Radical embodied cognition vs. «classical» embodied neuroscience

A. Reboul, L2C2 FRE3406

1. Introduction: the birth and development of cognitive science

Psychology, as most other humanities, took its independence from philosophy in the second part of the XXth century, under the impulsion of scholars such as William James in the USA and Wundt in Germany. In those early years, psychology was largely based on introspection and this led, at the beginning of the XXth century, to the behaviorist revolution. Behaviorism, in reaction against introspectionnism, tried to give psychology a scientific respectability by severely restraining itself, both in its explanations and in its chosen explananda, to observables, i.e. behaviors on the one hand and features of the environment on the other. This led to the stimulus-response unit, which was supposed to account for all behavior, from the most simple reflexes (e.g., the eyeblink) to the most complex human behaviors (e.g., linguistic communication).

Cognitive sciences were born as a revolution against behaviorism in the 50s, one major landmark being Chomsky's (1959) review of Skinner's (1957) book, *Verbal Behavior*. Chomsky pointed out that, contrary to Skinner's claims, verbal behavior (the production and interpretation of linguistic utterances) was too complex to be explained by stimulus-response mechanisms, even if these were chained in sequences. One needed hierarchical structures, and hence internal (or mental) rules and representations. Lashley generalized the argument to all complex cognitive behaviors. The time was favorable to a return of representations as the center of mental processes, as the first computers were put into use, leading to the birth of artificial intelligence. Early cognitive science (otherwise known as *«classical» cognitive science*) was thus strongly linked to so-called GOFAI (Good Old-Fashioned Artificial Intelligence).

Its strong link with GOFAI, which was the empirical side of cognitive science on a par with experimental psychology, directed classical cognitive science to *computationalism* (the idea that cognition was basically computational), *representationalism* (the idea that cognitive computations applied to representations), and functionalism (the idea that a given cognitive mechanism could be described in terms of its *function*, i.e., in terms of its inputs and outputs). Functionalism was *materialist* as far as cognition was concerned, that is, it considered cognitive processes to be material processes. However, one of its main

claim, the hypothesis that any cognitive mechanism could be realized in multiple ways and in different material substances, from computers to brains, and in different ways in different brains, meant that classical cognitive science was fairly disincarnated, and, indeed, was not very interested in the actual brain structures which subtended cognitive processes.

The next major step was due to the technological advances which led to the rapid development of sophisticated tools for brain imagery. In tandem with the relative failure of artificial intelligence and with the fairly moderate success of new departures in Artificial Intelligence (i.e., Distributed Parallel Processing, see Rumelhart et al. 1986), this put neurosciences as the empirical basis for cognitive science and indeed reoriented the field toward cognitive neuroscience. The rise of neuroscience meant that cognitive science, without entirely rejecting functionalism (for instance the computational level of description of cognitive processes described by Marr (1983) can be read in functionalist terms), became strongly incarnated in the brain. However, this was only the first step, and neuroscientists were responsible for a first extension of cognition beyond the brain, i.e., *embodied cognition*: researchers on emotion (see, e.g., Damasio 1994), perception (see, e.g., O'Regan & Noe 2001), or action (see, e.g., Jeannerod 1994, Rizzolatti & Sinigaglia 2007) have concluded that cognition went beyond the brain and implicated the whole body. The extension of cognition did not stop at the body boundaries, however, and philosophers (see, e.g., Clark & Chalmers 1998, Wilson 2004) proposed to extend the notion of cognition to encompass tools (such as notebooks and pens) as well as to other minds.

All of these changes have led to a new landscape of cognitive science, with classical cognitive science having lost a lot of ground on the one hand, and the dominant paradigm of cognitive neuroscience on the other. Cognitive neuroscience itself, however, can be divided, depending on what it encompasses, that is, on whether it is limited to the brain (so-called *«classical» cognitive neuroscience*), goes beyond the brain to encompass the whole body (*embodied cognition*), or beyond the body to features of the social or physical environment (*extended cognition*). Nevertheless, all these changes did not contradict the major characteristics of cognitive science since its inception, to wit, its *representationalism*.

This next major step was taken by *Radical Embodied Cognitive Science*, the eliminativist program defended by Chemero (2009). Chemero bases his claim on a definition of *representation* that crucially hinges on the notion of decoupling on the one hand and on a very selective choice of cognitive processes on the other. Given his definition of

representation, he shows, on the basis of his examples, that the notion of representation is not necessary to account for cognitive processes.

In the present paper, I shall try to show that Chemero's conclusion is unwarranted because it is based on a very restricted and oriented choice of cognitive processes. I shall examine further examples of cognitive processes, which have generally been seen as prototypical examples of «complex» cognitive processes, necessarily based on representation: the collaborative hunting found among Taï chimpanzees, and production and processing of fiction and thought experiments. I shall argue that, though the first does not in fact necessitate a representational account, the second does. I shall conclude to the validity of embodied and situated cognition, but to the invalidity of Radical Embodied Cognitive Science and, more specifically, of Chemero's eliminativism.

2. The Radical Embodied Cognition debate

Chemero describes «classical» cognitive science as *mental gymnastics*. He notes that, though embodied cognitive science is not anti-representationalist, its reliance on the environment dramatically reduces mental gymnastics, because the representations involved are action-oriented. This leads Chemero to a strong conclusion, a conclusion that is pivotal for his whole eliminativist argument: «Embodied cognition is necessarily situated» (Chemero 2009, 25). Let us call *situated* behavior any behavior that is *triggered and organized online by the environment*. In other words, it is *environment-dependent*: it does not happen in the absence of the relevant feature in the environment. By contrast, *non-situated* behavior is any behavior that is *not* triggered and organized *online* by the environment. In other words, it is *environment-independent*: it can happen in whatever environment the agent finds itself in.

Radical Embodied Cognition rests on three claims:

- representational and computational views of embodied cognition are mistaken;
- embodied cognition relies on a particular set of tools *T*, which includes dynamical systems theory;
- the tools in *T* do not posit mental representations.

This leads Chemero to a programmatic declaration for eliminativist Radical Embodied Cognitive Science:

> «I hereby define radical embodied cognitive science as the scientific study of perception, cognition, and action as necessarily embodied phenomenon, using explanatory tools that

> do not posit mental representations. It is cognitive science without mental gymnastics» (Chemero 2009, 29).

However, minimizing «mental gymnastics» makes it more difficult to account for complex, intelligent behavior (a difficulty not all that different from that which capsized the behavioral boat). Chemero proposes an answer to that problem: «in embodied cognitive science, some of the intelligence is «off-loaded» from the brain to the body and environment» (Chemero 2009, 27). This, basically, is the view propounded by extended cognition and it links quite well with the Gibsonian notion of *affordances* (see Gibson 1986), unsurprisingly, given that Gibsonian psychology is given a central role in Chemero's proposal.

One classical objection against extended cognition (see, e.g., Adams & Aizawa 2010), which could touch radical embodied cognitive science as well, given Chemero's position, is that the «cognitive system» formed by the agent and the environment (a tool or another individual, for instance) can be «decoupled», or, in other words, separated. According to Chemero, however, radical embodied cognitive science is a variety of extended cognitive science, where agent and environment are modelled as non-linearly coupled dynamical systems. The non-linearity of the coupling means that the agent and the environment form a unified, non-decomposable system. Thus, this specific objection is laid to rest.

Defenders of representationalist embodied cognition can propose either or both of the two following objections against Chemero's view:

- first, they might claim that it is impossible to explain *truly* cognitive phemomena without mental gymnastics;
- second, they can say that the models and theories used in radical embodied cognitive science actually do attribute representations to cognitive systems.

I will follow a modified version of the first path. I do not intend to claim that «truly» cognitive processes are necessarily representational, because I very much doubt that this can be argued in a non-question-begging way. Rather, I intend to claim (and show) that *some* cognitive processes are representational. In order to do so, I shall quickly outline the account Chemero gives of the notion of *representation*, which is clearly central to any discussion of his argument. Chemero discusses three possible relations between representations and their targets:

- a representation *R* and its target *T* are in *constant causal contact* just in case whenever *R* is present in a system, *T* is causing it;

- a representation *R* is *decouplable* just in case it can at least sometimes perform its function in a system when it is not in causal contact with its target *T*;
- a target *T* is *absent* just in case *T* has no local causal effects when a representation *R* of it is present in a system.

Quite reasonably, Chemero chooses the middle option, that the criterion of a representation is decouplability. It is, as he points out, neither too wide (as would be the first option, constant causal contact) or too strong (as would be the absence of the target). It also ties well with Chemero's argument, as we shall see, given his claim that all embodied cognition is situated.

Chemero himself points out that the anti-representationalism of radical embodied cognitive science can be interpreted in two ways:

- as a *metaphysical* claim: the nature of cognitive systems is non-representational (they do not involve representations);
- as an *epistemological* claim: our best explanations of cognitive systems will not involve representations.

Chemero chooses to argue for the epistemological interpretation of anti-representationalism, on the understanding that the truth of the epistemological claim may be evidence for the truth of the metaphysical claim, on the basis of a few examples of cognitive processes, falling either in the domains of perception (tracking objects), actions (grasping or exploring an object), or keeping rhythm. Chemero proposes dynamical systems theory accounts for each of these cognitive processes. Given that dynamical systems theory does not involve representations, explanations of cognitive processes do not involve representations. Chemero concludes that anti-representationalism is vindicated.

To contest Chemero's eliminativism, three possibilities come to mind:

- contesting that the dynamic systems accounts he proposes are indeed the *best explanations* one can come up with;
- contesting his definition of a representation in terms of decouplability and arguing in favor of one of the other alternative definitions – in terms of constant causal contact or of absent target – Chemero proposes (or for yet another alternative definition of a representation);
- contesting Chemero's choice of cognitive processes.

I will follow the third path and contest Chemero's choice of cognitive processes.

My approach will be based on what I take to be the obvious reading of Chemero's claim, that it is a *universal* claim. Anti-representationalism as a universal claim can be formulated as follows in both of its readings:

- *metaphysical* claim: *No* cognitive system involves *any* representation;
- *epistemological* claim: *None* of our best explanations of *any* cognitive system involves representations.

Concentrating, as Chemero does, on the epistemological claim, what this means is that showing that even one cognitive process has a best explanation that involves representations is enough to falsify Chemero's anti-representationalist claim.

3. An empirical approach

3.1. Introduction

Let me begin by summing up Chemero's argument, and presenting it as a logical deduction:

1. All cognition is embodied.
2. All embodied cognition is situated.
3. All situated cognitive processes have best explanations in terms of dynamical systems theory.
4. Dynamical systems theory does not involve representations.

From these four premises, Chemero draws the conclusion:

5. No best explanation of any cognitive process involves representations… hence eliminativism.

The premise I want to contest is the second one, according to which all embodied cognition is situated: the central notion here is not embodiment, but situatedness, and, given its importance, I want to remind the reader of the distinction between situated and non-situated cognition:

- situated behaviors and the cognitive processes that support them are *triggered and organized online by the environment* — in other words, they are *environment-dependent*;

- non-situated behaviors and the cognitive processes that support them are *not* triggered and organized *online* by the environment — in other words, they are *environment-independent*.

Given this reminder, I shall now examine two claims of Chemero's — there are *no* non-situated cognitive processes; *none* of our best explanations of *any* cognitive systems involve representations — through the analysis of two examples: collective hunting in Taï chimpanzees, and the human use of thought experiments and fiction.

3.2. Collective hunting in Taï chimpanzees

It is well-known that chimpanzees are the only apes to be omnivorous (as are humans, of course), and are in fact hunters, mostly of Colobus monkeys. Goodall (1986) first investigated the hunting behavior of chimpanzees in Gombe. In the past twenty years, Boesch (1994a, 1994b, 2002, 2009, Boesch and Boesch 1989) has completed her account by a series of descriptions of the rather different hunting behavior of chimpanzees in the Taï forest. Gombe chimpanzees live in a savannah environment, where there is a low underbrush and big trees are often separated from one another, allowing chimpanzees to trap Colobus monkeys on a given isolated tree. This leads Gombe chimpanzees to solitary short opportunistic hunts of smaller preys (young and infants), with no sharing of the meat afterwards. By contrast, Taï chimpanzees live in a tropical rainforest with very dense vegetation, where trees form a canopy, allowing preys to pass from one tree to the next. This more difficult environment has led Taï chimpanzees to collaborative long deliberate hunts, targeting bigger preys (adults) and ending in meat sharing. What is more chimpanzees not only hunt *collectively*, they actually *collaborate*, by adopting different roles during the hunt, and changing their roles depending on what is going on. Boesch first distinguishes *hunters* (actually participating in the hunt) and *onlookers* (who follow the hunt without participating in it). Hunters (male chimpanzees only) can play the following roles:

- the *driver*, who follows the prey without catching it;
- the *blocker*, who places himself under a tree to block the prey's progression;
- the *chaser*, who follows the prey quickly and tries to catch it;
- the *ambusher*, who goes to a location where the prey has not yet arrived, and where he can't be seen and he jumps on the prey as soon as it arrives.

This seems to argue for fairly complex cognitive processes, of the sort that are typically considered to be subtended by representations. An additional complexity is to be found in

meat sharing. Individual effort is not directly rewarded, but a specific participation, leading to the success of the hunt, is required for the individual to have a better access to meat. These socially enforced rules illustrate the complexity of the collaborative action, and, again, have been taken to argue for representational cognitive processes. As Chemero rightly points out, however, a complex behavior (such as that manifested by the Taï chimpanzees in their collaborative hunts) does not necessarily presuppose cognitive processes of representational nature. Indeed, Chemero is perfectly right to point out that this is one of the important contributions of dynamical systems theory: *apparently complex behavior can be produced by a few simple rules taking as input features of the local (or proximal) environment*. This, it should be noted, is in keeping with Chemero's contention that, in radical embodied cognitive science, much of the cognitive complexity is off-loaded on the body or the environment. In other words, with the notion that the apparently complex behavior (as well as the cognitive processes subtending it) is situated. *So the right question is not whether the behavior is complex, but whether it is situated*.

To answer it, let us go back to the difference in hunting behavior in different populations of chimpanzees. The four well-studied wild populations of chimpanzees, at the present date, are located in Ngogo, Taï, Mahale, and Gombe. All of them, except for the Taï chimpanzees, live in environments similar to Gombe, making catching Colobus monkeys fairly easy. Thus, though simultaneous hunts, where male chimpanzees all hunt at the same time, each on his own and in his own interest, are the rule in Ngogo and Mahale, and are the majority (up to 85% of all hunts in Gombe), collaborative hunts, agreeing with the description given above, are the majority in Taï, making up to 75% of all hunts (see Boesch 2009, figure 6.1, 122). Thus, there is no doubt that collaborative hunting is forced on Taï chimpanzees by their environment. This, however, does not necessarily mean that collaborative hunting, *as a behavior subtended by cognitive processes*, is situated. Basically, this very much depends on the extent of the information to which each chimpanzee participating in the hunt has access:

- If this information is *wide*, including the present position of each hunter, the present position of the prey, and the location of trees likely to facilitate the capture, this would suggest a partially non-situated process, for instance a process using an *allocentric* mental map (geometrically rather than action-based: on the distinction between allocentric vs. egocentric maps, see Jacob & Jeannerod 2003);

- If, however, this information is restricted to the immediate environment of the hunter, this rather hints at a situated process, in which the hunter will act based on changes in his immediate environment (e.g., the arrival of the prey).

Boesch describes the Taï forest as so dense that visibility is highly restricted. In other words, it is a safe bet that each hunter has only restricted access to what happens in the hunt at a given time. If this is so, it also seems that the collaborative hunt, as far a each participant is concerned, is situated. This raises the question of where the complexity apparent in the collaborative hunt comes from. To answer this question, I want to have a quick look at a parallel example. In the animal world, massive migrations of groups of several thousands of animals (herds of wildebeests and more generally herbivores on the African plains, shoals of fishes, groups of migrating birds, humans in big cities, etc.) are fairly frequent and raise questions as to how the individuals in such enormous groups coordinate their movements so as not to collide. Couzin and Krause (2004) have studied the collective moves of different types of animals and have showed that this apparently complex and organized behavior is the result of the application of a few simple rules by each individual in the group on the basis of features of its proximal environment. This, I think, gives a hint of the possible non-representational types of accounts that might be given of the collaborative hunting of Taï chimpanzees. One could, of course, object that the behavior of each hunter in Taï is more complex. The question, however, is whether it is qualitatively different, or merely quantitatively different. It seems to see that the number of variables in the proximal environment of the agent to be taken into account is greater in the Taï hunters than in group movements, but that is merely quantitative. Additionally, learning can offset the complexity due to the greater number of variables. And, indeed, Boesch (2002) describes the learning progress of hunting roles as slow: it begins when young males are 9 to 10 years old (adolescent by chimpanzee standards) and goes on for about twenty years. Boesch (2002, 38) compares this with what happens in Amerindian tribes of hunter-gatherers: "Young men begin hunting at around 15 years and their meat production peaks at around 35 year. This suggests a similar learning period of twenty years."

thus, in sum, Taï collaborative hunts have been taken to involve high-level (and clearly representational) cognitive processes, e.g., mindreading (see Boesch 1994a, 1994b, 2002, 2009, for such arguments), it is not clear that this is the case. Indeed, there are good readons to think that the behavior is strongly situated, as are the cognitive processes

that underlie it. And Chemero is certainly right to claim that there is no reason to think that situated processes involve representations in his, decouplable, sense.

3.3. Fiction and thought experiments

Let me begin by pointing out that, given that non-situatedness implies that the behavior concerned is neither triggered and organized online by features in the environment, an obvious possibility is that it is triggered and controlled online by decoupled representations, in other words, by representations that can perform their functions when they are not in causal contact with their targets, i.e., independently of the environment (i.e., representations in Chemero's sense). There are quite a few possible examples of uses of decoupled representations, from hypothetical reasoning to heterogeneous attention and long-term planning for collaborative action. I will concentrate on two clear examples of decoupled representations (where "clear" means that it is hard to see how they could be anything but decoupled): counterfactuals, as used in scientific, historical, and philosophical works; fiction, as in literature, movies, comics, TV soaps, etc.

Counterfactuals have been used in history, mostly to establish causal links between events (in the large sense, including not only individual events, but the emergence, disappearance or change of ideas, cultures, institutions). The use of counterfactuals in history has been the subject of philosophical inquiry as to its epistemological validity, but here I shall be concerned with how much such counterfactuals depart from reality. Mostly the use of counterfactuals in history is based on the idea that one can bring to light the causal chain that led to a given historical event and discriminate the relative importance of its links, by ratiocinating about what would have happened if one or another of these links had not occurred. This is where the departure from reality can be assessed: in how much reality has to be altered in the production of the antecedent of the counterfactual.

As my example, I shall take McNeil's (1999) paper on "the plague that saved Jerusalem". In this paper, McNeil deals with an event that took place in 701 BC: Assyrians led siege to Jerusalem, capital of the Kingdom of Juda. That siege was lifted by Sennacherib, King of Assyria, because his army succumbed to a mysterious, but lethal, contagion. According to McNeil, this allowed the religion of the Kingdom of Juda to develop into a monotheism, Judaism, which was the direct ancestor of the two other great monotheisms, Christianity and Islam. If Sennacherib had not lifted the siege, there is no doubt that the superior Assyrian forces would have taken the city and inflicted to it the treatment that the neighboring Kingdom of Israel received twenty one years before in 722 BC: its population

was exiled, and abandoned the cult of Jaweh (who had failed to protect it) and dissolved in the local population. But another factor was the "providential" contagion that struck the Assyrian army, and that was widely interpreted as a divine intervention.

How far the counterfactual antecedent differs from what actually happened will depend on how far back in time one has to go for the alteration. From McNeil's part of view, the two different factors, i.e., the contagion and the lifting of the Assyrian siege, are more or less simultaneous. Indeed, one could argue that the main factor was the contagion in as much as it both helped the development of monotheism by being interpreted as the results of a divine intervention; and it was so interpreted because it "providentially" led to the lifting of the siege. In other words, in the case of McNeil's argument, one does not have to go very far back in time. On a relative view of situatedness, McNeil's counterfactual argument going back only very close in time could be seen as situated. On an absolute view of situatedness, McNeil's counterfactual argument is non-situated. For what it is worth, I cannot see how a relative notion of situatedness could get off ground. However, the other case I want to examine, fiction, is undoubtedly non-situated.