

Running head: EMBODIED EMBEDDED COGNITIVE NEUROSCIENCE

Can There be Such a Thing as Embodied Embedded Cognitive Neuroscience?

Jelle van Dijk,

Human Technology,

The Hague University, The Hague

Roel Kerkhofs,

Statistics Netherlands,

Voorburg

Iris van Rooij,

Human-Technology Interaction,

Eindhoven University of Technology

and Pim Haselager

Artificial Intelligence/Cognitive Science,

NICI, Radboud University, Nijmegen

**Correspondence information:**

Pim Haselager

Artificial Intelligence/Cognitive Science,

NICI, Radboud University Nijmegen,

Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

Tel: +31 (0)24 3616148

Fax: +31 (0)24 3616066

Email: w.haselager@nici.ru.nl

### Abstract

Contemporary cognitive neuroscience, for the most part, takes its aim to be figuring out how cognitive processes are realized *in* the brain. This research goal betrays the field's commitment to the philosophical position that cognizing is something that the *brain* does. Since the 90s, philosophers and cognitive scientists have started to question this position, arguing that the brain constitutes only one of several contributing factors to cognition, the other factors being the body and the world. This latter position we refer to as *Embodied Embedded Cognition* (EEC).

Although cognitive neuroscience's research practice and EEC do not seem to fit well together at present, it is pertinent to ask if a variant of cognitive neuroscience can be developed that sets itself research goals that are more congenial to the EEC view. In this paper we investigate this possibility. We put forth a new guiding metaphor of the role of the brain in cognitive behavior to replace the current cognitivist metaphor of the brain as an information-processing device. We also identify a research agenda that naturally arises from our metaphor. In this way we hope to provide an impetus for cognitive neuroscientists to pursue an EEC-inspired research program.

**Keywords:** cognitive neuroscience, embodied cognition, embedded cognition, brain-behavior relationship, cognitivism, modularity.

## 1. Introduction

“*Where are the words?*” asked Friedemann Pulvermüller, at a meeting with fellow psycholinguists and cognitive neuroscientists. With this he meant to ask where in the brain the words we use are stored (see Pulvermüller, 1999, for an answer). Pulvermüller’s research goal is characteristic for, and symptomatic of, the internalist and localist intuitions underlying contemporary cognitive neuroscience. It is common practice to aim at mapping particular cognitive functions to neural structures; a practice that is motivated by the assumption that cognition is realized by functionally identifiable structures that are localized in the brain. In this paper we question this assumption; specifically, we question the very assumption that cognizing is something that the *brain* does. We furthermore discuss the possible research goals of a cognitive neuroscience that rejects this assumption.

To be clear, we do subscribe to a materialist view of cognition, but we believe there is more to cognition than the processes inside the brain alone. Besides the brain, also the organism’s body and the world form part of the physical substrate that underlies behavior and cognition. In this respect, our position is exemplary of the theoretical framework that we call *Embodied Embedded Cognition*, or EEC for short (Brooks, 1999, Clark, 1997, Chiel & Beer, 1997). EEC proposes that cognition and behavior emerge from the bodily interaction of an organism with its environment. According to EEC, the physical structure of body and world, and the internal milieu of the organism’s body, all provide important constraints that govern behavioral interactions. From this perspective cognitive states are best explained by a physical system of interacting components, where the brain is only one such component. In other words, the brain is best viewed not as a commander or director of behavior, but rather as only one of the players among equally important others (i.e., the body and the world). As a result, according to

EEC, in a great number of cases, higher cognitive functions cannot be directly mapped onto brain structures. This observation raises two important questions, one conceptual and one methodological: (1) If the goal of an EEC neuroscience is not exclusively to map cognitive functions to brain structures, then what are its (other) explanatory goals? (2) By what methods can the research goals of an EEC neuroscience be realized? Since an answer to (2) presupposes an answer to (1), we first set out to provide a possible answer to the conceptual question. Drawing on existing research in the fields of robotics, biology, and neuroscience, we put forth a perspective of the role of the brain in the formation and coordination of cognitive behavior, and we identify the kinds of research questions that present themselves on this perspective.

Although we leave a proposal for a concrete answer to (2) for another occasion, our analysis does make clear that existing methods for cognitive neuroscientific research will unlikely suffice for implementing a full-blown EEC neuroscience research program. Hence, our answer to (1) establishes a need to the development of new tools and methodologies.

Many scientists and philosophers have tried to make the case for the usefulness, validity or even necessity of an alternative to the internalist paradigm for cognitive science (see a.o. Varela, Thompson & Rosch, 1991; Edelman, 1992; Thelen & Smith, 1994; Bickhard & Terveen, 1995; Kelso, 1995; Hutchins, 1995; Port & van Gelder, 1995; Hendriks-Jansen, 1996; Clark, 1997; Clancey 1997; Agre, 1997; Juarrero, 1999; Pfeifer & Scheier, 1999, to name but a few). Let us note that participating in this debate is explicitly *not* the main focus of our paper. We believe the case for EEC has already been made by other researchers well enough for it to be considered at least a viable alternative to cognitivism. Given that pluralism in approaches makes for healthy science, it seems sensible to try to pursue a EEC neuroscience, not instead of cognitivist neuroscience, but in addition to it. Besides, even though cognitivism has its successes,

it is also known to have major problems, in particular with accounting for common sense behavior in environments of real-world complexity (Pylyshyn, 1987; Haselager 1997). EEC may be able to overcome the traditional problems plaguing cognitivism. Because EEC and cognitivism are not rivaling theories, but differing philosophical paradigms, the divergences are too general and fundamental for a small set of experiments to directly test between them. The relative success of EEC as a framework for the study of cognition can only be assessed in the long run, once a consistent EEC neuroscience research program has been established. At this moment in time, it seems prudent to first try and formulate at least a rough indication of the explanatory goals of such an EEC neuroscience research program. Therefore, in the remainder of this paper we will assume an EEC perspective and then set out to investigate what this might imply for our understanding and study of the main tasks the brain is performing. Thus, we aim to tentatively sketch the outlines of an empirical EEC-inspired cognitive neuroscience.

The paper is organized as follows. We start, in Section 2, by examining the standard explanatory strategy of contemporary cognitive neuroscience, and identifying its philosophical commitments to cognitivism and modularity of mind. In Section 3, we consider neuroscientific data that can be seen as supporting the assumption that cognition is fundamentally embodied in nature. Although this research provides empirical evidence for the central claim of EEC, it still adopts much of the constraints and conceptual assumptions characteristic of traditional cognitive neuroscience. In sharp contrast, Sections 4 and 5, lay out an embodied embedded conception of the brain and its contribution to human cognition. Drawing on work in robotics, biology, and neuroscience we propose a conceptualization (a metaphor) of the relationship between behavior, body and brain activity in real-world contexts. In Section 6 we explicate the kinds of research questions that our conceptualization raises.

## 2. Cognitivism and modularity in cognitive neuroscience

Traditionally, cognitive scientists have conceived of the mind/brain as an organ for building internal representations of the external world. By performing internal computations over these representations the mind/brain comes to a decision about what actions are appropriate. This philosophical position is still dominant in contemporary cognitive science and is referred to as *cognitivism*. Cognitive neuroscience has naturally inherited the cognitivist conception of the brain as a programmer for, and decider of, overt behavior. An apparent advantage of this conception is that it affords the study of the brain's functioning in isolation. If the brain implements an encapsulated mechanism for cognizing (perceiving, planning and deciding) then it makes sense to try and understand this mechanism by studying the brain independent from embodied embedded interaction with the world. After all, from this perspective the body and world are merely external factors (related to the input and output) to cognition. Cognizing is something that the brain does.

So how may we then understand the mechanism by which the brain cognizes? The answer to this question draws on yet another traditional conception of the mind/brain as essentially a modular system. According to this view, also known as the *modularity of mind* thesis (Coltheart, 1999, Fodor, 1983, 2000; Pinker, 1997), the brain's mechanism for controlling action is decomposable into functional units, called modules. Each such module is believed to be responsible for computing an independent cognitive function conceived of as an information-processing task. In practice, any given task will likely draw on many different modules at the same time, so modules typically work in concert and potentially even in parallel. But this does nothing to weaken the assumption that the inner workings of these modules are encapsulated and largely unaffected by the workings of other modules. The only way in which modules can

influence each other is by exchanging representations of the results of their processing. Note how the modularity thesis is a natural extension of cognitivism and mirrors the cognitivist assumption that brain processes are encapsulated and disconnected from concurrent bodily and environmental processes. Given the modularity thesis, understanding how the brain cognizes means figuring out the basic set of dedicated functional modules (and their interconnections) that make up the brain. That is why so many psychological theories are presented in the form of flow-charts, accompanied by the attempt to localize the individual components in the brain.

In line with the modularity thesis, current cognitive neuroscientific methodologies are geared towards isolating the workings of independent modules. Consider, for example, *double dissociation* methodology (e.g., Coltheart, 2002): A person, known to have damage to brain area *X*, performs normally on task *A* but is impaired on task *B*; another person, known to have damage to brain area *Y*, performs normally in task *B* but is impaired on task *A*. The traditional inference made from this pattern of results is that brain areas *X* and *Y* implement different functional modules, with the functional module implemented in *X* being responsible for the performance of task *B* and the functional module implemented in *Y* being responsible for the performance of task *A*. The well-known *subtraction method* employed in fMRI research (e.g., Hernandez-Carcía, Wager, & Jonides, 2002) displays a similar form of logic: A person performs two tasks *A* and *B*, where it is assumed that task *A* is a subtask of *B*. Activation levels (in terms of blood flow) observed for task *A* are subtracted from the observed activation levels for task *B*. Brain areas that show a significant *difference* in activation for the two tasks are then inferred to implement the modules necessary for performing task *B* but not task *A*.

We are aware that the standard cognitive neuroscientific methodologies are subject to criticism regards the (potentially problematic) methodological assumptions that underlie them

(e.g., Uttal, 2001; Hardcastle & Stewart, 2002; Dunn & Kirsner, 2003; Van Orden, Pennington, & Stone, 2001). Our concern here, however, is not so much with the *methodological* assumptions involved, but more with the *conception* of the brain that motivates the use of them as the way to study the relationship between brain and cognitive behavior. Clearly, the motivation to use such methodologies makes sense only when one subscribes to a cognitivist and modular view of the brain, whose modules can in principle be studied in isolation.

In the next section (Section 3) we will present neuropsychological evidence that higher-order (semantic) representations are intricately connected to the neural substrate underlying basic low-level sensory and motor processes.

### **3. The body is reflected in how the brain represents meaning**

Cognitivism conceives of sensory and motor systems as peripheral transducers of proximal stimulation into internal abstract codes. These codes serve as the input to more central cognitive modules. The representations that serve as inputs to and outputs of central systems are abstract representations that are potentially amodal in nature.

A prime example of a cognitive ability that is widely believed to be performed by central cognitive modules is the representation of the meaning of words. From a cognitivist perspective, one would expect the neural substrate underlying the representation of word meanings to be distinct from the neural substrate underlying sensation and motor control. In contrast, we will present two studies that show the interconnectedness of basic sensory and motor experiences on the one hand and language comprehension on the other.

Wiemer-Hastings and Kurby (in prep.) have shown in a series of experiments that the visual cortex plays a central role in the representation of content words. In their first study they selectively “turned off” cells in the visual cortex while they were measuring response times to

words. The cells in the visual cortex were selectively turned off by means of habituation. In this procedure participants have to look at a computer screen on which either horizontal or vertical bars were depicted. After looking at these grids for some time, the cells in the visual cortex that have selectively vertically aligned receptive fields become insensitive and stop responding to input with this alignment. When the participants were habituated to a particular direction they were presented a word on the screen, e.g., a “vertical” word like TOWER for the vertical habituation, or a “horizontal” word like TRAIN for the horizontal habituation. They had to push a button as soon as they had recognized the word. The results showed an interaction between habituation type and target word, showing that participants had longer recognition times for words like TOWER when they were habituated on a vertical pattern than when they were habituated on a horizontal pattern. Likewise, recognition times for words like TRAIN were longer for the horizontal habituation than for the vertical habituation. The striking fact about these results is that the cells with the specific receptive fields are very low-level feature detectors. These results suggest that even these low-level cell assemblies contribute to the meaning of concepts.

A second important study is the study by Pulvermüller, Härle, and Hummel (2001). This study also focuses on the idea that representations, in addition to the classic “language areas” over the temporal lobe of the cortex, consist of brain areas associated with primary processing such as motor processing or perceptual processing. In this study they studied the processing of action verbs that are related to different extremities (e.g., SMILE, which is head-related, KICK, which is leg-related, and THROW, which is arm-related). Participants performed a lexical decision task, in which they read letter strings on a computer screen, and had to decide by pressing a button whether the word they were reading was an existing word or not. In addition to

the recording of reaction times the electroencephalogram (EEG) was measured. The hypothesis was that face-related words would be processed faster than arm-related words, which on their turn would be processed faster than leg-related words. Reason for this hypothesis is the somatotopic organization of the motor cortex. The motor cortex is so organized that leg functions are localized more to the top of the brain (closer to the vertex), face functions are more localized towards the temporal lobe, and arm functions somewhere in between. As a result of this, leg-related words have wider cortical distributions than arm-related words, which have wider cortical distributions than face-related words. Wider cortical distributions and more narrow distributions predict a difference in processing times because longer cortico-cortical connections imply longer traveling times for action potentials when the network ignites (Pulvermüller et al., 2001).

The reaction time results showed that indeed leg-related words are responded to slower than face- and arm-related words, although the difference between face- and arm-related words was not significant (1). In addition, the analyses of the EEG data showed that the processing of the action verbs produced differential activation along the motor cortex. Leg-related verbs showed more activation close to the vertex, face-related verbs showed more activation near the temporal lobe, close to the representation of the articulators. This study suggests that the motor cortex plays a vital role in the representation of action verbs. The very areas that are used by the brain to execute movements are involved in the representation of words. Therefore it can be claimed that an important part of the meaning of these verbs is *embodied* in the sense that these verb-representations involve neural structures associated with low level sensorimotor processes.

The results of these studies appear to be counterintuitive from a cognitivist perspective of cognitivism. As such, these studies can fuel the motivation to reconsider the distinction between central and peripheral systems and the strict informational encapsulation of modules. The studies

also demonstrate that it is possible to collect data congenial to an embodied embedded cognition framework with a traditional neuroscience toolkit. However, we believe that these studies are also somewhat limited, for two reasons.

First of all, the results of these studies are not, strictly speaking, contradicting the cognitivist conception of the brain. The subfield of linguistics — tellingly known as *cognitive linguistics* (e.g., Lackoff, 1987; Langacker, 1987, 1990) — that fuels the type of research reviewed here emphasizes studying the nature of internal representations and their physical realization in human brains (see also Edelman, in press; Veale & O'Donoghue, 2000; Veale, O'Donoghue, & Keane, 1999). Cognitive linguistics does not conceive of internal representations as abstract and amodal entities, but rather as concrete conceptualizations achieved via imagination or internal simulation of bodily experiences and behaviors. Nonetheless, the approach is cognitivist in that it views cognition as first and foremost a form of internal information-processing. In other words, the view seems to be a blend of both cognitivist and EECist intuitions, which is why we will refer to it as “weak” EEC. In contrast, “strong” EEC rejects the very idea that the brain's primary purpose is to represent the world and instead views the brain as first and foremost a regulator of embodied embedded behavior.

A second limitation is that methodologies such as adopted in weak EEC neuroscience research will unlikely suffice to implement a full-blown, strong EEC neuroscience. A strong EEC neuroscience, after all, would not so much be interested in associations between (lower-level) sensory-motor processes and (higher-level) cognition per se, but rather it would want to understand *how* the brain contributes to the creation and maintenance of *adaptive real-world behavior* in concert with contributions from the body and the environment. This requires a more

fundamentally alternative view on the relation between brain and behavior. To this issue we will now turn.

#### 4. Behavior before the brain

To consider the contribution of the brain to the emergence of behavior, it may be useful to examine an evolutionary old example. Let us take as an extreme case (in terms of simplicity) the *E. coli*, a single-cell organism without nucleus (prokaryote). This creature is capable of moving its flagella in two ways such that its behavior can change from tumbling about randomly to swimming in one direction. Without specific stimulation it changes between these two modes every few seconds, thereby remaining more or less in place. Like all biological creatures, the *E. coli* needs to maintain itself. It therefore has to have the capacity to distinguish things both in the external world (food vs. non-food, safe vs. unsafe environments) and in its internal milieu (e.g. chemical imbalance or ‘hunger’ vs. homeostatic equilibrium or ‘satiation’). The *E. coli* has these capacities. Once a chemical gradient (e.g. an increase in sugar level or a decrease in toxic substances) in its environment changes, it increases the amount of swimming and decreases the random tumbling. In effect, these behaviors make it swim upwards along a stream of increasing nourishment towards a food source, downwards along a stream of decreasing toxics, and to move about randomly in a non-stimulating or neutral environment (Cairns-Smith, 1996, p. 90--94).

The *E. coli* has no brain but performs its tasks in virtue of its molecular machinery. Evolutionary speaking, *behavior came before the brain*. When the complexity of a creature increases (its capacity to receive inputs, perform actions, as well as its internal structure), and the tasks it performs become more varied, some sort of integration mechanism becomes necessary. Parallel processes need to be combined, simultaneously occurring events that can be sensed need to be ranked in terms of priority of relevance, multiple motoric outputs need to be orchestrated in

order to achieve fluent movements, etc. This is where the (central) nervous system comes in: to facilitate behavior, not (at least not necessarily) to create it.

Neural control is essentially grown out of the need for maintaining proper *internal* value systems. The control of external behavior may be seen, to exaggerate it somewhat, as a ‘side effect’ to that main function. At the same time, controlling the inner economy successfully is impossible, for any organism, without adequate interaction with the environment. At the interface between the internal household and the need for external action lies the function of the brain.

The relation between bodily values and sensorimotor processes has been explicated by Damasio, (1994, 1999, 2003). In his theory, the prefrontal cortex plays a pivotal role. The prefrontal cortex receives input not only from other cortical areas (e.g., sensorimotor information) but also from subcortical emotional structures, a.o. the amygdala. The amygdala in turn receives input originating in the body’s internal milieu, in the form of homeostatic variance (e.g. body temperature, heart-rate, the metabolism and several hormonal fluctuations). According to Damasio, sensorimotor processing streams in the cortex are *laden* with positive or negative value by means of input from the prefrontal areas. The prefrontal cortex works like a filter, labeling sensory input with so-called ‘somatic markers’ (Damasio, 1994). The filtered information then becomes a constraining factor on the dynamics of the ongoing interaction with the environment. Because of this, interaction with the environment is fundamentally subjective (or ‘somatic’).

A bodily-based bias on sensorimotor processing is crucial for making quick decisions in a dynamic, complex environment. Prefrontal patients have difficulty in taking such ‘split-second’ decisions, based on what is aptly called our ‘gut-feeling’. Such patients are generally unable to

successfully organize their life and run into all kinds of organizational problems, especially in everyday business like planning daily activities. This is not a problem of intelligence.

Experiments show that patients can *reason about* common sense problems, if such problems are represented in the form of a puzzle or a game with strict rules. They are simply unable to effectuate their reasoning capacities in the real world (Damasio, 1994). In other words, these patients are particularly bad at everyday ‘common sense’ behavior.

In line with this is the embodied mind hypothesis of Varela, Thompson and Rosch (1991), drawing on the concept of autopoiesis. Autopoiesis emphasizes the idea that all biological systems are engaged in maintaining their own integrity, and that the components of a system contribute to the proper functioning of the whole on which their own continued existence depends. From this fundamental principle of life, cognition emerges. In that sense there is no fundamental difference between the *E. Coli* and *homo sapiens*. Given the specific ‘niche’ in which the organism finds itself (i.e. its embeddedness), the system will evolve an adequate compensatory response in the form of inner structure, in order to sustain itself. For the *E. Coli*, chemical taxis turned out to be a viable strategy achievable without a nervous system, for the survival of many other organisms a brain appears to be necessary.

### **5. The brain as a traffic facilitator**

We now present our alternative conception of the role of the brain. We begin with several insights from reactive robotics. (Brooks, 1999). Basically, a reactive creature consists of behavioral layers that each instantiate a direct input-output coupling. No intermediate (in between input and output) world modeling, planning and decision making takes place. Layers compete for dominance on the basis of the input received by the system. Figure 1 presents an illustration.

-----

Insert Figure 1 about here

-----

From this perspective a creature can be seen as a repertoire of behavioral dispositions. It is inclined, in virtue of its bodily possibilities and its history of interactions with its environment, to respond to stimuli in specific ways without high level thought or planning. Perception, action and world are structurally coupled to form a temporarily stable behavioral pattern that is functional with respect to the task. We call this structural coupling a ‘basic interaction cycle’. A creature carries its set of potential behaviors with it across contexts, and if these contexts ‘fit’ with the creatures’ behavioral repertoire, its overall conduct may be satisfactory for a long time.

The workings of a basic interaction cycle may be particularly relevant for common sense behavior, i.e. the normal things we do in daily life, such as making breakfast, going to work, shopping for dinner, etc. Common sense behavior actually consists in quite complicated sequences of behavior, even though it does not require the type of thinking characteristic of say playing a chess game. Instead we seem to operate more or less on ‘autopilot’. Behavior flows naturally out of our interactions with the environment (2). We propose that during autopilot behavior the environment selects appropriate behaviors from the behavioral repertoire, without any internally computed behavioral plans.

In the reactive robots of the early 90’s, behavioral layers and the precedence relations between them were set beforehand and were hardwired into the system (3). This results in creatures not unlike the *E. coli* discussed above. However, once the behavioral capacities of a creature become larger, and its sensori-motor capacities quite rich, a more flexible and integrated way of setting up behavioral layers and their interrelations becomes necessary. This, we suggest,

is the main task of the brain: Instead of interpreting the brain as the driver or pilot of the body, we see it as assisting the environment driven selection from the behavioral repertoire. It helps in facilitating the display of the *currently most relevant* behavioral dispositions and delaying or blocking others, by temporarily loading the dice, so to speak. These temporary applicable behavioral layers are then called upon by stimuli from the environment. (see Figure 2.)

-----  
 Insert Figure 2 about here  
 -----

A concept that may help to elucidate how the brain is involved in the temporary creation of a relevant behavioral repertoire is Edelman's (1992; Edelman & Tononi, 2000) notion of functional clusters. A functional cluster consists of "elements within a neural system that strongly interact among themselves but interact much less strongly with the rest of the system" for a certain amount of time (Edelman & Tononi, 2000, p.120, see also pp. 184--185). Several neuronal groups form a strongly integrated assembly for brief periods (most likely to be measured in the range of 50-100 milliseconds). In other words, functional clusters exist only temporarily, consist of various contributing areas that are recruited for the specific occasion and are changeable over contexts. Such clusters, in the robotics terminology discussed earlier, constitute parts of the behavioral layers (input-output links) and, at the same time, set precedence relations between them. This kind of extreme flexibility is not yet to be found in current reactive robots, because the robots' layers are fixed (hard-wired) instead of temporary.

Functionally speaking, the brain is helping the creature to readjust on the spot, becoming a more adequate behavioral system, by selecting a set of relevant behavioral dispositions (and their interrelations) out of its entire repertoire, from which the environment can then select.

Functional clusters are very different from cognitivist *modules*. In the way we use Edelman's concept of clusters here, functional clusters are momentary periods of integration that serve to form a compensatory force to an environmental change, in the context of the local situation. Functional clusters *presuppose* the existence of a lower level interaction cycle on which they put a bias. Without the basic interaction cycle, activity in functional clusters has no meaning. In contrast, cognitivist modules do what they do irrespective of their context. Their functionality is determined in terms of the input-output mappings they compute and these mappings are fixed. Their internal representations retain their meaning even if the module is disconnected from the world. On the neural level, cognitive neuroscience has sought to map cognitive modules onto specific anatomical structures. In our model, there is no fixed mapping between the observable patterns of behavior and the contributing functional clusters in the brain.

A global picture of the role of the brain is presented in Figure 3. Here we first of all see how activity of the perceptual system, the action system and the state of the world are structurally coupled in a basic interaction cycle. There is no strictly serial processing 'from perception to behavior'. Instead, the components form an interactive system whose local interactions result in global stabilities that are often sufficiently adaptive.

-----

Insert Figure 3 about here

-----

It can be seen how the basic interaction cycle might to some extent be sustained, or even enhanced, by the brain, as depicted by the large arrow. One could say that the brain helps to increase efficiency and maximize desired behavior by creating a *functional cluster* that

temporarily sets the appropriate behavioral repertoire from which the environment can select. The function of this neural bias is to help the basic interaction cycle to do things better, faster, and easier. It does not do so by *replacing* the basic interaction cycle, but by providing a useful *addition* to it, functioning as a temporarily ‘*traffic facilitator*’. This traffic facilitation is based on the brain’s monitoring of the body’s internal states (body maps) and the ensuing labeling of the saliency of external objects (e.g. by means of Damasio’s somatic markers)

In summary, goal- or homeostasis-directed interactions or behaviors between organism and world are primary. The brain is a secondary structure, though important, influencing rather than creating these interactions. The primary function of the brain is to intervene in the ongoing display of behavior in the light of the body’s needs and its history of world interaction. A similar view, in which several internal constraints influence, rather than prescribe, embodied and embedded action-selection processes, can be found in Cox and Smitsman (this issue). In sum, more causal credit is given to the basic interaction dynamics between the body (the sensors and actuators) and the world (Clark, 1997; Wheeler & Clark, 1999; Norman, 2002). Simply put, the brain is more of a servant and less of a director than traditionally thought.

Having outlined a strong EEC view of the brain that is at odds with the cognitivist and modular conception of the brain, the next step is to indicate the kinds of questions this view raises that could constitute a basis for EEC-inspired research in cognitive neuroscience.

## **6. Embodied embedded cognitive neuroscience**

In this section we put forward some foundations for empirical questions that emerge from the general framework presented above. As we see it, the task of a new (and strong) embodied embedded cognitive neuroscience, is to find out what the different *kinds* of neural influences are that alter the nature and course of the basic interaction cycle. The most basic question at this

point might be this: What are the ways in which the brain can influence the ongoing interaction between agent and world? How can the brain contribute, i.e., *add to*, the interaction that is already taking place? The more general question is: what is it, ultimately, that the brain does, if it is *not* (necessarily) representing knowledge of the world and if it is not (necessarily) computing programs for action?

Consider the role of the brain during autopilot behavior. The general function of the brain's behavior facilitating influence is to sustain the necessary conditions for the desired interaction between body and world to take place. The functional clusters, discussed in Section 5, can be thought of as temporarily setting the values of internal control parameters (Keijzer, 2001, Kelso, 1995) that do not prescribe but bias the inclinations of a system, 'nudging' its coordination tendencies by facilitating certain dispositions and hampering others. A general way to think about control parameters is that they influence the layout of an attractor landscape reflecting a variety of stable and unstable behaviors (attractors and repellers). Here are some suggestions for the functional kinds of neural control we envision:

1. Suppressing (blocking) potential behaviors that would otherwise (i.e., due to the basic interaction cycle) have emerged. Suppressing a behavior amounts to the process by which internal control parameters change the dynamical landscape such that a certain behavior is no longer reachable from the current state of the system. By effect, the system will converge onto another stable state.
2. Maintaining behaviors that would otherwise have died out, for instance by delaying other behaviors, or by generating positive or suppressing negative emotions. Maintaining a behavior means that whenever external forces would otherwise induce a change in the dynamical landscape such that the stable state becomes instable, the brain compensates

for this by changing internal control parameters such that the current stable state remains a stable state.

3. Enhancing behaviors, for example speeding up the start of a behavioral sequence that would otherwise have taken more time to start. In effect, this amounts to shaping effective behavior as a nearby attractor.
4. Favoring actions for the effect they will have on the brain's traffic regulation function.

An interesting notion here is Kirsh and Maglio's (1994) 'epistemic action', a behavior that affects the environment in such a way that it will ease the need for excessive control, i.e. by modifying the stimuli in the environment in such a way that they will address the behavioral repertoire in a more straightforward way. In other words, epistemic actions change the state of the world in such a way that the task of acting intelligently in it becomes *easier*. Epistemic actions are in effect changing external control parameters: that is, these are behavioral interventions that change the potential landscape of the organism-environment dynamics, such that 'solutions' are easily reached in the form of a natural convergence upon an attractor.

Neural control need not be taken as creating a procedure (a motor plan, for instance) to be executed, but might often be seen as simply *suppressing* the emergence of some natural behavioral pattern, thereby generating circumstances in which other behaviors can emerge.

Suppressing the first, automatic impulse, in favor of other, perhaps more 'intelligent' behavior, is an important quality of intelligent agents. 'Stop! Don't act now!' might be a very effective way of letting the environment re-stimulate the behavioral repertoire and thereby re-selecting a more appropriate behavior. The prefrontal cortex is a prime candidate for being involved in this form of control (Garavan, Ross, Murphy, Roche, & Stein, 2002). A complementary form of influence

might be neural control that enhances or strengthens some behavioral pattern. Sometimes the basic interaction cycle might be ‘tempted’ to kill a behavioral pattern on the basis of local environmental influences. Neural control from the brain might help to maintain the behavior, which can turn out to be beneficial for the agent in the long run. The common psychological terms associated with these ideas are ‘attention’ and ‘concentration’.

Given the assumption that the system recruits available resources in order to maintain functional stabilities several new questions arise: How are brain biases implemented at the neural level? How do they arise and change? What is the relation between this model of automatic, common sense behavior and classic, representational views on reasoning, i.e. deep thought? A different set of questions concerns the extent to which one can push the lack of control by one of the other components, while still maintaining the stability in behavior. In what specific situations does the brain fail to ‘compensate’ for the body and/or environment, and vice versa? When does that lead to radical break-down of the behavioral stability? And finally, can we identify situations where representational activities of the brain are required for sustaining behavioral stability? In other words, when are representations really necessary? (Haselager, Bongers, van Rooij, 2003; van Rooij, Bongers, & Haselager, 2002).

## **6. Conclusion**

One of the strengths of contemporary dominant cognitivist neuroscience seems to lie in the availability of an agreed upon conception of the brain, viz., as a modular information-processor. Given this conception of the brain the explanatory goals of cognitivist neuroscience are also clear, viz., to figure out what are the basic set of modular systems that comprise the brain. At present, the embodied embedded cognition (EEC) framework seems to be lacking such an agreed upon conceptualization of the brain, and accordingly the explanatory goals of an EEC

neuroscience program remain unclear. By providing a guiding metaphor--the brain as traffic regulator--and laying out a set of associated explanatory goals, we aim at contributing to the conceptual foundations of a strong EEC-inspired empirical neuroscience program.

Are the questions raised by our strong EEC perspective of brain and behavior any different from the kinds of questions traditional cognitivist neuroscience is trying to answer today? We believe they are. Classically, empirical data from brain scanning devices are used to find out what different kinds of *cognitive functions* the brain performs. Here ‘cognitive functions’ are explicitly conceptualized as information-processing tasks that depend on representations. That is: The success of the brain’s functioning is formulated in terms of how well the brain is able to model the outside world internally and how well it is able to construct on the basis of this model a solution to a problem, extracted from perceptual information, in the form of appropriate actions to be executed. Experimental conditions are constructed to ‘tap into’ some particular task. This task, present in one condition and absent in the other, is then believed to be carried out by a specific cognitive module. Variation in neural activity across conditions is seen as evidence of the neural mechanism that underlies the presumed ‘cognitive function’. Traditional cognitive neuroscience is fundamentally cognitivist in its conception of the brain and its role in the genesis of cognitive behavior.

In our proposal, neural mechanisms perform completely different *kinds* of ‘functions’. These functions are ‘cognitive’ in the sense that they help in the emergence of intelligent behavior, but fundamentally different in how they relate to the outside world and lower level interactions. The function of an embodied embedded brain is best described as creating several *controlling influences* on an existing *basic interaction cycle*. The success of this kind of brain functioning is assessed in terms of how well the brain is able to sustain a healthy balance

between environmental changes and body-internal demands. The goal of a strong EEC neuroscience is to find out what kinds of controlling influences exist, and what the underlying neural mechanisms are that generate these influences.

The strong EEC neuroscience program that we envision may well turn out to be methodologically more challenging than the current cognitivist neuroscience (the popularity of the latter may in part be explained by its methodological niceties, see also Section 2), but we believe that it is nevertheless a valuable option to pursue, if not instead of cognitivist neuroscience than at least in addition to it. Current methodologies for isolating and locating modules in the brain seem to be designed to answer an incomplete set of questions. Existing neuroscientific methods can be (and are being) used to generate evidence that representational content is grounded in bodily experience, what we refer to as weak EEC (as reviewed in section 3). But so far they do not generate penetrating insights into the way in which the brain actively supports and interconnects with ongoing intelligent and adaptive behavior as conceived by strong EEC. Weak EEC theories still maintain the cognitivist metaphor of mind as a decider and programmer of behavior and the methodological tools they adopt are not designed to address questions raised by a strong EEC perspective of the brain.

So one may ask: Can there be such a thing as a *strong* embodied embedded cognitive neuroscience? Clearly there is an opposition between the brain-centered research and strong EEC, *at the moment*, but we feel that this opposition is not necessary and even unfortunate. There are many questions raised by a strong EEC perspective of the brain that naturally invites neuroscientific study and experimentation. In this paper we have attempted to provide a first characterization of, or guiding metaphor for, the role of the brain in behavior. We also identified a set of global questions that a strong EEC neuroscientist would ideally set out to answer. By

what specific research methods this can be done and how far this research would take us remain open issues. But one thing seems clear: without incentive such methods are unlikely to be developed.

### Footnotes

(1) The words in the three categories were matched on word length and word frequency.

(2) Two caveats are important here: First, of course the brain is involved in human and lots of animal auto-pilot behavior. Our point is not that we are like *E. coli* in such cases. Still, the main tasks the brain is confronting in these kinds of cases are more of the ‘servant-type’ than of the ‘director-type’. Second, of course we are capable of rapidly switching from auto-pilot to deep thought behavior, e.g. when the environment obstructs the routine flow of things and some sort of problem-solving will have to take place before things can run their course again.

Phenomenologically such situations are characterized by experiences expressible like: ‘Whoa, what’s going on?’ One stops and starts thinking: Where did I put that knife? It was here only a minute ago! Perhaps some kind of representational model is needed in order to explain the brain’s interventions of the second kind. However, theoretical models are being constructed that explicitly deal with higher order cognition while remaining firmly within a dynamically situated context (Christensen & Hooker, 1994; Haselager, Bongers, van Rooij, 2003; van Rooij, Bongers, & Haselager, 2002).

(3) In later robots, artificial evolution and/or learning are used to create or enhance the behavioral repertoire (e.g., Nolfi & Floreano, 2000).

## References

- Agre, P.E. (1997). *Computation and human experience*. Cambridge: Cambridge University Press.
- Bickhard, M.H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: Impasse and solutions*. Amsterdam: Elsevier.
- Brooks, R. (1999). *Cambrian intelligence*. Cambridge: MIT Press.
- Cairns-Smith, A.G. (1996). *Evolving the mind: On the nature of matter and the origin of consciousness*. Cambridge: Cambridge University Press.
- Chiel, H.J. & Beer, R.D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends In Neurociences*, 20, 553-557.
- Christensen, W.D. & Hooker, C.A. (1994). Representation and the meaning of life. In H. Clapin, P. Slezak and P. Staines (Eds.), *Representation in mind: New approaches to mental representation*. Westport: Praeger
- Clancey, W.J. (1997). *Situated cognition: On human knowledge and computer representation*. Cambridge: Cambridge University Press.
- Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge: MIT Press.
- Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, 3, 115-120.
- Coltheart, M. (2002). Cognitive neuropsychology. In H. Pashler and J. Wixted (Eds.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (pp. 139-174). New York: John Wiley & Sons.
- Cox and Smitsman (this issue) Embedded Action Selection.

- Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Avon Books.
- Damasio, A.R. (1999). *The feeling of what happens: Body, emotion and the making of consciousness*. London: Vintage.
- Damasio, A.R. (2003). *Looking for Spinoza: Joy, sorrow and the feeling brain*. Orlando: Harcourt.
- Dunn, J. & Kirsner, K. (2003). What can be inferred from double dissociations? *Cortex*, 39, 1-7.
- Edelman, G. (1992). *Bright air, brilliant fire: On the matter of the mind*. New York: Basic Books.
- Edelman, G. & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.
- Edelman, S. (in press). Bridging language with the rest of cognition: Computational, algorithmic and neurobiological issues and methods. *Proceedings of Ithaca EMCL workshop*. Amsterdam/Philadelphia: John Benjamins.
- Fodor, J. (1983). *The modularity of mind*. Cambridge: MIT Press.
- Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge: MIT Press.
- Garavan, H., Ross, T.J., Murphy, K., Roche, R.A.P., & Stein, E.A. (2002). Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection and correction. *NeuroImage*, 17, 1820-1829.
- Hardcastle, V.G. & Stewart, C.M. (2002). What do brain data really show? *Philosophy of Science*, 69, 72-82.

- Haselager, W.F.G. (1997). *Cognitive science and folk psychology: the right frame of mind*.  
London: Sage
- Haselager, W.F.G., Bongers, R.M., & van Rooij, I. (2003). Cognitive science, representations and dynamical systems theory. In W. Tschacher and J-P. Dauwalder (Eds.), *The dynamical systems approach to cognition* (pp. 229-242). Singapore: World Scientific.
- Hendrik-Jansen, H. (1996). *Catching ourselves in the act: Situated activity, interactive emergence, evolution, and human thought*. Cambridge, MA: MIT Press.
- Hernandez-García, L., Wager, T., & Jonides, J. (2002). Functional brain imaging. In H. Pashler and J. Wixted (Eds.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (pp. 175-221). NY: John Wiley & Sons.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT-Press.
- Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. Cambridge: MIT Press.
- Keijzer, F.A. (2001). *Representation and behavior*. Cambridge: MIT Press.
- Kelso, J.A.S. (1995). *Dynamic patterns: The self-organization of brain and behavior*.  
Cambridge: MIT Press.
- Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18, 513-549.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: Chicago University Press.
- Langacker, R.W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*.  
Stanford: Stanford University Press.
- Langacker, R.W. (1990). *Concept, image, and symbol: The cognitive basis of grammar*. Berlin: Mouton de Gruyter.

- Nolfi, S. & Floreano, D. (2000). *Evolutionary robotics: The biology, intelligence and technology of self-organizing machines*. Cambridge: MIT Press.
- Norman, D.A. (2002). *The design of everyday things*. New York: Basic Books
- Pfeifer, R. & Scheier, C. (1999). *Understanding intelligence*. Cambridge: MIT Press.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Port, R.F., & Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge: MIT Press.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253-336.
- Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking? Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, 78, 143-168.
- Pylyshyn, Z.W. (ed.) (1987). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex.
- Thelen, E., & Smith, L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge: MIT-Press.
- Uttal, W.R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge: Bradford/MIT Press.
- Van Orden, G.C., Pennington, B.F., & Stone, G.O. (2001). What do double dissociations prove? *Cognitive Science*, 25, 111-172.
- van Rooij, I., Bongers, R.M., & Haselager, W.F.G. (2002). A non-representational approach to imagined action. *Cognitive Science*, 26, 345-375.
- Varela, F.J., Thompson, E. & Rosch, E. (1991). *The embodied mind*. Cambridge: MIT Press.

- Veale, T. & O'Donoghue, D. (2000). Computation and blending. *Journal of Cognitive Linguistics*, 11, 253-281.
- Veale, T. O'Donoghue, D. & Keane, M. (1999). Computability as a limiting cognitive constraint: Complexity concerns in metaphor comprehension about which cognitive linguists should be aware. In M. Hiraga, C. Sinha and S. Wilcox (Eds.), *Cultural, psychological and typological issues in cognitive linguistics* (pp. 129-155). Amsterdam/Philadelphia: John Benjamins.
- Wheeler, M. & Clark, A. (1999). Genic representation: Reconciling content and causal complexity. *British Journal for the Philosophy of Science*, 50(1), 103-135.
- Wiemer-Hastings, K., & Kurby, C.A. (in prep.). Selective adaptation effects on lexical access: Perceptual concept primitives?

**Figure captions**

*Figure 1.* An illustration of the structure of a reactive agent. Sensation and action are coupled by several behavioral layers, each competing for dominance. It is the state of the world that determines which layers get “selected” at a particular time.

*Figure 2.* Illustration of how brain contributes to the selection of behavioral layers. Here world and brain co-constrain behavioral tendencies.

*Figure 3.* A sketch of our view of the brain-behavior relationship. Behavior is a continuously ongoing process, represented by the basic-interaction cycle at the bottom. The brain does not create behavior, but instead serves as a biasing influence on the nature of the basic-interaction cycle.

Figure 1.

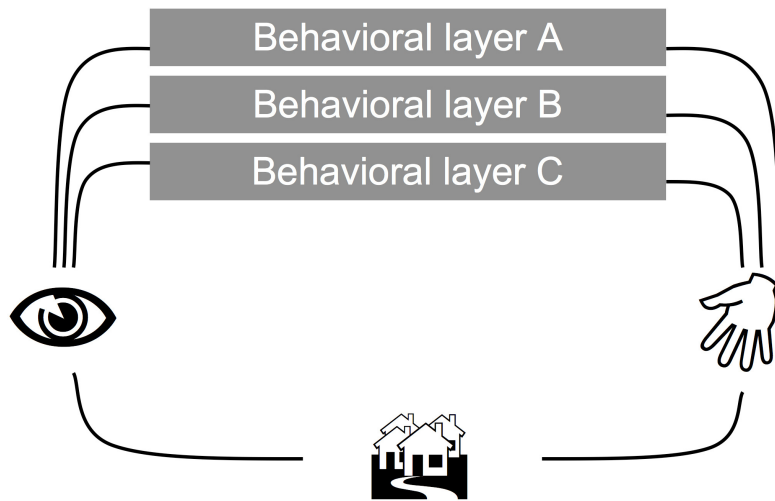


Figure 2.

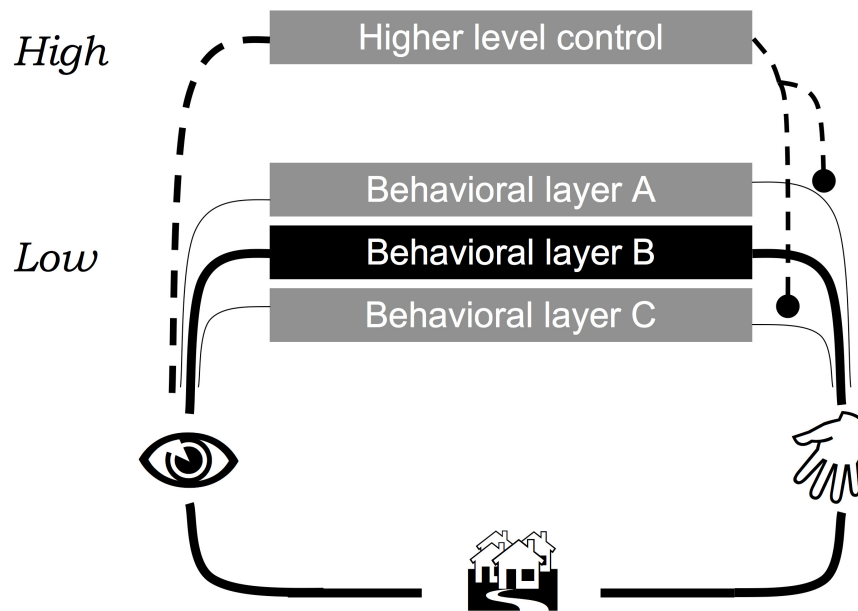


Figure 3.

