuous reciprocal causation’ (CRC): “CRC occurs when some system S is both continuously affecting and simultaneously being affected by activity in some other system O” (Clark 2008, 24). In such cases (and for reasons to be explored in section 3), in order to model the temporal evolution of the two systems S and O, *Dynamical Systems Theory* further postulates a *coupled system* E consisting of both S and O.

So the claim is that when one’s internal cognitive capacities are combined (*i.e.*, mutually interact) with some environmental element O, they form an extended cognitive whole, E, whose behavioral competence will drop if one removes the external

component, just as it would drop if one removed part of its brain.

Consider, for example, the use of pen and paper when trying to solve a complex, say, a three-digit multiplication problem such as 987 times 789. It is true that few, if any, of us can solve this problem by looking at or contemplating on it. We may only perform the multiplication process by using pen and paper to externalize the very problem in symbols. Then, we can serially proceed to its solution by performing simpler, iterative multiplications, starting with 9 times 7, and externally storing the results of the process, for use in later stages. In this way, the pen and paper compensate for our limited working memories allowing us to perform a task that is otherwise infeasible. Accordingly, if one tried to describe how a regular human mind may perform such a cognitive task then, apart from the states and properties of a typical human brain, one should also factor in both the normative aspects of the notational/representational system involved (that we cannot multiply by infinity, for instance), and the properties and ongoing states of the mediums with which the manipulation of the representations was performed.<sup>2</sup>

Similarly, think about the role of language when writing a philosophy paper. According to Clark, language too is “an external epistemic artifact designed to complement, rather than recapitulate or transfigure, the basic processing profile we share with other animals” (1998, 169). As I write down this essay...

...I am continually creating, putting aside, and re-organizing chunks of text [...] I have source texts and papers full of notes and annotations. As I (literally, physically) move these things about, interacting first with one, then another, making new notes, annotations and plans, so the intellectual shape of the chapter grows and solidifies. It is a shape which does not spring fully developed from inner cogitations. Instead, it is the product of a sustained and iterated sequence of interactions between my brain and a variety of external props (Clark 1998, 173).

To return, now, to the discussion of the core tenets of HEC, we can take a look at the most famous example that Clark and Chalmers (1998) use in order to argue for the Extended Mind hypothesis. Before moving to the example, however, we may first recall that whereas HEC is motivated on the basis of extended cognitive processes, the hypothesis of the Extended Mind is usually formulated on the basis of extended mental states. This makes the latter position somewhat more provocative, because the existence of extended mental states—such as extended dispositional beliefs—is a claim that is more counterintuitive and thereby less easily motivated than the claim that there are

---

<sup>2</sup> For the importance of the normative aspects of the external representational systems in explaining cognition see (Menary 2007).

extended cognitive processes. In fact, as we shall see in a while, arguments for the Extended Mind hypothesis usually rely on a sort of common-sense functionalism that does not seem to be necessary in arguments for cognitive extension.<sup>3</sup>

Here is the example: First, think about a normal case of a belief stored in biological memory. Inga learns about an interesting exhibition in MOMA. She thinks, recalls that the museum is on 53<sup>rd</sup> street and starts walking to the museum. Now consider Otto who suffers from Alzheimer's disease; as a consequence, Otto has to rely on information in the environment to help structure his life and so carries a thick, well-organized notebook everywhere he goes. When he learns new information he writes it down, when he needs some old information he looks it up. Otto hears about the same exhibition and decides to go see it. He opens the notebook, finds the address of the museum and starts heading towards 53<sup>rd</sup> street.

Clark and Chalmers (1998) claim that Otto walked to 53<sup>rd</sup> street because he wanted to go to MOMA and believed that MOMA was on 53<sup>rd</sup> street. What is more, if one wants to say that Inga had her belief before she consulted her memory, then one could also claim that Otto believed that the museum was on 53<sup>rd</sup> street even before looking up the address in his notebook. This is because the two cases are functionally on a par; given our everyday, common-sense understanding of how memory works, we can make the following claim: "the notebook plays for Otto the same role that memory plays for Inga; the information in the notebook functions just like the information [stored in Inga's biological memory] constituting an ordinary non-occurrent belief; it just happens that this information lies beyond the skin" (1998, 13). The conclusion, therefore, is that Otto's dispositional beliefs, and thereby his mind, are extended to his notebook.<sup>4</sup>

Although the postulation of extended mental states is not necessary for making the case for HEC, allowing us to bypass the long debate that Otto has generated (I will return to the Otto case in the concluding section), the discussion of this example is helpful as it has produced some very interesting intuitions on what is required for an external artifact to count as a putative part of one's overall cognitive economy. In particular, investigating the case in more detail, Clark (2010a) notes that the availability

---

<sup>3</sup> This does not mean that HEC is incompatible with common-sense functionalism, or that HEC is anti-functionalism on the whole. In so far as a cognitive process is a function, HEC is compatible with functionalism. Running ahead of myself, I can here only note that the present approach appears to favor a 'systems' version of functionalism that focuses on the functional-geometrical properties of abstract systems as understood by dynamical systems theory. One of the advantages of such functionalism is that it can provide very precise multiple realizability claims, informed by the parameter spaces of the abstract systems under study. Also see fn. 19.

<sup>4</sup> For an overview and rejection of the claim that external memory is real memory see (Michaelian 2012). Michaelian notes, however, that his arguments against the idea that external memory is real memory are not arguments against HEC—which may still be true of the Otto case.

and portability of the resource of information might be crucial. Accordingly, he offers the following set of additional criteria to be met by non-biological candidates for inclusion into an individual's cognitive system (2010a, 46)<sup>5</sup>:

- 1) "That the resource be reliably available and typically invoked".
- 2) "That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny. [...] It should be deemed about as trustworthy as something retrieved clearly from biological memory".
- 3) "That information contained in the resource should be easily accessible as and when required".

These criteria have also come to be known in the literature as the 'glue and trust' criteria and they are primarily meant to ensure the effect of 'transparent equipment': "equipment (like the carpenter's hammer) with which we are so familiar and fluent that we do not think about it in use, but rather rely on it to mediate our encounters with a still-wider world" (Clark 2006, 106). In other words, an external element is part of one's ongoing cognitive loops when it is not part of the problem space but is instead one of the mediums manipulated in order to complete the cognitive task at hand.

The problems, however, for the HEC theorist begin with the observation that these criteria may be far too easily satisfied, meaning that they are insufficient to ensure that some external element can count as part of one's cognitive system.<sup>6</sup> Rupert (2004, 401-5), for example, argues that when a person has access to a phonebook, or a directory service through the use of her cellular phone, she can be said to satisfy the criteria that Clark has set forth. Admittedly, however, it would be counterintuitive to conclude that the phonebook, or the directory service is part of that person's overall cognitive system, allowing her to have non-occurrent true beliefs about the phone numbers of everyone whose number is listed.

In other words, if any external element that both satisfies the 'glue and trust' criteria and causally affects one's cognitive processes is to count as part of one's cognitive system, we are going to be led to a 'cognitive bloat' (Clark 2001, Rowlands 2009) whereby cognition will seem like leaking all the way out in implausibly many directions. Eventually, the worry further goes, we will be led to an "unacceptable proliferation of systems (many of them extremely short lived)" (Rupert 2004, 396).

---

<sup>5</sup> This paper was first published in *The Extended Mind*, (2010), Menary (ed.) Cambridge, Massachusetts, MIT press, but it has been available online since 2006. The 'glue and trust' criteria, however, had already made their appearance in (Clark and Chalmers 1998), although the phrasing was somewhat different.

<sup>6</sup> In (Clark & Chalmers 1998, 17) the authors consider a further criterion: "Fourth, the information in the notebook has been consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement". As the authors further note, however, "the status of the fourth feature as a criterion for belief is arguable (perhaps one can acquire beliefs through subliminal perception, or through memory tampering?)", so they subsequently drop the said criterion.

Now notice that such a ‘cognitive bloat’ would actually be the outcome of repeatedly committing the ‘causal-constitution’ fallacy that Adams and Aizawa have pointed out. The objection, this time, is that proponents of HEC often put forward their view by arguing that since an external process (*e.g.*, the directory service) causally affects a cognitive process (*e.g.*, the search for a phone number), then the external process is a genuine part of the overall cognitive process. But this, Adams and Aizawa note, is fallacious: “it simply does not follow from the fact that process X is in some way causally connected to a cognitive process that X is thereby part of that cognitive process” (2008, 91). They call this the simple version of the fallacy. Parenthetically, however, we should also note that Adams and Aizawa have identified a second version of it—*viz.*, the ‘systems version’—which unfolds in two steps: “The first is to move from the observation of some sort of causal connection to the claim that the brain, body and relevant parts of the world form a cognitive system. The second step is a tacit shift from the hypothesis that something constitutes a system to the hypothesis that it is an instance of extended cognition” (Adams & Aizawa 2008, 92). This, however, is again fallacious: “It simply does not follow from the fact that one has identified an X system in terms of a causal process of type X that that process pervades every component of the system” (*ibid.*, 125). Nevertheless, Adams and Aizawa are worried that many times proponents of cognitive extension rely on at least one of these two fallacious moves.

Accordingly, instead of arguing for the constitutive contribution of the external artifacts to one’s overall cognitive economy (which seems to rely on fallacious reasoning), one should simply endorse the much less provocative idea that cognition is many times *merely dependent* on external elements. In other words, one should better opt for the less radical position that has come to be known as the Hypothesis of Embedded Cognition (HEMC) (Rupert 2004, 393).

HEMC: Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place.

This hypothesis is close to HEC as it acknowledges the dependence of cognition on its environment. It is, however, a conservative view because it denies that environmental aspects are proper parts of cognition; external factors may only serve as tools and props to cognition, which is restricted within the organismic brain or, at most, the organismic body as a whole. According to HEMC, then, cognition is organism-bound, potentially aided by environmental factors, but not extended to them. It denies that cognitive mechanisms are external, but it also denies that a mechanistic explanation of how



psychological processes work should be a purely internal story. “An advocate of HEMC may claim that cognitive mechanisms are internal, but that the mechanistic explanation of how they work is a complex story involving both internal activity and environmental resources” (Sprevak 2010, 356).

As a result, it has been further argued that since both accounts are concerned with the way agents interact with their environments, both views will produce the same causal explanations with respect to cases where agents employ artifacts. But if this is true, then HEMC allegedly wins the day on the basis of conservatism and simplicity: “If two theories embrace structurally equivalent explanations (with or without the same labels), but one of those theories simply tacks on commitment to an additional kind of entity [*e.g.*, extended cognitive processes, or extended cognitive systems], of no causal significance, then the relative simplicity comparison is straightforward” (Rupert 2009, 18).<sup>7</sup>

Should we then abandon HEC on the face of HEMC, or are there reasons for not giving up so easily? Could there be a principled way to individuate systems in general, and cognitive systems in particular, such that we can avoid the ‘causal-constitution’ fallacy and the related ‘cognitive bloat’ worry? Moreover, is it true that HEC and HEMC can provide the same mechanistic explanations and that the only difference between the two is that the former unnecessarily postulates the existence of extended cognitive systems?

To answer, we must first focus on dynamical systems theory (DST)—the best available tool for modeling and understanding systems that continuously interact with their environment, which is precisely the kind of systems that both HEC and HEMC are interested in. Remarkably, DST has lately gained an impressive momentum in providing cognitivist explanations of processes that may be restricted within the brain, processes that extend to the agent’s body, or even processes that span brain, body and environment.<sup>8</sup> Despite this, however, the reason I take this approach to be particularly helpful in modeling the kind of systems that both HEC and HEMC focus on does not depend on its empirical success—especially when compared with its main competitor, *i.e.*, computationalism. Given the amount of time this rival approach has been around, such a comparison is bound to lead to the opposite direction. My point is rather technical. Both HEC and HEMC focus on the continuous interaction between the agent and his environment, and this interaction is inherently temporal in such a way that

---

<sup>7</sup> Mark Sprevak (2010), however, argues that such a straightforward comparison is not really possible.

<sup>8</sup> See, for example, (Bressler and Kelso 2001); (Thompson and Varela 2001); (Varela et al. 2001); (Kelso and Engström 2006); (Dale and Spivey 2006); (Spivey and Dale 2006); (Spivey 2007).

infinitesimal changes in time can be central to understanding the behavior of the target systems. As is known, however, the sequential nature of computational techniques cannot accommodate temporal considerations at this level of complexity (Gelder 1995). HEC and HEMC, therefore, can best (and perhaps exclusively) be modeled by using the techniques of DST.<sup>9</sup> With that said, we may now turn to the mathematical theory itself.

### 3. Systems

#### 3.1. Dynamical Systems Theory

One of the primary activities of several scientific disciplines, such as physics, chemistry, biology and the social sciences as well, is the study of systems. Systems are sets of interdependent elements, objects, entities, or items standing in interrelations on the basis of specific processes they take part in and give rise to, thereby forming a unified whole. Of course, an element, object, entity, or item can be part of several systems at the same time, depending on the kind of processes it engages in. Thus, whether some object counts as a component of a system always depends on the phenomenon under study and, more in particular, on the processes that are thought to give rise to the relevant phenomenon.

With respect to the scientific study of systems, on one hand, we have *dynamical modeling*, which is the part of applied mathematics that is concerned with understanding natural phenomena by providing abstract dynamical models for them—*i.e.*, mathematical entities also known as abstract state-dependent systems. *Dynamical Systems Theory* (DST), on the other hand, is a branch of theoretical mathematics, which is concerned with the properties of such abstract dynamical systems. The general strategy of DST is to conceptualize systems geometrically, in terms of positions, distances, regions and trajectories within the space of a system's possible states. DST is thus concerned with the geometrical properties of the *flow* of the system, which is the entire range of the possible *trajectories* of an abstract dynamical system.

Now, before moving on we should note again that the primary focus of this paper is on *continuous mutual interactions*, which are distinctive of *coupled systems*. Specifically, the aim is to understand what those continuous mutual interactions and their resulting *coupled systems* are, which will, in turn, allow us to provide two arguments that will explain why the postulation of such systems is necessary for understanding certain

---

<sup>9</sup> However, I will not here take stands in the further debate between connectionist and dynamicist approaches to the mind (see Gelder 1995). I am interested in the ramifications of DST as a mathematical tool, and both of these approaches employ DST as their primary modeling tool.

behaviors. In order to properly appreciate the nature of coupled systems, however, we must first talk about some basic concepts—including the concepts of *autonomous* and *non-autonomous* systems—that are necessary for their definition.<sup>10</sup>

So to start with, in general, every dynamical system is characterized by a set of *state variables*  $x$  and a *dynamical law*  $L$ —a set of differential equations—that regulates the change of those *state variables* across time. Starting from some initial state  $x_0$  the law  $L$  generates a sequence of states, which is called the *trajectory* of the system. The set of all *trajectories* through every point in the *state space* is called the *flow* and, as previously mentioned, DST is primarily interested in the geometrical structure of the entire *flow* of the system; *i.e.*, the geometrical, or topological properties of all the possible behaviors the system might exhibit across time.

Here are some important behaviors that a system may exhibit. Sometimes, a system can converge to certain *limit sets* called *attractors*. *Attractors* have the interesting property that they gravitate *trajectories* passing through all nearby states. Accordingly, the set of initial states that converge to a given *attractor* is called its *basin of attraction*. The portions of the *trajectories* that are found within a *basin of attraction*, but which do not lie in the *attractor* itself are termed *transients*. Now, the reason why *attractors* are important is because they govern the long-term behaviors of a physical system; regardless of its initial state, a physical dynamical system will always settle near an *attractor* after *transients* have passed. In contrast, *repellers* are *limit sets* that are unstable in that some nearby *trajectories* diverge from them. So the *state space* of dynamical systems contains multiple *repellers* and *attractors* (each surrounded by its own *basin of attraction*), which determine the trajectories the system may take.

To move now to the definition of the systems that we will be focusing on, let us start with the *autonomous* ones. If the *dynamical law*,  $L$ , depends only on the values of the system's *state variables* and the values of some *parameters*,  $u$ —that remain fixed during the operation of the system ( $x' = L(x(t), u)$ )—then the system is called *autonomous*. Most *autonomous systems* are structurally stable, in that changing the values of their fixed *parameters*,  $u$ , will produce small changes in the resulting *flow*.<sup>11</sup> Other systems, however, can become unstable in that very small changes in *parameter* values can produce substantial changes in their *flow*, bringing about *phase portraits* that are qualitatively different from the initial one. For example, new *attractors* may appear and old *repellers* may disappear. These *qualitative* changes in the system's *flow* are called *bifurcations*.

<sup>10</sup> In what follows, the definition of italicized technical terms can be found in table 1, on p. 35. Note that the order of appearance in the table follows the order of appearance in the main text.

<sup>11</sup> "Limit sets and basins of attraction may deform and move around a bit, but the new flow will be qualitatively similar (*i.e.*, topologically equivalent, or *homeomorphic*) to the old one" (Beer 1995, 180).

Changing, therefore, the parameters of a system can bring about both quantitative and qualitative changes. What has been said so far, however, concerns *autonomous* dynamical systems (*i.e.*, systems whose *parameters* are held constant for the duration of any particular *trajectory*). Allowing, however, *parameters* to change across time, as the *trajectory* unfolds, gives rise to a second type of systems, *viz.*, *nonautonomous systems*. *Nonautonomous* dynamical systems are, therefore, systems in which one or more *parameters* vary in time:  $x' = L(x(t), u(t))$ , and since “the flow is a function of the parameters, [...their states are] governed by a flow which is changing in time (perhaps drastically if the parameter values cross bifurcation points in parameter space)”<sup>12</sup> (Beer 1995, 180).

Now, it is also important to distinguish between two kinds of parameters and their representational roles. On one hand, as we saw, we have *parameters*  $u$ , which remain constant. This kind of parameters refers either to one of the internal features of the system that may be manipulated (but which remain fixed during the system’s operation), or to the stable background conditions the system operates in. On the other hand, we have *parameters*  $u(t)$ , which change over time, and which represent the *inputs* to the system. These inputs might originate from the dynamical environment, or some other well-defined system that causally affects the system under study.

Now, keeping all this in mind, it is very interesting to see what happens when these *inputs* (*i.e.*, *parameters*  $u(t)$ ) do not just originate from the system’s dynamical environment, but from another system with which the system under study *mutually interacts*. This *mutual* interaction gives rise to the third and last type of system we will be here concerned with, namely *coupled* systems. Typical examples of such systems include two mutually interconnected pendulums, the watt governor and a rotation engine, and, possibly, cognitive agents and their epistemic artifacts. Following Beer (1995), I will here present the case by focusing on a (blind) agent mutually interacting with a specific aspect of his environment, such as a tactile visual substitution system.<sup>13</sup> In other words, we will be focusing on the two continuous-time *nonautonomous* dynamical systems whose *dynamical law* will be  $A$  and  $TVSS$ , respectively.<sup>14</sup>

---

<sup>12</sup> A parameter space is the set of all the values of the parameters encountered in a particular mathematical model.

<sup>13</sup> See Bach-y-Rita and Kerzel (2003) for a recent review on TVSS.

<sup>14</sup> Important note: In the example Beer offers in his paper, the environmental aspect,  $E$ , actually refers to the agent’s body, which mutually interacts with the agent’s neural network,  $A$ . Several authors, however, who have been inspired by Beer’s understanding of autonomous agents take  $E$  to either refer to some ambient feature of the environment (*e.g.*, light), or to some particular object of perception like a tree. The problem is that ambient features of the environment cannot be treated as systems in their own right such that they can mutually interact with the agent, and even though objects of perception like trees are systems in themselves they do not mutually interact with the agent. Therefore, neither ambient features of the environment nor

Now, to say that the agent and the substitution system engage into continuous mutual interaction just is to say that the two systems are *coupled nonautonomous* dynamical systems. In more technical terms, however, two (*nonautonomous*) systems are *coupled* when the changing *parameters*  $u(t)$  of each system function as some of the *state variables* of the other, and vice versa.

We can further represent this mutual interaction as a function  $E$  from the substitution system's *state variables* to the agent's *changing parameters*—such that it captures all the ways in which the substitution system can affect the agent—and a function  $I$  from the agent's *state variables* to the substitution system's *changing parameters*—such that it includes all the possible ways in which the agent may have an effect on his epistemic artifact. Thus  $E(x_{TVSS})$  represents the effects of the tactile visual substitution system on the agent and  $I(x_A)$  represents the effects of the agent on the environmental aspect. Formally, then, we have the following coupled *dynamical laws*:

$$x_A' = A(x_A, E(x_{TVSS}), u_A^*)$$

$$x_{TVSS}' = TVSS(x_{TVSS}, I(x_A), u_{TVSS}^*)$$

(Where  $u_A^*$  and  $u_{TVSS}^*$  represent any parameters of  $A$  and  $TVSS$ , respectively, that are not affected by the coupling).

Describing the case, Beer (1995, 182) “cannot overemphasize the fundamental role that feedback plays in this relationship”. Any action that the agent takes affects the visual substitution system in some way through  $I$ , which in turn affects the agent himself, through the feedback he receives from the visual substitution system via  $E$ . Similarly, the visual substitution system's effects on the agent through  $E$  are fed back through  $I$  to in turn affect its own operation.

Thus, each of the two dynamical systems is continuously deforming the flow of the other (perhaps drastically if any coupling parameters cross bifurcation points in the receiving system's parameter space) and therefore influencing its subsequent trajectory. [...] It is therefore perhaps most accurate to view [the] agent and his [tactile visual substitution system] as mutual sources of perturbation, with each system continuously influencing the other's potential for subsequent interaction. (*ibid.*)

Accordingly, given 1) this kind of reciprocal direct dependence between  $A$  and  $TVSS$ ,

---

objects of perception can be coupled to an agent. This is an important note, because by not paying attention to the fact that *coupled systems can only consist of mutually interacting non-autonomous systems*, we will first have a wrong understanding of what a coupled system is supposed to be, which will, in turn, lead to wrong and, unsurprisingly, counterintuitive conclusions. As I note in section 4.2. Chemero appears to fall victim of such a mistaken understanding of coupled systems.

and 2) the definition of systems as sets of interdependent elements standing in interrelations on the basis of specific processes they participate in and give rise to, we can view the two *coupled nonautonomous* systems *A* and *TVSS* as a unified *autonomous* dynamical system, *ATVSS*.

Moreover, as Beer further notes, the state variables of this overall unified system is the union of the state variables of the agent and his visual substitution system. Accordingly, its dynamical laws are governed by all the interrelations (including the mutual interactions through *I* and *E*) among this larger set of state variables, *and the new dynamical properties these interrelations give rise to* (for example, the new *quasi-visual* experiences produced). In addition, any trajectories observed in the interaction between the agent and his tactile visual substitution system must be trajectories of the overall coupled system, *ATVSS*. So, “after transients have died out, the observed patterns of interactions between *A* and [*TVSS*] must represent an attractor of [*ATVSS*]” (Beer 1995, 183) (such as the *quasi-visual* perception of an object).

### 3.2. Two arguments for the ontological postulation of coupled systems.

Before proceeding further, here are the concluding remarks of the previous subsection again: the *coupling* of *nonautonomous* dynamical systems into one *autonomous* unified system can give rise to behavior that goes beyond the sum of the behaviors the individual subsystems can produce on their own; some of the geometrical properties of the *flow* of the larger system will not be attributable to either subsystem alone. “An agent’s behavior properly resides only in the dynamics of the coupled system [*ATVSS*] and not in the individual dynamics of either [*TVSS*] or *A* alone” (*ibid.*, 183).<sup>15</sup>

In other words, the mutual interaction between the agent and his tactile visual substitution system gives rise to new systemic properties (such as the new *quasi-visual* experiences produced, or new possibilities for interaction with the environment) that do not belong to any of the subsystems alone, but to the overall coupled system, *ATVSS*. We also noted that system individuation does not depend on any physical boundaries, but, instead, on the processes (and their properties) one is interested in, and which emerge out of component interactions. So, taking these last two points together provides us with a first reason to think that the (ontological) postulation of *coupled* systems is far from redundant.

Put another way, in such cases of continuous mutual interaction, the postulation of a single coupled system brings explanatory value. That is, the postulation of coupled

---

<sup>15</sup> Theiner (2011, sec. 3.3.2) appears to rely on similar considerations for his argument for HEC.

systems is necessary with respect to the explanation of certain systemic properties, which we would otherwise be at a loss how to account for. Accordingly, coupled systems are not open to the common eliminativist line that *Xs* do not exist because our best explanations are not committed to the existence of *Xs* (*i.e.*, that positing *Xs* does no explanatory work). Coupled systems such as *ATVSS* must, therefore, be taken as real.<sup>16</sup> We can call this the ‘systemic properties’ argument for the existence of *coupled* systems.

There is, however, yet another reason for which the postulation of *coupled* systems appears to be necessary, and which concerns the nature of the components’ interaction involved in such systems. Let me try to spell this out. As noted before, the *law* of a *nonautonomous* system is a function of both the system’s *state variables*  $x$ , and its changing *parameters*  $u(t)$ :  $x' = L(x(t), u(t))$ . That is, the set of possible behaviors of the system (*i.e.*, the *flow* of the system) depends on the changing aspects of the system (*i.e.*, the system’s *state variables*) and its *changing parameters*  $u(t)$ . As far as the parameters are concerned, we further noted that when they remain constant for the duration of the system’s operation they refer either to some of the system’s internal features (*e.g.*, the material it is made of), or to the stable background conditions the system operates in. When parameters change over time, however, they refer to the *inputs* to the system, which might originate from the dynamical environment, or some other well-defined system.

Accordingly, when we have two causally (but not mutually) dependent systems the *input* refers to the effects of the affecting system on the affected system. The *output*, on the other hand, refers to the affected system’s reaction (*i.e.*, the system’s behavior) to its *input*, but which, by hypothesis (remember there is only one-way dependence), has no substantial direct effects on the affecting system’s dynamics. Thus, it will only be represented by quantitative changes in one, or more of the affected system’s *state variables*. Overall, then, in cases of two causally (but not mutually) dependent systems, we can clearly tell the behaviors of each system apart, in terms of distinct inputs and outputs from the one system to the other.

Notice, however, that in cases of *nonautonomous coupled* systems, where some of the *changing parameters*  $u(t)$  of each system function as state variables of the other and vice versa, talk of inputs and outputs is inapplicable. The reason, as we saw Beer (1995, 182) pointing out, has to do with the fundamental role that *feedback loops* play in this relationship. The effects of the environment on the agent are *partly determined* by the agent’s own ongoing activity *at that time*, and *vice versa*. It is, therefore, impossible to decompose the ongoing causal effects in terms of distinct inputs and outputs from the

---

<sup>16</sup> Thanks to Mark Sprevak for pointing out this alternative formulation to me.

one system to the other.<sup>17</sup> The feedback loops between the two systems give rise to behavior, which cannot be broken down in terms of distinct inputs and outputs from the one subsystem to the other, and which therefore belongs only to the larger system, comprising of both the agent and the relevant environmental aspect.

Put another way, in cases where two *nonautonomous* systems mutually interact on the basis of feedback loops, there is an ongoing *causal amalgam* between the two units that disallows their decomposition into two separate systems on the basis of distinct *inputs* and *outputs*.<sup>18</sup> The reason is that the way each component is affected is not exogenous to the component itself, and so cannot be properly thought of as its *input*. Likewise, the way each component affects the other is directly and synchronically related to the component to be affected and so cannot be properly conceptualized as *output* of the affecting component. So, again, since we cannot disentangle the behavior of the two components in terms of distinct inputs and outputs from the one to the other, we must accept they constitute an overall system comprising of both of them. We can call this the ‘ongoing feedback loops’ argument for the (ontological) postulation of *coupled* systems.

So, to close this section, we now have two distinct arguments for the postulation of *coupled* systems. First, the properties that arise out of the interaction of *coupled systems* cannot be attributed to any of the contributing systems alone, but to the *coupled* system as a whole. Accordingly, we *have to* postulate the coupled system. (Alternatively, coupled systems are necessary for accounting for these systemic properties, so they cannot be ontologically eliminated). Second, in cases of ongoing feedback loops between *coupled* systems, there is a dense non-linear causal interdependence that disallows us to decompose systems in terms of distinct *inputs* and *outputs* from the one to the other (the reason being that the effects of each component to the other are not entirely endogenous to the affecting component, and vice versa). Accordingly, we *cannot but* postulate the coupled system.

---

<sup>17</sup> The simultaneity of the effects is crucial here. Specifically, it is very important to resist the temptation to visualize the case in terms of two components A and TVSS, whereby at time  $t_0$ , A is in state  $x_{A0}$ , which makes TVSS, at time  $t_1$ , enter state  $x_{TVSS1}$ , which then, at time  $t_2$ , makes A enter state  $x_{A2}$ , and so on. If this linear story were correct, then we would indeed be able to identify state  $x_{A0}$  as output of A and input to TVSS, and state  $x_{TVSS1}$  as output of TVSS and input to A, and so on. The differential equations, however, describing continuous time dynamical systems refer to infinitesimal differences (hence the name *differential* equations) in time, and so, theoretically,  $t_0 = t_1$ . Therefore, since  $x_{A0}$  and  $x_{TVSS1}$  are supposed to occur at the same time, there is no way in which they could be conceptualized in terms of inputs and outputs from A to TVSS, and vice versa. More precisely, system A’s subsequent state  $x_{A2}$  depends both on TVSS’s state  $x_{TVSS1}$  and its own state  $x_{A0}$  (which affects TVSS’s state  $x_{TVSS1}$ ) at time  $t_0 = t_1$ , and so on. Likewise, at  $t_3$ , TVSS’s state  $x_{TVSS3}$  depends both on A’s state  $x_{A2}$  and its own state  $x_{TVSS1}$  at time  $t_1 = t_2$ .

<sup>18</sup> Following the suggestion of an anonymous referee, I should here note that the benign form of causal circularity to be found in a *causal amalgam* is distinctly different from the metaphysically dubious forms of causal circularity to be found in the ideas of backwards and self-causation. Instead, it seems to be more akin to Lewis’ (1986) notion of ‘piecemeal causation’. See also (Theiner 2011, sec. 6.4.2)



Overall, then, we might say that the *constituents* of a system are the interdependent components, which, on the basis of feedback loops, give rise to the processes (and their properties) one is interested in, and which attracted the observer's attention to the relevant components in the first place. Before closing this section, however, let me also note how we can distinguish the overall system from other environmental aspects that may affect the system's performance but not constitutively so. For example, sunlight, the ambient temperature, or pressure might affect a system's temporal evolution. So long, however, as the system, on its part, has no (measurable) direct effect on those environmental aspects, such that they will not, in turn, affect back the system's performance, they can be safely considered as background conditions or inputs with a causal, but not constitutive effect on the system under consideration. Such causal effects will only be represented in the system's *dynamical law* as its (changing, or constant) *parameters* *u*.<sup>19</sup>

#### 4. Feedback Loops and Cognitive Extension

##### 4.1. Feedback Loops, the 'Causal-Constitution' Fallacy, 'Cognitive Bloat', HEMC, and HEC

The discussion of HEC was paused on the interrelated objections of the 'causal-constitution' fallacy and the 'cognitive bloat' worry; if we take the 'glue and trust' criteria to be both necessary and sufficient for cognitive extension then we are led to 'cognitive bloat', the reason being that the satisfaction of the said criteria cannot ensure that we have not committed the 'causal-constitution' fallacy. In other words, unless we have a principled way on the basis of which we can determine whether external (to an agent's brain, or organism) components are related to a cognitive process in a merely causal as opposed to constitutive way, 'cognitive bloat' will ensue. Moreover, since 1) no principled account that could draw such a distinction is in sight, and 2) HEMC is a rival hypothesis well suited to provide the same mechanistic explanation as HEC but without making the extra claim that whatever is causally related to a cognitive process is also part of it—thereby avoiding the 'causal-constitution' fallacy and the ensuing 'cognitive bloat' worry—we should follow HEMC theorists in retaining the term cognitive for the

---

<sup>19</sup> This apparently generates a problem with respect to the constant parameters that refer to the system's inherent features. Shouldn't the inherent features of a system count as constitutive of the system? Notice, however, that these parameters take their values from the realization basis of the system (*i.e.*, what it is made of). Thus, saying that those parameters of the system are not constitutive of it, really gives rise to the notion of multiple realizability and reveals what are the essential aspects of the system: *i.e.*, the relational properties of the components that arise out of the components' interactions. This also indicates that parameter spaces can provide information for more specific 'multiple realizability' claims.

neural, or at most the organismic processes, while fully recognizing the deep ways in which cognition is affected by extraorganismic factors; HEMC is supposed to recognize *all* kinds of cognitive embeddedness, while avoiding the ‘cognitive bloat’ worry. Cognitive scientists should, therefore, give up HEC on the face of HEMC.

Premise 1) of the previous reasoning, however, now seems questionable. The above considerations on system individuation provided us with strong support for claiming that when two systems are mutually interdependent on the basis of ongoing feedback loops, the interactive processes those systems engage in, and the properties they give rise to, belong to an overall coupled system consisting of both of them. In other words, *ongoing mutual interdependence on the basis of feedback loops is the criterion by which we can judge whether two seemingly distinct systems constitute an overall system, consisting of both of them*. Conversely, when no such mutual interaction is in play, but instead a system affects another one in an one-way dependence (*i.e.*, the activity of the affected system has no ongoing direct effect on the affecting system, such that no feedback loops are exhibited), then we have a paradigmatic case of a merely causal—as opposed to constitutive—dependence.

How is this supposed to help with respect to the ‘cognitive bloat’ worry? Recall the directory service case again. Even though the agent appears to satisfy the 3 ‘glue and trust’ criteria, no feedback loops between him and the directory service are in play. The person employing those resources does not engage in continuous reciprocal causation (CRC) with them. Those resources are simply causally, as opposed to constitutively, related to the agent in that they deliver her information whose formulation was entirely independent of her; there is no continuous mutual interaction between the agent and the epistemic artifact. If an account were to be provided for this case, then it would have to be a linear one, probably in testimonial terms.

Overall, then, we may claim that we have a case of a *cognitive* system being genuinely extended only in those cases where we face a task (*i.e.*, a process—and remember we individuate systems on the basis of the processes we are interested in) that we would intuitively like to call a cognitive one, and which was accomplished on the basis of continuous mutual interactions between the agent and his artifact. Otherwise—if the assumed cognitive task is not completed on the basis of CRC between the agent and his artifact—we may only talk about cognition being at most embedded or scaffolded.

Now, notice that using the *existence of continuous mutual interaction on the basis of feedback loops as a criterion of constitution* does not run against the spirit of HEC. As we noted earlier, it is only in cases where the CRC phenomenon between the internal and

external elements is manifested that we can safely talk about extended cognition. As Clark (2008, 131) notes “when we confront a recognizably cognitive process, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that recycled as inputs drive the process along, [...] any intuitive ban on counting *inputs* as parts of *mechanisms* seems wrong”.<sup>20</sup> The reason, we saw in the previous section, is that in cases where the outer and the inner contributions mesh with each other on the basis of feedback loops, DST treats the two components as a *coupled system* in its own right.

As the previous quote indicates, however, Clark treats CRC only as a sufficient condition on cognitive extension. This is probably because it appears to be too demanding a criterion with respect to certain cases that are meant to motivate the Extended Mind hypothesis (*e.g.*, the Otto and the shopping list cases; I will return to them in section 5). Instead, the only conditions that Clark treats as necessary—but which are insufficient—for cognitive extension are the 3 ‘glue and trust’ criteria. Since, however, the ‘feedback loops’ criterion provides an objective way for drawing systemic boundaries—overcoming in effect the ‘causal-constitution’ fallacy and the ‘cognitive bloat’ worry—I here wish to accentuate its importance by treating it as both necessary and sufficient for cognitive extension.

Apparently this makes the ‘glue and trust’ criteria seem redundant—which strictly speaking they are—but they still have a role to play. Reliability of the external resource, as well as frequent and easy access to it seem to be *practical preconditions* for the agent to be able to engage in continuous mutual interaction with it. The extent, however, to which each one of these criteria may need to be satisfied is going to differ from case to case (subjects can be very fast accustomed to the tactile visual substitution system, for example), but many times they will all have a significant guiding effect. The effect of transparent equipment, however—the feeling of ‘seeing-through’ the artifact; becoming one with it—will take place only after the agent has actually started mutually interacting with it. All we need, then, to accept the constitutive status of external elements within one’s overall cognitive mechanism is that the phenomenon of continuous reciprocal causation between the outer and the inner parts take place.

We may then treat the three ‘glue and trust’ criteria as conducive conditions on

---

<sup>20</sup> I must note, here, that because of the ‘ongoing feedback loops’ argument for the ontological postulation of coupled systems (section 3.2), the quoted phrase—literally interpreted—is actually incorrect. Properly speaking, when CRC is manifested, one of the points is that there are no distinct inputs and outputs between the interacting systems. Clark uses this wording as a metaphor to describe the feedback loops that generate the CRC phenomenon. Admittedly, however, it is a good way to visualize CRC so I may allow myself to use this phrasing to indicate CRC, but, when I do, I suggest that it be interpreted loosely.

HEC, but note that the only necessary and sufficient condition on cognitive extension (however this may be satisfied) is the manifestation of continuous mutual interactions between the agent and his artifact.<sup>21</sup> These ‘3 to 1’ criteria will ensure the integration of the external artifacts within the overall cognitive mechanism. Once the favorable conditions are there and feedback loops are in play, the ongoing interaction between what we would previously call ‘internal’ and ‘external’ deeply affects the ‘tendencies’, or even causes *bifurcations* in the agent’s or the artifact’s *phase portrait*, giving rise to behaviors that are distinctive of the extended cognitive system.

“To avoid thus begging the question, we should not operate with prior assumptions about where to place the causal-constitutive boundary, but wait on the results of explanation” (Hurley 2010, 106).<sup>22</sup> As it happens, within the framework of DST, the boundaries are not “exogenous to explanatory aims. In cognitive applications, the state space can extend to include dimensions whose variables are bodily and environmental as well as neural, as brain, body and environment interact in mutually shaping patterns” (*ibid.*, 130).

When those mutual interactions between the agent and his artifact are present,

---

<sup>21</sup> Think about a patient who has lost his sight and who comes out of a comma with an advanced tactile visual substitution system wired to his neural network. The substitution system is not currently activated (so he currently has neither constant nor easy access to it), but the patient has been hypnotized to trust the system when it will be activated, say, in a year’s time. I think there are no reasons why the substitution system shouldn’t count as part of his cognitive system on the day of activation. Although, admittedly, the fact that the agent is able to mutually interact with it indicates that easy access will be present after the activation date. For a discussion of a series of similar though experiments evoking our intuitions regarding the idea of cognitive integration see (Pritchard 2010) and (Palermos 2011).

<sup>22</sup> Adams & Aizawa (2001; 2008, 2010) claim that the mark of the cognitive is the manipulation of representations with underived content, which is plausibly (at least for the time being) not a feature of any external process. Thus they avoid begging the question against externalism when they put forward the ‘causal-constitution’ fallacy.

It is not clear, however, what Adams and Aizawa have in mind and how the said criterion is supposed to promote internalism. One complication is that there is no commonly accepted and non-problematic theory of underived/intrinsic/original content. Second, as Menary (2006) points out, the way Adams and Aizawa make use of the idea that the mark of the cognitive is the manipulation of representations with underived content is either too strict—such that it rules out many of the paradigmatic cases of cognition—or too permissive—such that it does not disallow cognitive extension.

Finally, after a lot of argumentation, Adams and Aizawa (2010, 70) write: “Clearly we mean that if you have a process that involves no intrinsic content, then the condition rules that the process is non-cognitive”. Focusing, however, on the extended cognitive processes that emerge out of the continuous reciprocal causation between extra-neural elements and the organismic brain would surely guarantee the involvement—courtesy of the latter component—of manipulations of representations with underived content, were such entities to exist.

For more details on the debate on underived content as the mark of the cognitive, see (Clark 2008; 2010b), (Menary 2006), (Adams and Aizawa 2001; 2008; 2010), (Ross and Ladyman 2010).

Another significant consideration on this topic is that within the dynamical approach to cognition, whether representations exist at all, or what may count as a representation are open questions whose answers might be conceptually alien to the classical computationalist approach that Adams and Aizawa seem to embrace. Indicatively, van Gelder (1995), Pollack (1990), Spivey (2007) and Petitot (1995) have suggested that attractors in the state space of the cognitive system are plausible candidates for bearing representational significance. It is, therefore, an open question whether attractors that arise from the agent’s interaction with some artifact could bear non-derived content, and it is by no means clear why we should pre-theoretically exclude such a possibility.

DST provides a clear case for HEC. In fact, what the above discussion demonstrates is that whenever the environment affects the agent's cognitive loops in a one-way dependence, cognition is indeed embedded à la HEMC. But whenever one's organismic cognitive apparatus interacts with the environment on the basis of ongoing feedback loops, then the HEMC theorist must recognize that cognition is not only embedded, but also extended à la HEC.

Therefore, "the issues between internalism and externalism should be", and can be I may add, "resolved bottom up by such scientific practice, not by advance metaphysics" (Hurley 2010, 107).<sup>23</sup> The two arguments we considered from DST—the best conceptual tool for modeling dynamical systems interacting with their environment—demonstrate that coupled (*a.k.a.* extended) cognitive systems are not only ineliminable, but they are also necessary for making sense of certain cognitive behaviors.

#### 4.2. Similar Arguments

Invoking DST in favor of HEC is a move that, one way or another, has been made by some authors in the past. Accordingly, I would here like to make clear how the attempts of some of those authors relate to the present arguments, and comment on the extent to which I share their interpretations. Two of the three lines of reasoning I have in mind resemble the two arguments I offered in favor of the postulation of coupled systems in the previous section. The third concerns a solution to the 'causal-constitution' fallacy by focusing on nonlinear relations between elements, but it is problematic in that its author attempts an unwarranted interpretation of it.

To start with, the first point comes from Clark who notes that—contrary to the spirit of HEMC—the properties of the overall interdependent system are not always equal to the mere addition of the participating systems' properties (2008, 116):

A further reason to resist the easy assimilation of HEC into HEMC concerns the nature of the interactions between the internal and the external resources themselves. Such interactions, it is important to notice, may be highly complex, nested, and non-linear. As a result there may, in some cases, be no viable means of understanding the behavior and potential of the extended cognitive ensembles by piecemeal decomposition and additive reassembly. To understand the integrated operation of the extended thinking system

---

<sup>23</sup> For an objection to the claim that scientific practice can resolve this debate see (Sprevak 2010). In so arguing, Sprevak has in mind that both HEC and HEMC will produce the same causal explanations—since they are both interested in the interaction between the agent and her environment—and so scientific explanation will not be able to produce a clear verdict regarding the success of HEC over HEMC, or the other way around. My reasons for thinking that it can are quite different: Should there be cases where the accomplishment of some cognitive task is the product of the mutual interaction between the agent and some artifact, then given the DST framework and the concept of coupled systems (as well as the absence of a plausible mark of the cognitive) any theorist will have to conclude that there is a coupled cognitive system that operates on the basis of extended cognitive processes.

created, for example, by combining pen, paper, graphics programs, and a trained mathematical brain, it may be quite insufficient to attempt to understand and then combine (!) the properties of pens, papers, graphics programs, and brains.

Previously we noted that when continuous mutual interactions are present, the HEMC theorist must accept that cognition is not only embedded, but also extended à la HEC. In line with that comment, Clark explains above that when mutual interactions are in play, the extra neural elements and the organismic brain should not be considered—as HEMC, in effect, only allows—as distinct systems whose interactive performance could be linearly analyzed.<sup>24</sup> Instead, in such cases, the only possible way to study all their properties will be as one unified system. This, we can see if we go back to section 3.2, was also the point of the ‘systemic properties’ argument for the existence of coupled systems.

The second consideration comes from Varela who writes: “If one says there is a machine M in which there is a feedback loop through the environment, so that the effects of its output affect its input, one is in fact talking about a larger system M’ which includes the environment and the feedback loop in its defining organization” ((Varela 1979, 158); quoted by (Hurley 1998, 104). Again, we may notice how this reasoning reverberates the ‘ongoing feedback loops’ argument for the postulation of coupled systems, presented in section 3.2.

And, finally, the third suggestion comes from Chemero (2009) who notes—without providing any further arguments though—that the existence of non-linear relations between two components is what is required for them to constitute an overall system comprising of both of them. Specifically, his comments go like this (*ibid.*, 31-32):

Agents and environments are modeled as nonlinearly coupled dynamical systems. Because the agent and environment are nonlinearly coupled, they form a unified, nondecomposable system, which is to say that they form a system whose behavior

---

<sup>24</sup> One may think that the HEMC theorist would be happy to accept that an agent might be in a continuous and reciprocal causal relation with some aspect in the environment, but still deny cognitive extension. This is actually not an option for the HEMC theorist, however: Given the conceptual framework of DST, in such cases, HEMC collapses into HEC. In fact, in chapter 7 of his book (2009), Rupert concedes that, in such cases, DST can provide strong support to HEC (2009, 131-4). It is telling that none of the dynamical models that he considers in favor of HEMC concerns a two-way interaction between the organism and some particular environmental aspect. For more details see (Rupert 2009, 137-149). For more on HEMC’s failure of aggregativity see (Theiner 2011, 126-130). Wheeler is one more theorist who has claimed that CRC between brain, body, and environment is a powerful indication of extended cognition. Specifically he writes (2005, 265): “[W]here the complex channels of continuous reciprocal causation cross back and forth over the physical boundaries of skull and skin, the cognitive scientist, operating from a functional or organizational perspective, may face not (i) a brain-body-environment system in which brain, body, and environment form identifiable and isolable subsystems, each of which contributes in a nontrivial way to adaptive success, but rather (ii) just one big system whose capacity to produce adaptive behavior must be understood in an holistic manner”.

cannot be modeled, even approximately, as a set of separate parts. [...] When the agent and environment are nonlinearly coupled, they, together, constitute a nondecomposable system, and when that is the case, the coupling-constitution fallacy is not a fallacy. In other words, the coupling-constitution fallacy is only a fallacy when the coupling is linear.

So Chemero, who also draws from DST, offers a solution to the ‘causal-constitution’ fallacy, which is technically equivalent to the one we offered before (I will explain why shortly).<sup>25</sup>

Chemero, however, further takes his view—*Radical Embodied Cognitive Science*—to be a version of the extended cognition hypothesis (Chemero & Silberstein 2007). In addition, he takes his view to be anti-representationalist and in favor of direct perception. Specifically, the way he explains the lack of need for representations is by claiming that since the agent is coupled to his environment, then we do not need to claim that the agent needs to represent the objects of his perception; instead, due to his coupling with them, both the agent and the objects of his perception are part of an overall cognitive system. In order to be perceived, therefore, objects do not need to be *re-presented*; they are already present within the cognitive agent, so to speak. Specifically he writes (2009, 114-115):

An outfielder effectively tracks a fly ball when the light reflecting off the ball makes contact with her eyes, and she moves her eyes and head so as to maintain that contact. In terms of the physics of the situation, the ball, the outfielder, and the intervening medium are just one connected thing. In effective tracking, that is, the outfielder, the ball and the light reflected from the ball to the outfielder form a single coupled system. [...] There are two relevant consequences of taking tracking as the model of direct perception. First, we can see that perception is, by definition, direct. Perception is always a matter of tracking something that is present in the environment. Because animals are coupled to the perceived when they track it, there is never need to call upon representations during tracking.

There is a serious problem with Chemero’s reasoning, however; arguing in this way, he actually commits the fallacy that he has previously offered a way out of. Although he is correct to claim that the existence of non-linear relations is both necessary and sufficient for the emergence of coupled systems, Chemero misses a crucial point: These *non-linear relations arise only out of cooperative or inhibitory feedback loops between interacting parts*. Contrary to what he claims, therefore, objects of perception are neither non-linearly related nor, thereby, coupled to their perceiver; no feedback loops are present in cases of perception. Instead, the agent is only linearly dependent on the objects he perceives.

Even if sensorimotor theories of perception—to which I am favorably disposed—

---

<sup>25</sup> For an additional overview of the ‘causal-constitution’ fallacy and some further attempts to its solution—some of which are not entirely independent of DST—see Kirchhoff (*forthcoming a*; *forthcoming b*).

are true (Noë 2004), the point remains the same.<sup>26</sup> The agent might need to move around or track the object of perception in order to actually perceive it, but in so doing he does not change the object of perception in any way that affects him back, and so on. Instead, the perceived environment remains unaffected by the agent's actions and, so, there are no feedback loops between the agent and the object he perceives. Accordingly, since it is only such loops that can give rise to non-linear relations, it would be incorrect to say that I am non-linearly related and thereby coupled to the Kadinsky poster that hangs on the wall opposite of where I am sitting.

This might be bad news for Chemero's arguments for anti-representationalism and direct perception—though there might be other ways to argue for these positions from the DST perspective—but it's certainly good news for dealing with the 'causal-constitution' fallacy and the 'cognitive-bloat' worry. For if Chemero's above reasoning were correct, it would mean that almost anything I interact with would be part of my cognitive system, which is exactly what the 'cognitive bloat' worry warns us against. It is, therefore, important that care be taken in how one interprets the idea of 'non-linear relations' and the claims of constitution they lead to, or else counterintuitive and—even worse—misleading conclusions will follow.

Focusing, instead, on *continuous mutual interactions* (which entail the non-linear relations that Chemero is appealing to) provides a safer way for individuating cognitive systems. This is not the only advantage of the approach, however. On the basis of the 'ongoing feedback loops' and 'systemic properties' arguments we can further understand *why* Chemero holds that any non-linearly coupled (that is, properly coupled) systems will form an overall "unified, nondecomposable system [...] whose behavior cannot be modeled, even approximately, as a set of separate parts": 1) Coupled systems bear systemic properties that belong to none of the interacting subsystems, but to the coupled system as whole. 2) Coupled systems form *causal amalgams* whose behavior cannot be decomposed in terms of distinct inputs and outputs from one subsystem to the other.

---

<sup>26</sup> I should here note that I do not mean to criticize enactive theories of perception—at least not when interpreted as providing support to merely embodied cognition. It is true, however, that, by invoking DST, the perceiver's interaction with his environment (on the basis of sensorimotor contingencies) might be interpreted as a case of cognitive extension. Such an interpretation would indeed be problematic as it would commit the same mistake that Chemero makes in his support for direct realism/anti-representationalism (*i.e.*, that the perceiver and the object perceived are parts of an overall extended system).



### 4.3. Thermostats, Air-Conditioners, and Dynamical Systems Theory

Having drawn attention to how easy it is to draw the wrong boundaries, I will here attempt to demonstrate how focusing on feedback loops helps distinguish between what constitutes and what merely causally affects a system. I will do so by way of examining two examples that Adams and Aizawa appeal to in order to illustrate the two versions of the ‘causal-constitution’ fallacy. Notice, however, that since we have said that systems are individuated on the basis of the processes they give rise to—such that systems may extend in virtue of interactive processes that extend beyond the boundaries of systems—it is not clear which of the two versions of the ‘causal-constitution’ fallacy we would commit in the absence of the ‘feedback loops’ criterion of constitution. To demonstrate, the force of the criterion, however, here is how it rules with respect to both of the following two examples.

Simple version: “Consider the expansion of a bimetallic strip in a thermostat. This process is causally linked to the motion of atoms of the air in the room the thermostat is in. The expansion of the strip does not, thereby, become a process that extends into the atoms of the air in the room” (2008, 91). Agreed, but the reason for this is that the bimetallic strip has no direct effects on the air molecules such that it will in turn affect itself. Of course, Adams and Aizawa could further point out that as time goes by the expansion of the bimetallic strip will turn the heating component off and so will, in turn, indirectly affect the motion of atoms of the air in the room. Notice, however, that this is not the kind of differential interaction that the coupling arguments of the previous section are referring to. Rather, if one provided a model for this case, the temperature of the room would only appear as one of the changing parameters,  $u(t)$ , in the thermostat’s *dynamical law*; it is one of the *inputs* to the thermostat. The thermostat, in other words, is a *nonautonomous* system, affected by—amongst other things—the temperature of the room, but it is not coupled to it. The ongoing feedback loops criterion rules correctly in this case.

Systems version: “The Liquid Freon<sup>TM</sup> in an older model air conditioning system evaporates in the system’s evaporation coil. The evaporator coil, however, is causally linked to such things as a compressor, expansion valve, and air conditioning ductwork. Yet the evaporation does not extend beyond the bounds of the Freon<sup>TM</sup>. So a process may actively interact with its environment, but this does not mean that it extends to its environment” (*ibid.* 91). Again agreed; the evaporation process does not extend to the rest of the system. But the whole process of *conditioning the air* is an overall process whose realization very much depends on the *mutual interaction* between the evaporation

coil and the rest of the components of the air conditioning system. Surely, no one would like to identify the air conditioning system with the evaporation coil.<sup>27</sup>

## 5. Discussion and Conclusion

The claim therefore is that the ‘3 to 1’ criteria—the 3 ‘glue and trust’ criteria perceived as conducive preconditions to the ‘ongoing feedback loops’ (or CRC) criterion—are crucial for cognitive extension. It should be interesting, however, to ask what are the ramifications of treating the CRC phenomenon as both necessary and sufficient for HEC, and what is the underlying reason that motivates this addition. Moreover, could this last question concern a fundamental feature of (‘higher-order’) cognition?

So, first, let’s begin by asking whether CRC is too stringent a criterion. Earlier we noted that it excludes the use of the phonebook and directory service as plausible candidates for cognitive extension. But what about other alleged cases of extended cognition? Although answering this question will each time require to submerge in the details of the target case, the CRC criterion will generally exclude any case where there are no ongoing feedback loops between the agent and the target environmental aspect. So, for instance, shopping lists will not count as cognitive extension; the agent may have created the list for use at some point in the future, but, crucially, during the process of retrieval, the agent does not perform any operations on the basis of the list, which ‘recycled as inputs’ drive his cognitive loops along. Since there are no feedback loops between them, the list is not coupled to the agent.

What, then, about the Otto case, which seems to be similar to the shopping list case? Otto does not continuously interact with his notebook. Instead, when he needs some old information he looks it up and when he needs to store some new information he writes it down. It might, therefore, be true that Otto’s relationship with his notebook is best described as a case of one-way dependence and thus, by the lights of the CRC criterion, not count as a case of cognitive extension. Depending, of course, on one’s intuitions about the Otto case, this might count both as an advantage and a

---

<sup>27</sup> To provide a more charitable reading of Adams and Aizawa’s objection, what they seem to suggest is that the *essential* process of cooling the air in the whole process of air conditioning does not pervade all aspects of the air conditioning system. In the first place, however, it seems quite suspicious to claim that the evaporation coil is performing the only essential process in air conditioning. Second, in order to draw the parallel between air conditioning and cognitive systems, as Adams and Aizawa wish to do, they need to find the analogue of the ‘cooling component’ in the case of cognition, such that cognition is restricted within the agent’s brain (assuming that the relevant component, or processes can only be found, at least for the time being, within organismic brains). In the absence of such a plausible ‘mark of the cognitive’, however, when Adams and Aizawa run this version of the ‘causal-constitution’ fallacy against the HEC theorist, they either beg the question, or the ball is in their side of the court to convincingly argue for such a ‘mark of the cognitive’. See also fn. 30.

disadvantage. In any case, disallowing the Otto and his notebook from counting as an instance of cognitive extension should not particularly trouble proponents of extended cognition, as it was primarily meant to motivate the related, but more provocative hypothesis of the Extended Mind.

Even so, however, it should be noted that it is not crystal-clear whether Otto's cognitive processing does not extend to his notebook. For as far as the CRC criterion is concerned, the problem is that Otto's interaction with the notebook is not continuous, but is instead interrupted. Notice, however, that lots of instances of normal memory appear to be just like this: a one-step process of storage, or retrieval which hardly ever extends over long periods of time, and while the whole process may appear to be inactive for most of the time, it is always in the background, ready to reliably bring about the desired storage, or retrieval effect. But now notice that Otto's situation is no different. He may not continuously interact with the device—though, probably, he very rarely lets go of it—but as far as the process of memory is concerned, whenever he needs to store or retrieve information, he always finds himself interacting with the notebook. In effect, the answer to the question whether, by the lights of the CRC criterion, Otto's distinct time (but still ongoing) interaction with his notebook counts as a case of cognitive extension depends on whether *distinct time dynamical systems* can produce the same coupling arguments that *continuous time dynamical systems* did in section 3.2.<sup>28</sup>

As far as the rest of the paradigmatic cases of extended cognition are concerned, it seems that when writing or solving a mathematical problem by using pen and paper, employing microscopes and telescopes, perceiving the world through a tactile visual substitution system and so on are cases where the agent continuously interacts with his artifacts in a two-way direction.

Consider, for instance, telescopic observation. Making observations through a telescope is a dynamic process that requires a great deal of experience in operating the artifact. Moving the telescope around while adjusting the lenses, generates certain effects (*e.g.*, shapes on the lens of the telescope), whose feedback *drives* the ongoing cognitive loops along. Eventually, as the process unfolds, the coupled system of the *agent and his telescope* is able to identify—that is, *see*—recognizable objects in space (*e.g.*, stars, planets, comets, galaxies et cetera).

---

<sup>28</sup> As Clark suggests (2010c, 1060) “the case of Otto is perhaps best seen as a rather simple, slow timescale, version of such loop”. It should be further noted, however, that if we imagine Otto being very well trained and organized, constantly having his notebook open, going back and forth over its pages, looking up its contents as well as taking notes about almost every single detail that would normally be registered in his biological memory, then we can say that he does continuously interact with it in a way that satisfies the CRC criterion. Accordingly, depending on how we imagine the Otto case it may already count as a case of cognitive extension, without the need for any coupling arguments from distinct time dynamical systems.

So, crucially, the epistemic artifact and the agent do interact in a continuous and mutual way. Therefore, this is not a case of an agent merely using an instrument; the interaction between the agent and the telescope is not linear such that the two systems can be neatly decomposed by having their function described in terms of distinct inputs and outputs from the one system to the other. Consequently, it would be a vain attempt to analyze telescopic observation by providing a linear story, whereby the telescope provides the agent with fully articulated pieces of information. Instead, agent and telescope process the relevant information in cooperation, thereby forming a coupled system. In a technical but also intuitive sense the cognitive process of telescopic observation (and its properties) belongs to the overall extended system, and not to any of the contributing systems alone.

This sort of description, of course, can fit a wide range of other cases as well. In section 2.1 we encountered the problem of long multiplication and in section 3.2 we considered the case of an agent perceiving his environment through a tactile visual substitution system. Less well-known, but equally interesting examples are tactile vestibular substitution systems (Tyler et al. 2003), auditory visual substitution systems (e.g., vOICe) (Auvray et al. 2007), and magnetic perception (Nagel et al. 2005). Accordingly, even though the CRC criterion excludes shopping lists and similar cases from the list of possible candidates for cognitive extension, and while it does not provide a conclusive verdict on the Otto case, it is not in danger of being so strict as to render cognitive extension a metaphysically possible, yet so far unrealized phenomenon.

Second, another related worry would be to ask whether the CRC criterion is aberrant in the sense that it is not something that is normally required by the rest of the processes to be found within one's head, and which are paradigmatically recognized as cognitive. If that is the case, then the CRC criterion would seem like an ad hoc move, which might be necessary—but not naturally motivated—for adjudicating cognitive extension.<sup>29</sup>

Notice, however, that as scientific observations of the human and of many primates' brain indicate, it is never the case that only one brain area is in operation. Instead, several parts of the brain always work in parallel. Therefore, far from being an exception to the rule, it appears that ongoing feedback loops between different areas in the brain *are* the rule. Take for example the rather simple cognitive task of reaching for

---

<sup>29</sup> Adams and Aizawa (2008) and Sprevak (2009) have argued that even the 'glue and trust' criteria are too strong, such that they may exclude several in-the-head processes, which are nevertheless considered to be paradigmatic parts of one's overall cognitive economy. The present approach can overcome this problem since it treats the 'glue and trust' criteria as merely 'conducive practical preconditions' for CRC, which is the only necessary and sufficient condition on HEC.

and grasping a mug sitting before you on the desk. In order for this to happen, what is actually required is the delicate use of visually received information by functionally and neuro-anatomically distinct subsystems operating together.

In fact, such considerations have led Dennett and Clark to oppose the traditional conception of the mind as a central, homoncular chooser towards whom all significant information is directed such that it can make the final choice. As they note, if that were the case, it would be deeply problematic; suppose only your frontal lobes have the final say. “Does this shrink the physical machinery of the mind and self to just the frontal lobes? What if, as the philosopher Daniel Dennett suspects, no mental subsystem has always and everywhere the final say? Has the mind and self simply disappeared?” (Clark 2006, 19).

Dennett and Clark, instead, put forward a view of the mind as ‘tools all the way down’ operating in collaboration without the need of a central user. Genuine cognition just emerges from the interactions between these various kinds of mind tool. “No single tool amongst this complex kit is intrinsically thoughtful, ultimately in full control, or plausibly identified as the inner ‘seat of the self’. We (we human individuals) just are these shifting coalitions” (Clark 2006, 20).<sup>30</sup>

But then if that’s how cognition arises—out of the interaction between different neural components—CRC appears again to be the rule. Far from being under-motivated, then, it now seems that it captures a fundamental aspect of cognition. And, in return, if cognition just is the parallel processing of different tools operating in collaboration, it is even more compelling to count *any* tool, which when appropriately coupled enters this coalition of tools, as a genuine part of one’s cognitive economy.

Claiming, of course, that every process, which may qualify as cognitive, must involve multiple components operating in parallel on the basis of ongoing feedback loops is a strong claim that may fly on the basis of empirical evidence. More specifically, as things stand, it is difficult to deny that there could be lower-level cognitive processes in the brain, which are wholly performed by just one brain area. Lower-level cognitive processing might indeed not involve CRC between different components. We may, however, want to call such processes mere ‘brain functions’, as opposed to full-fledged cognitive processes. In any case, ‘higher-order’ cognitive processing, such that it may qualify as a ‘cognitive task’ (and here I set the barrier quite low so as to include ‘reaching for and grasping’ an item), will most probably have to always involve CRC

---

<sup>30</sup> Although it is beyond the scope of the paper to expand on this issue, notice, again, that this picture of the mind gives rise to new ways of understanding what kinds of representation might be central for genuine cognition that cuts across the distinction between derived and underived content. For more on this see (Clark 1999) and Dennett (1987; 2000; 2001)

between different brain areas. Therefore, strictly speaking, the claim that ‘CRC is the rule’ does not necessarily mean that all cognition must involve CRC. Instead, it is intended to convey the idea that ‘higher-order’ cognitive processes (and let me note again how low the bar is set)—*i.e.*, cognitive processes that cognitive scientists paradigmatically focus on—can only arise on the basis of dense and mutual interactions between different processing components.

The ‘feedback loops’ criterion clearly captures this multiple-components nature of the mind, allowing at the same time for cognitive extension to take place. The overall picture, then, is that cognition typically involves the parallel interdependent operation of tools, and it probably comes in degrees. The more tools, the more cognitive power—but, again, there is no reason to exclude from this cognitive coalition all extra-organismic processes. To be sure, the superiority of human minds over the minds of other mammals, or primates may still be the outcome of bigger brains,<sup>31</sup> but as human culture evolves, the human beings’ cognitive capacities evolve as well, and that hardly ever could be seen as a difference in the organismic apparatus. Instead, establishing appropriate patterns of causal connections with bio-external structures, under specific circumstances appears to *transform* our effective cognitive architecture. Of course,

One might reply by denying that there really are many such cases, or by denying that they are of any great cognitive significance. But neither of these deflationary moves seems warranted. We humans are constantly engineering and re-engineering our worlds so as to provide for more and more of these ‘loopy opportunities’. And we go to great lengths to ensure that the right non-biological materials and circuits are mostly available whenever the relevant cognitive task is encountered (Clark 2010c, 1061).

To close, then, it is the nature and theoretical significance of these ‘loopy opportunities’ that the present paper was meant to clarify and accentuate. Cognitive science is a wildly growing and flourishing field that expands in every dimension, sometimes revealing the existence of loopy cases. The discovery of such cases and the present arguments from DST indicate not just the theoretical and philosophical possibility of cognitive extension, but its actual realization as well.

---

<sup>31</sup> Opposable thumbs appear to be another very important ingredient, which, by the way, is also associated with complex tool-use.

**Acknowledgments**

I am particularly thankful to Andy Clark, Duncan Pritchard, Mark Sprevak, and Andrew Wilson for very useful discussions and comments on previous drafts. I am also thankful to two anonymous referees for *Cognitive Systems Research*.

## References

- Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*. Blackwell Publishing Ltd.
- (2010). 'Defending the Bounds of Cognition'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (2001). 'The bounds of cognition'. *Philosophical Psychology*. Vol. 14, No.1, 43-64.
- Auvray, M., Hannequin, S., and O'Regan, J., K. (2007) 'Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with 'The vOICE', *Perception*, 36(3): 416-430.
- Bach-y-Rita, P., and S. W. Kercel. (2003). 'Sensory substitution and the human- machine interface', *Trends in Cognitive Science* 7, no.12 541-6.
- Beer, R. (1995). 'A dynamical systems perspective on agent-environment interaction'. *Artificial Intelligence* 72, 173-215.
- Bressler, S., and Kelso, J. (2001). 'Cortical coordination dynamics and cognition'. *Trends in Cognitive Sciences* 5, 26-36.
- Burge, T. (1993). 'Content Preservation', *The Philosophical Review*, 102, 457-88.
- Clark, A., and Chalmers, D. (1998). 'The Extended Mind', *Analysis* 58, no. 1: 7-19.
- Clark, A., (1998). 'Magic Words, How Language Augments Human Computation'. In *Language and Thought: Interdisciplinary Themes*, P. Carruthers and J. Boucher (Eds) (Cambridge University Press: Cambridge, 1998).
- (2001). 'Reasons, Robots, and the Extended Mind', *Mind and Language*, 16;2: 121-145.
- (2006). 'Soft selves and Ecological Control'. In *Distributed Cognition and the Will*, D. Spurrett, D. Ross, H. Kincaid and L. Stephens (eds). MIT Press, Camb. MA
- (2007). 'Curing Cognitive Hiccups: A Defence of the Extended Mind', *The Journal of Philosophy*, 104: 163-192.
- (2008). *Supersizing The Mind*. Oxford University Press.
- (2010a). 'Memento's Revenge: The Extended Mind, Extended'. In the *Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (2010b). 'Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (2010c). 'Much Ado About Cognition'. *Mind*, Vol. 119, Issue 476, Pp. 1047-1066.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT press.



- Chemero, A. & Silberstein, M. (2007). 'Defending Extended Cognition'. In Proceedings of the 30th Annual Meeting of the Cognitive Science Society, ed. B. Love, K. McRae, and V. Sloutsky. Austin, Texas: Cognitive Science Society.
- Dale, R., and Spivey, M. (2006). 'Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation'. *Language Learning*, 56, 391–430.
- Dennett, D. (1987). *Intentional Stance*. MIT press.
- (2000). 'Making Tools for Thinking'. In Dan Sperber (ed.), *Metarepresentations*. Oxford University Press.
- (2001). 'Things about Things'. *Foundations of Cognitive Science*.
- Hurley, S. (2010). 'The Varieties of Externalism'. In *the Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- Kelso, J. A. S., and Engström, D. (2006). *The Complementary Nature*. Cambridge, Mass.: MIT Press.
- Kirchhoff, M. (forthcoming a). 'Extended Cognition and the Causal-Constitution Fallacy: In Search of a Diachronic and Dynamical Conception of Constitution', *Philosophy and Phenomenological Research*.
- (forthcoming b). 'Extended Cognition and Constitution: Re-evaluating the Constitutive Claim of Extended Cognition'. *Philosophical Psychology*.
- Lewis, D. (1986). 'Postscripts to "Causation"'. In D. Lewis (Ed.), *Philosophical papers* (Vol. 2, pp. 172-213). New York: Oxford University Press.
- Menary, R. (2006). 'Attacking the Bounds of Cognition', *Philosophical Psychology*. Vol. 19, No. 3, June 2006, pp. 329-344.
- (2007). *Cognitive Integration: Mind and Cognition Unbound*. Palgrave MacMillan.
- (2010a). 'Introduction: The Extended Mind in Focus'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- (2010b). 'Cognitive Integration and the Extended Mind', In *the Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- Michaelian, K. (2012). 'Is external memory memory? Biological memory and extended mind', *Consciousness and Cognition* 21(3): 1154-1165.
- Nagel SK, Carl C, Kringe T, Märtin R, König P. (2005). 'Beyond sensory substitution - learning the sixth sense', *Journal of neural engineering* 2(4):R13-26.
- Noë, A. (2004). *Action in Perception*. Cambridge: MIT Press.

- Palermos, S., O. (2011). 'Belief-Forming Processes, Extended'. *Review of Philosophy and Psychology*, 2: 741–765.
- Pollack, J. (1990), 'Recursive Distributed Representations', *Artificial Intelligence*. 77-105.
- Putnam, H. (1975). 'The Meaning of "Meaning"'. In K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.
- Petitot, J. (1995). 'Morphodynamics and Attractor Syntax', in *Mind as Motion: Explorations in the Dynamics of Cognition*, Port, R., F. and van Gelder, T. (eds). MIT press.
- Pritchard, D. (2010). 'Cognitive Ability and the Extended Cognition Thesis', *Synthese*, 175: 133-151.
- Ross, D. and Ladyman, J. (2010). 'The Alleged Coupling-Constitution Fallacy and the Mature Sciences'. In *The Extended Mind*. (2010), Menary (ed.) Cambridge, Massachusetts, MIT press.
- Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. New York: Cambridge University Press.
- Rowlands, M. (2003). *Externalism: Putting Mind and World Back Together Again*. Montreal: McGill-Queen's University Press.
- (2009). 'Extended Cognition and the Mark of the Cognitive'. *Philosophical Psychology*, 22(1); 1-19.
- Rupert, R. (2009). *Cognitive Systems and the Extended Mind*, Oxford University Press, New York.
- (2004). 'Challenges to the Hypothesis of Extended Cognition'. *Journal of Philosophy*, 101: 389-428.
- Spivey, M. (2007). *The Continuity of the Mind*. Oxford University Press.
- Spivey, M., and Dale, R. (2006). 'Continuous temporal dynamics in real-time cognition'. *Current Directions in Psychological Science*, 15, 207–211.
- Sprevak, M. (2009). 'Extended Cognition and Functionalism'. *The Journal of Philosophy*. 106, 503-527.
- (2010). 'Inference to the hypothesis of extended cognition'. *Studies in History and Philosophy of Science*.
- Theiner, G. (2011). *Res Cogitans Extensa. A Philosophical Defense of the Extended Mind Thesis*. Peter Lang GmbH, Europaischer Verlag der Wissenschaften
- Thompson, E., and Varela, F. (2001). 'Radical Embodiment: Neural dynamics and consciousness'. *Trends in Cognitive Sciences*, 5, pp. 418–425.
- Tyler M, Danilov Y, Bach-y-Rita P (2003). "Closing an open-loop control system: vestibular substitution through the tongue". *Journal of Integrative Neuroscience*, 2:159-164.

- Van Gelder, T. (1995). 'What Might Cognition Be, If Not Computation?', *The Journal of Philosophy*, Vol. 92, No. 7 (Jul., 1995), pp. 345-381.
- Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). 'The brainweb: Phase synchronization and large-scale integration'. *Nature Reviews Neuroscience*, 4, pp. 229-239.
- Wheeler, M. (2004). 'Is Language the Ultimate Artifact?', *Language Sciences* 26: 693-715.
- (2005). *Reconstructing the Cognitive World*. MIT Press, Cambridge, Massachusetts.
- Wilson, R. A. (2000). 'The mind beyond itself'. In D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*, New York University Press, 31-52
- (2004). *Boundaries of the Mind: The individual in the Fragile Sciences: Cognition*. New York: Cambridge University Press.

<b>Dynamical Systems Theory Technical Terms</b>	<b>Definition</b>
<i>state variables, x</i>	the values of the changing aspects of the system
<i>dynamical law, L</i>	a set of differential equations, $x' = L(x, u)$ that regulates the change of the <i>state variables</i>
<i>trajectory</i>	a sequence of states generated by the <i>dynamical law</i> , starting from some initial state $x_0$
<i>state space</i>	the set of all possible values of <i>state variables</i>
<i>Flow</i>	the set of all possible <i>trajectories</i> through every point in the <i>state space</i>
<i>limit set</i>	sets of points that are unaffected by the <i>dynamical law</i>
<i>attractor</i>	a <i>limit set</i> that gravitates <i>trajectories</i> passing through all nearby states
<i>basin of attraction</i>	a set of initial states that converge to a given <i>attractor</i>
<i>transient</i>	the portion of a <i>trajectory</i> that is found within a <i>basin of attraction</i> but which does not lie in the <i>attractor</i> itself
<i>repellor</i>	an unstable <i>limit set</i> whose nearby <i>trajectories</i> diverge from it
<i>phase portrait</i>	the graphical representation of <i>attractors</i> , <i>repellors</i> and <i>basins of attraction</i> (i.e., the different phases) the system can enter into
<i>parameters, u, (fixed)</i>	the values of the <i>dynamical law</i> that are determined by the internal features of the system (e.g. the material it is made of), or the background conditions it operates in
<i>parameters, u(t), (changing)</i>	the input to a <i>nonautonomous system</i>
<i>autonomous system</i>	a system whose <i>dynamical law</i> ( $x' = L(x(t), u)$ ) depends only on the values of the <i>state variables</i> and the values of some set of <i>fixed parameters, u</i>
<i>bifurcation</i>	a qualitative change (e.g., the appearance or disappearance of <i>attractors</i> and / or <i>repellors</i> ) in the system's <i>phase portrait</i> , caused by a change in its <i>parameters</i> (either <i>fixed</i> or <i>changing</i> )
<i>nonautonomous system</i>	a system whose <i>dynamical law</i> ( $x' = L(x(t), u(t))$ ) depends both on the values of the <i>state variables</i> and the values of some set of <i>changing parameters u(t)</i> .
<i>coupled system</i>	an <i>autonomous system</i> consisting of two mutually interacting <i>nonautonomous systems</i> , whose <i>parameters</i> function as some of the <i>state variables</i> of the other, and vice versa

Table 1