To appear in P. C. Garzón and A Gomila (Eds.), Handbook of Cognitive Science

******Prepublication draft******

A Lazy Brain? Embodied Embedded Cognition and Cognitive Neuroscience

Pim Haselager[1,*], Jelle van Dijk[2], and Iris van Rooij[1]

[1] Nijmegen Institute for Cognition and Information,
Radboud University Nijmegen, The Netherlands
[2] Mediatechnology, Utrecht University of Applied Sciences

*Corresponding author: Artificial Intelligence/Cognitive Science, NICI, Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands (Email: w.haselager@nici.ru.nl)

Abstract

Over the last decades, philosophers and cognitive scientists have argued that the brain constitutes only one of several contributing factors to cognition, the other factors being the body and the world. This position we refer to as Embodied Embedded Cognition (EEC). The main purpose of this paper is to consider what EEC implies for the task interpretation of the control system. We argue that the traditional view of the control system as involved in planning and decision making based on beliefs about the world runs into the problem of computational intractability. EEC views the control system as relying heavily on the naturally evolved fit between organism and environment. A 'lazy' control structure could be 'ignorantly successful' in a 'user friendly' world, by facilitating the transitory creation of a flexible and integrated set of behavioral layers that are constitutive of ongoing behavior. We close by discussing the types of questions this could imply for empirical research in cognitive neuroscience and robotics.

Keywords: cognitive neuroscience, embodied embedded cognition, abduction, computational intractability, control architecture, reactive robotics

1. Introduction

The *E. coli* shines in its simplicity. This single-cell organism can locate food in its environment without having any plan on how to look for it, nor having any beliefs about the world it finds itself in. Instead it finds food, and avoids toxics, by moving its flagella in one of two ways: it either tumbles about randomly or it swims straight ahead. Without specific stimulation it changes between these two modes every few seconds, thereby engaging in a random exploration of its environment. Once a chemical gradient in its environment is sensed (e.g. an increase in sugar level or a decrease in toxic substances), it increases the amount of swimming and decreases the random tumbling resulting in a process called *chemotaxis*. In effect, the bacterium swims upwards along a stream of increasing nourishment towards a food source and downwards along a

stream of decreasing toxics (Cairns-Smith, 1996, p. 90--94). The behavior of the *E. coli* could in principle be described in terms of the folk psychological concepts of beliefs, desires and intentions (Jonker, Snoep, Treur, Westerhoff, & Wijngaards, 2001), but these would be superfluous metaphorical ascriptions at best. It seems implausible and unnecessary to attribute such mental states to the *E. coli* as its behavioral success is readily explained in terms of the direct perception-action couplings in which sensed chemical gradients trigger different behavioral patterns (tumbling or swimming).

Humans are much more complicated organisms than *E coli*. Humans have much richer behavioral repertoires than the *E. coli* do, and humans can apply this repertoire with an exceptionally high degree of flexibility and sensitivity to environmental conditions, both past, present and future. It is this flexibility that is seen as a mark of human intelligence and what has proven so difficult to replicate in robots. On the one hand, the increased flexibility makes humans' lives easier, as it allows them to survive under wider environmental conditions than E. coli. But, on the other hand, it also seems to make things more difficult for humans, because it confronts them with a challenging control task. The challenge seems to be that humans need to *decide what* to do *when*. The E. coli do not have this problem (these bacteria do exactly what is triggered by the gradient of chemicals in their environment).

The received view in cognitive science and artificial intelligence is that cognitive systems can come to display the kind of intelligent behavior that is characteristic of human beings only by maintaining more or less accurate mental representations of the world (i.e., *beliefs*), which they derive from perceptual information. Based on their beliefs about states of the world, humans are assumed to make plans (i.e., *intentions*) with the aim of guiding motor behaviors in a way that meets certain goals (i.e., *desires*). This internalist, cognitivist view of the relationship between cognizing and behavior is inherited by much of contemporary cognitive neuroscience, resulting in the explanation of intelligent cognitive behavior as the product of powerful brains that can maintain world models and devise plans. In other words, contemporary cognitive neuroscience tends to see cognizing as something that the *brain* does.

We think that by construing the control problem posed to the brain in this way, cognitive neuroscience, like artificial intelligence, may be making a mistake. Maintaining a stable and approximately correct set of beliefs about the world that is sufficient for programming more or less successful behavior in situations of real-world complexity seems to pose a computationally too demanding task for a human (or any kind of) brain to perform. This computational intractability problem, long known to plague cognitivist models of cognition (Pylyshyn, 1987; Haselager 1997), clashes with the observation that people make split second decisions in everyday contexts, typically with good results. Hence, cognitive neuroscience may do well to consider alternative views of the control architecture of humans.

In this chapter we make the case for one such alternative control structure. Our control structure is inspired by the theoretical framework of *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 the body, the physical and social structure of the world, and the internal milieu of the organism's body all provide important constraints that govern behavioral interactions. From this perspective behavior is best explained by a 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, the processes subserving cognitive behavior cannot be directly mapped onto brain structures.

We are well aware of the apparent tension between an EEC perspective of the brain and contemporary cognitive neuroscience research (see also van Dijk et al. in press). Much of current cognitive neuroscience's methodology (e.g., brain imaging and single cell recordings) is built on the idea that the brain implements an encapsulated mechanism for cognizing that can be understood by studying the brain in almost complete isolation, independent from any realistic bodily interaction with the world. Accordingly, much experimental effort in cognitive neuroscience is devoted to figuring out *which* of the cognitive subprocesses (perception, abduction, planning, deciding) are performed *where* in the brain and *how* these processes are neurally implemented. This research aim makes sense if one presupposes that the body and world are merely external factors (related to the input and output) to cognition. But it is exactly this presupposition that is questioned by EEC.

In this chapter we review arguments *against* the exclusive adoption of the cognitivist conception in cognitive neuroscience, and *for* extending it with an EEC view. Of course, empirical researchers are not easily swayed by theoretical or philosophical argumentation alone, nor should they be. If EEC is to inspire cognitive neuroscience to extend its research methodology, so that it aligns with an EEC view of the role of the brain in cognitive behavior, then EEC may do well to formulate concrete research questions that are amenable to empirical testing by cognitive neuroscientist in the near future. In this chapter we therefore try to take some steps towards the generation of such concrete questions.

1.1. Overview

The chapter is organized as follows. We start in Section 2 by explaining the computational intractability problem, why it poses a formidable problem for cognitivism, and why we think that existing attempts to overcome the problem within a cognitivist framework fail or are otherwise problematic. In Section 3 we put forth some arguments for, and speculations about, how organisms can come to inhabit, and be adaptive in, relatively complex environments without the need for continuous high-level world modeling, planning and decision-making. Basically we argue that due to a natural fit between organism and environment, most of the time organisms can be 'ignorantly successful' in their 'user-friendly' environments. In Section 4 we sketch the contours of a 'minimalistic' control structure that could suffice for such ignorant successfulness by introducing the metaphor of traffic facilitation as a way of conceiving the main task for higher-level control mechanisms in the brain. According to this view, the brain does not primarily produce (through modeling, planning and deciding) behavior but rather, at least most of the time, inhibits or disinhibits perception-action loops that are constitutive of ongoing behavior. We discuss the types of questions this traffic facilitator metaphor could imply for empirical research in cognitive neuroscience experimentation as well as robotics.

2. The computational unfeasibility of a brain in complete control

We examine the computational demands of the task attributed to the brain by the cognitivist. Cognitivist accounts typically assume that central control systems work in two general stages: first, based on the information provided by the sensory input systems, 'higher' cognitive processes (sometimes referred to as 'central systems') form beliefs about how the world is; and second, the central system selects from the entire repertoire of possible actions a sequence of actions that when performed in the world as it is believed to be, will lead to the realization of certain goals. Both stages can be shown to run into the problem of computational intractability

(Bylander, 1994; Joseph & Plantinga, 1985), but for ease of presentation we will focus on the computational task posed by the first stage only.[1] Clearly, the beliefs generated in the first stage cannot be guaranteed to be true, always and everywhere, but assuming that behavioral success is to be explained by plans based on these beliefs they cannot be arbitrarily wrong either. It seems then that for a cognitivist account of adaptive behavior to work, one needs to assume that brains have a capacity for forming at least more or less accurate beliefs, at least sufficiently accurate to support the success of planned behavior most of the time. We present the following quote as just one example of the cognitivist idea that higher processes are involved in trying to make sense of the world on the basis of imperfect information in order to decide on action:

> "Action selection is a fundamental decision process for us, and depends on the state of both our body and the environment. Because signals in our sensory and motor systems are corrupted by variability or noise, the nervous system needs to estimate these states." […] "The approach of Bayesian statistics is characterized by assigning probabilities to any degree of belief about the state of the world ... Bayesian statistics defines how new information should be combined with prior beliefs and how information from several modalities should be integrated." (Körding & Wolpert, 2007, p. 319)

On this view, then, higher cognitive processes involved in planning and decision-making are engaged in generating abductive hypotheses that make (the most) sense of the perceived information, given everything else the cognitive system knows (Rock, 1983; Shanahan, 2005; see also Fodor, 1983, 2000). The word 'abduction' is not often used in neuroscientific literature. However, in order to ensure that one's beliefs about the world more or less correspond to what is actually the case in the world one seems to minimally require a capacity for domain-general abduction. Here, by 'abduction' is meant an inferential process that takes as input partial information about the world, or data (as produced by sensation and perception) and gives as output hypotheses about which states of the world are believed to currently hold and which ones not. For example, if an object looks like a duck (vision) and quacks like a duck (audition), then we might (or might not – depending on what else we perceive and believe) abduce that the object in front of us is a duck. We furthermore, might or might not abduce that the object is eatable, a bird, 2 feet long, etc. By 'domain-generality' is meant *both* that the abduction process can be informed by information coming from all of the input systems (vision, audition, olfaction, proprioception, etc.) *and* that the entertained hypotheses, and the information relevant to maintaining them, can span all kinds of content domains that are potentially relevant for human activity (the hypotheses can be about ducks, about people, about atoms, about the weather, etc.). This domain generality is also expressed sometimes by saying that human abduction processes are not informationally encapsulated (Pylyshyn, 1980, 1984).

It is the requirement of domain-generality that in a sense causes trouble when one wishes to devise computational procedures for abduction. The reason is that it implies that we cannot, in general, have good abductions by considering only a handful of observational facts and only a

[1] Alternatively, one may assume that the two steps are collapsed into one, in the sense that the probability of plan success is being evaluated by the central system for all possible plans against the background of all possible worlds consistent with current perceptions, and the plan that has the largest (or a large enough) probability of success is selected. For our purposes, the simplified two-step scenario suffices to make our points about the computational intractability of centralized (disembodied) inference, planning and decision-making. The same points would also apply to the collapsed-steps scenario sketched here, since its computational complexity is at least that of the two steps considered separately.

handful of relevant beliefs. In contrast, whether or not one should entertain belief *p*, given observational facts $d_1, d_2, \ldots, d_m$, depends also on one's whole system of background beliefs about the world, $p_1, p_2, \ldots, p_n$. Such belief systems may contain hundreds or thousands of beliefs, and hence $n >> m$. Moreover, these beliefs are not set in stone (neither are the observational facts by the way, which may be abduced to be misperceptions or illusions; see e.g. Thagard, 2000) and each and everyone of them is a potential candidate for updating when new observations are made. Given that the number of possible updates of beliefs (i.e., combinations of held beliefs) grows exponentially as a number of potentially held beliefs, efficient updating of the whole web of beliefs seems computationally prohibitive for minds/brains with finite computational resources.

To give a numerical illustration of the problematic nature of such an exponential growth, assume that there are in total, say, $n = 100$ beliefs in ones entire system of beliefs (a gross underestimation, we would think). Then there are already $2^n = 2^{100} > 10^{30}$ many possible truth assignments ('true' or 'false') possible; allowing values of believability between 'true' and 'false', as preferred by probabilists, makes the number of possibilities even larger. Clearly, exhaustively searching this space to find which truth assignment is supported by the observations at hand is impossible. Even if a brain (or a super-computer) has at its disposable as many parallel computational channels as there are neurons in the human brain (about 10^{14}), and if each such channel were capable of considering millions (10^6) of possible truth assignments per second, still the computation would require more than ten centuries to complete ($>10^{10}$ seconds). More importantly, there seems to be no other way possible to ensure that updating results in a stable and more or less accurate set of beliefs. This follows from the observation that all attempts to formally define the computational problem underlying abduction have resulted in a problem that is *NP-hard* (Abdelbar & Hedetniemi, 1998; Bylander, Allemang, Tanner, & Josephson, 1991; Thagard, 2000). What this means we explain next.

NP-hard problems are problems for which no practicable (i.e., polynomial-time) algorithm is known and it is strongly conjectured that no such algorithm can ever exist. In other words, it is conjectured that NP-hard problems can only be solved by some variant of exhaustive search (i.e., exponential-time) algorithms, which is why these problems are considered *computationally intractable* (Garey & Johnson, 1979, p. 8). Although the conjecture is so far unproven, it has strong empirical support.[2] There are currently thousands of NP-hard problems known (see e.g. the available online compendia). Moreover, it is known that if any one of these problems were computable in polynomial-time then all of them would be. Despite sustained efforts by mathematicians and computer scientists over the last four decades, nobody to this day has succeeded in devising a polynomial-time algorithm for an NP-hard problem--hence, the conviction that no such algorithm exists. Unless one would want to ascribe oracle-computing

[2] As an aside, we note that the conjecture is also strongly supported by mathematical intuition. The mathematical intuition derives from the believed inequality of two problem classes, called NP and P. Here, informally, NP can be thought of as a class of problems whose solutions can be easily checked, and P can be thought of as a class of problems whose solutions can be easily found. Now, the mathematical intuition (and perhaps the layperson intuition as well) says that NP may contain problems that are not in P (not all easily checkable problems need be easily solvable). To assist the non-mathematician's intuition, think of crossword puzzle or a game like Sudoku. For each such puzzle it is easy to check if a proposed solution is correct, but it is not clear that a solution is also always easy to find, i.e., there may be *hard* puzzles. Now, for technical reasons we cannot go into here, it is known that if an NP-hard problem would be computable without some form of exhaustive search, then this would imply that NP = P, which would violate mathematical intuition (see Garey & Johnson, 1979, for more details).

powers to central brain systems (something that would be akin to the avowed 'homunculus' in psychological explanation), it seems implausible that central brain systems have the capacity for efficiently computing NP-hard problems.

The theoretical obstacle posed by the computational intractability of abduction is greater than many cognitivists seem to realize. First of all, the problem cannot be detracted by choosing a different formalism for modeling abduction. Oaksford and Chater (1996), for example, argued for a switch from non-monotonic logics to Bayesianism for modeling human abductive inference based on the computational intractability of the former. But such a move seems in vein given that Bayesian models of abduction are as computationally intractable (if not more than) all other existing models of abductive inference---such as, non-monotonic logics, covering models, constraint satisfaction models, and neural network models (Abdelbar & Hedetniemi, 1998; Bruck & Goodman, 1990; Bylander et al, 1991; Cooper, 1990; Thagard & Verbeurt, 1998; Thagard, 2000).

Second, the computational intractability problem also cannot be detracted by loosening the quality of the abductions. It is often suggested in the cognitive science literature that computationally intractable problems can be approximately computed efficiently (e.g., Chater, Oaksford, Nakisa, & Redington, 2003; Chater, Tenenbaum, & Yuile, 2006, Love, 2000), but this seems at best a misrepresentation of the state of the art. It is well known that many NP-hard problems cannot be efficiently approximated (Arora, 1998; Yoa, 1992), and almost all are inapproximable if only a constant sized error is allowed (Garey & Johnson, 1979). Moreover, models of abduction are NP-hard to approximate even for quite liberal criteria of approximation (Abdelbar & Hedetniemi, 1998; Roth, 1991), and where claims are made of polynomial-time 'approximation' algorithms for abduction problems (e.g., Thagard & Verbeurgt, 1998) those algorithms do not approximate the required solution itself (i.e., the truth assignment), but instead its associated value, e.g., coherence or probability (see Hamilton et al., 2007, for a discussion).

Third, computationally intractable problems cannot be rendered tractable by a divide and conquer strategy. For example, in the cognitive science literature it is sometimes suggested that the computationally intractability problem plaguing a single, central abduction / planning system can be overcome by postulating the existence of a large set of 'modules' each being able to efficiently update beliefs, or make plans, for a specific domain of situations or 'contexts' (cf. the 'massive modularity' of Cosmides and Tooby, 1994, the 'toolbox of heuristics' of Todd and Gigerenzer, 1999, and the 'multiple models' of Wolpert and Ghahramani, 2000; see also Carruthers (2003a/b) and Sperber (2002) for discussions). If each such module implements a tractable computation, then it may seem that the whole system could tractably update our beliefs, and make plans, in all psychologically relevant contexts. However, even granting the number of required modules could be efficiently stored in the human brain (think of the potentially quite large number of possible contexts), a modular system cannot tractably compute any computationally intractable problem at risk of contradiction. If a problem were to be tractably computable by a modular system, then this would imply that that problem does not belong to the class of computationally intractable problems.[3] If a problem is intractable, as seems to be the case for domain-general abduction, then *no* algorithm for tractably computing it can exist.

[3] It would also mean that Q is tractably computable by a single non-modular system, because a non-modular system could tractably compute Q by (i) simulating the process by which the modular system selects the right module for the current context, and (ii) simulating the workings of the selected module. If both steps are tractable for the modular system, then the simulation is also tractable for the non-modular system.

Much more can be said about the topic of computational intractability, its proposed solutions and their failings (see, e.g., van Rooij, in press), but for our purposes the point is merely this: The computational intractability problem is not going to go away for cognitivist models of planning and action control. The only way to achieve tractability of control, so it seems, is to assign an easier computational task to control processes than domain-general abductive inference, and in effect make the explanation of success of behavior to a large extent independent of the success of our abductions of beliefs about the world. The question, of course, is how the successfulness of behavior in the world can be explained if not by an appeal to a control system that plans on the basis of beliefs about the world. In the next section will put forth an argument for why it may be plausible to assume that organisms with control structures that maintain no internal model of the world can nevertheless behave successfully and adaptively in the world. Moreover, we argue, that such organisms may very well come to inhabit the most complex or challenging worlds that their control structures can successfully handle, or approximations thereof.

3. Ignorantly successful in a user-friendly environment

No animal is behaviorally adapted to react in appropriate ways to all possible changes of all possible variables existing in the 'world out there'. Consider an ant, stamped upon by a casual pedestrian: this poor creature 'has no idea' what hit him, and, more importantly, it has no means whatsoever to counteract such occasions. The ant is either extremely lucky, or it dies. From the perspective of the ant, a passing pedestrian is a true Deus ex Machina. Still ants are successful creatures. On the whole, every organism seems to get by pretty successfully, using the behavioral capacities it possesses. So how is it that organisms can be successful in a complex and unpredictable world? The speculative idea we pursue in this section is based on the assumption that the *local*, or *personal* environment in which an intelligent creature is situated is not formed independent from the organism's own behavioral and evolutionary history. "Environments" are not simply pre-given, arising out of nothing. Organisms do not wake up to find themselves in completely new, unfamiliar, and hostile worlds. In a confined region of the global chaos we call reality, each creature 'makes a living', based on its sensory capacities and its behavioral repertoire, thereby creating its own *Umwelt* (Von Uexkull, 1934; Ziemke & Sharkey, 2001). In the words of Varela, Thompson and Rosch (1991) one might say that the organism, by its own actions, *brings forth*, or *enacts*, a world. In yet other words, organisms and their environments can be said to *co-evolve* (Chiel & Beer,1997; Deacon, 1997), or as Mead (1934) put it:

> "The sort of environment that can exist for the organism, then, is one that the organism in some sense determines. If in the development of the form there is an increase in the diversity of sensitivity there will be an increase in the responses of the organism to its environment, that is the organism will have a correspondingly larger environment.(…) In this sense it selects and picks out what constitutes its environment. It selects that to which it responds and makes use of it for its own purposes, purposes involved in its life-processes. It utilises the earth on which it treads and through which it burrows, and the trees that it climbs; but only when it is sensitive to them." (Mead, 1934, 245, quoted in Jarvilehto, 1999).

Our suggestion is that it is this interdependency between organisms and their environments that makes these environments generally facilitative to interaction. It is this intimate 'fit', we speculate, which ensures that actions, once taken, will generally prove to be successful/adaptive.

Moreover, under most, ordinary, circumstances, inappropriate actions will generally prove to be repairable: we are allowed to make mistakes in our Umwelt, so that we may even learn something along the way. For example, think of the way in which parents provide safe environments for their offspring to explore and learn in. In other words, the naturally emerging embodied embedded behaviors of an organism generally tend to be quite effective for the survival of *that* organism in *that* Umwelt. Consider that most ants live their lives successfully, without knowing about, nor having had to deal with, stamping feet. Most ants are ignorantly successful. And so, we claim, are we humans.

An ignorantly successful interaction with a by and large user-friendly environment might very well be an apt description of what takes place during ongoing behaviors of individual human beings in daily life. As an illustration of this, consider a situation in which a human being is in need of 'locating an often used object in the kitchen during cooking', e.g. a milk-beater. In cognitivist theory this is a problem of search, involving not only inspecting the visual scene, but also memory, as when we try to remember (or form hypotheses about) where we may have put the object. In practice, however, memory search is often not needed. In many situations the structure of the environment naturally constrains the kind of actions that can be performed, and one may question whether the brain needs to search through mental models and memory stores at all. For instance, in a kitchen, some drawers and shelves are more easily reached by the agent than others. Such drawers and shelves will be among the first to be inspected, that is, if the natural flow of body-world interaction is followed. Note that this is a physical constraint that exists because of the bodily characteristics of the person and the physical organization of the kitchen, and independent from any potential deliberation in the person's mind. Chances are that a daily used object like a milk-beater is also *put* on one of those easily reachable shelves or drawers, perhaps even by the person herself. So when we experience ourselves doing a seemingly random inspection of drawers and shelves instead of a rational search, we are actually being constrained both by body and world, leading to higher chances of behavioral success even if these actions in isolation would seem to be senseless or random. The example illustrates how success of behavior can follow from the 'fit' between the person's behaviors and the local environment. Where you can put objects most easily is also where you can look for them most easily, which in turn is where you have a high chance of finding the object that you where looking for.

Environments can thus be 'user-friendly', not unlike a well-designed interface. An agent's natural tendencies for action can tend to match the environmental structure in ways that turn out to be functional with respect to the agent's needs. This would be the case, for example, when the agent's behavioral repertoire and the structure of the situated environment co-developed with one another[4]. In such a process of co-evolution organism and situated environment (Umwelt) are mutually affected by one another. Evolution is sometimes seen as a one-way effect in which an animal adapts to changes in its environment. What is less often recognized is the reversed process, in which changes in an organism's structure might also lead to changes in the (situated) environment. Have polar-bears turned white because their environment became snowy? Or did the whiter-colored polar-bears use their skin-colour to their advantage, leading them to travel ever further up north into snowy territory? Or consider the human eye. From a traditional perspective, the eye would be viewed as the animal's *solution* to

[4] Incidentally, such a fit between organism and environment might, at least for human beings, emerge not only for a species on an evolutionary timescale but also, for an individual, from the ongoing interaction with the environment during his lifetime.

an environmental *problem*. An evolutionary explanation might begin by stating that, at some point, due to a change in the environment, the ability to detect the visible spectrum of light became relevant for survival (where previously it had not been). How to acquire the capacity to use light can be seen as the *problem*. Selection forces then procure a sensor that is able to detect light, in humans the eye. This is the *solution*. We think that such a view need not be correct. For one thing, it has been argued that sometimes structural properties of organisms emerge and persist (over numerous generations) long before the property in question becomes adaptive (Goodwin, 1994). In other words, evolution creates exaptations (Gould & Vrba, 1982; Gould, 1991), which in a way can be seen as 'solutions' for problems that don't even exist (yet). A perhaps even more fundamental question is why the visible spectrum of light became relevant for survival in the first place. In many situations, it is not unreasonable to suggest that such aspects of the environment co-evolve with changes in the behavioral repertoire of the organism itself. Consider, as a hypothetical example, a blind creature that has developed the means to move significantly faster than before. Now speed may be a useful adaptation, but it also presents dangers, such as a fatal collision. For this animal, sensitivity to distal (e.g. visual) rather than proximal sources now becomes adaptive, whereas its slow ancestor would have had no use for it. Hence, once the eye has evolved, the system relaxes into a stable relation between animal and environment, in which its new eyes team up nicely with its fast legs. But that is not the end of it. Once there is vision, the environment 'broadens up' once more. A 'visual environment' might help the animal in dealing with the dangers of going fast (the original 'problem'), but it also creates new challenges. As Lock (2003, p.105) states:

> "Simpler organisms can handle their simpler worlds by less complex means, but once evolution has come up with the where-withal for simpler organisms to handle their somewhat simpler selection problems, then it effectively creates for itself a new problem. That is, as organisms find ways of sustaining themselves, they create new potential sources of energy that can be preyed upon. And as new sources of energy, they present more complex worlds for their possible prey to operate in."

That is, when compared to its blind ancestors, the eyed creature faces some challenges of its own: How to cross that distant river, how to climb that far-away tree, how to fight that approaching competitor, and so on. The idea is thus that behavioral capacities co-evolve with changes in the organism's environment in a corresponding manner. New capacities enable the animal to be adaptive in that new environment. But the new situation has both 'advantages' and 'disadvantages'. The advantage is that the new extension to the behavioral repertoire helps the organism in dealing 'better' with some aspects of the environment than before. On the other hand, the disadvantage is that the animal has now projected itself into a new environment and this environment poses new cognitive challenges as compared to the previous situation. The development of new capacities, seen as a means to resolve some tension between organism and environment, can therefore also be seen as *generating* new challenges as well: new kinds of behaviors lead to an extension of the environment, which poses new demands. Therefore, instead of saying that animals become *more* adaptive with each step in evolution, we would rather formulate it as animals becoming *equally adaptive again and again*, at each new critical equilibrium (Goodson, 2003), albeit in a broader range of (more complex) environments. For a related view of the co-evolution of psycho-linguistic capacities and socio-linguistic environments, see Deacon (1997).

In sum, we propose that the local, situated environments in which organisms are embedded are relatively comfortable and safe environments. Organisms and their environments co-develop, making environments generally 'user-friendly' life-worlds. We argued that success of behavior follows from the 'fit' between the embodied embedded repertoire of the organism and the structure of the situated environment. Next, we showed how new capacities in effect broaden up the situated environment, which has both upsides as well as downsides: new possibilities for action and perception may be useful in dealing with certain existing challenges, but they also generate new challenges as well.

4. Generating research questions for cognitive neuroscience and robotics

As indicated in the introduction, EEC it would be highly desirable for EEC to formulate concrete research questions that can provide the basis for research in cognitive neuroscience. A first apparent obstacle is that in the current neuroimaging methodology the movements of subjects have to be restricted almost completely in order to reduce noise. This prevents anything like the occurrence of the natural organism-environment fit, discussed above, that forms the basis for the view on brain control to be outlined in this section. Another problem is that the perspective of EEC tends to get formulated at a rather abstract, philosophical or even generally descriptive, level. Hence, most statements (including our own so far) about the value of EEC tend to be far removed from concrete empirical research questions in cognitive neuroscience. A third problem is that existing theories and models of EEC commonly deal with relatively low-level organism-environment interactions, usually as far removed from the complexity of daily life behaviors as the research of the often scorned cognitivist perspective (hence, e.g., Clark & Toribio's (1994) challenge to deal with 'representation-hungry' cases of behavior; see also van Rooij et al. (2002)). These problems are indeed formidable and cannot be solved within one chapter. However, we do feel that there are enough ingredients available, from the area of robotics as well as from neuroscience, in order to at least tentatively sketch a view that might lend itself to empirical testing. In this section, then, we will try to work our way from a metaphorical depiction of high-level brain functioning during common sense behavior to its consequences for empirical research in robotics and cognitive neuroscience.

Brooks (1999, p.81) suggested that it is fundamental for an organism to have "the ability to move around in a dynamic environment sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction." He modeled this capacity by means of his well-known layered architecture: reactive creatures consisting of behavioral layers that each instantiate a direct input-output coupling. According to Brooks, it is a major advantage of his approach that no intermediate (in between input and output) world modeling, planning and decision making takes place. Instead, layers compete for dominance on the basis of the input received by the system. From this perspective a creature can be seen as a repertoire of behavioral dispositions and the environment selects from it. A creature 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 (as well may the case, as indicated in Section 3) its overall conduct may be satisfactory for a long time.

The fit between environment and behavioral repertoire might to a large extent underlie the relative success of most of our common sense behavior in daily life, such as having a drink in a bar, going home, or making dinner, etc. Common sense behavior actually consists in quite complicated sequences of behavior, even though it does not require the type of planning and decision making characteristic of say playing a chess game or buying a house. Instead we seem to operate more or less on 'autopilot'; our behavior flows naturally out of the stimulations from the environment.

In the reactive robots of the early 90's, the number of distinct behavioral layers was typically small and the precedence relations between them were set beforehand and were hardwired into the system. This resulted in creatures not unlike the *E. coli* discussed earlier. However, once the set of basic behavioral capacities of a creature become larger, and its sensorimotor capacities quite rich, a more flexible and integrated way of setting up behavioral layers and their interrelations becomes necessary. To illustrate, consider the following: If an organism has *n* basic behavioral layers available, then it could come to display, in principle, as many as 2^n distinct behavioral patterns by simply turning "on" some layers and turning others "off". With even as few as 32 layers this could result in as many as $2^{32} = 10^{10}$ distinct potential behaviors, which would, to quote Wolpert and Kawato, be "sufficient for a new behavior for every second of one's life" (1998, p. 1318). If additionally quantitative adjustments are possible---i.e., states in between "on" and "off", possibly implementing dominance relations---then the same organism would have the capacity for displaying an even larger number of possible behaviors. To help regulating the selection (or dominance relations) of behavioral layers, we suggest, is the main task of the high-level control function of the brain. In other words, instead of interpreting the brain's control system as the driver or pilot of the body, we see it as a *traffic regulator* (van Dijk et al., in press)---it is (merely, but importantly) assisting the environment driven selection from the behavioral repertoire. Notably, we do not propose that this traffic facilitation is achieved by computing the best (or even, a good enough) behavior from the set of possible behaviors given the current context (see, e.g., Körding & Wolpert, 2007; Wolpert. & Ghahramani, 2000; Wolpert & Kawato, 1998), because doing so would lead us right back to the computational intractability problem discussed in Section 2. Therefore, contrary to the traditional view of the control system as involved in world modeling, planning and decision making, we would like to hypothesize that the control function of the brain works in a, dare we say, more 'lazy' way.

There may be several ways in which one could conceive of a 'lazy' control system. We will describe just one such possibility here, drawing on an analogy with the control system of the *E coli*. Recall that the *E coli* can perform two modes of behaviors (tumbling or swimming), and the probability with which it switches between these two modes depends on chemicals (food or poison) it picks up from the environment. In a similar vein, our lazy control system may work by stochastically sampling from the set of behavioral options with a non-uniform *bias*, i.e., not every behavioral option is equally likely to be selected. The bias can be represented by a probability distribution *P* over the set of possible behaviors (e.g., combinations of "on" and "off" layers and/or combinations of dominance relations), where P(*t*, *i*) would denote the probability that behavioral disposition *i* is sampled at time *t*. Here, the bias *P* may be fixed, but more likely it is variable over time, e.g., as a function of experience and the organism's internal (homeostatic) milieu. This proposal raises several (more or less) concrete questions for cognitive neuroscience: How is *P* implemented in the human brain? What is the shape of the distribution *P* for humans? Is *P* fixed or variable? If *P* is variable, what is it a function of? If *P* is a function of experience

and/or homeostatic states, how do these factors contribute to changes in the distribution *P* over time, both descriptively and mechanically?

It seems to us that these questions can in principle be researched using (existing and developing) cognitive neuroscientific methods. Consider for instance the question of how such a lazy control system could be implemented in the brain. A concept that could help to elucidate how the brain might be 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. A similar concept, that of neuronal assemblies, is discussed by Chakraborty, Sandberg & Greenfield (2007, p.491):

> "Large-scale, coherent, but highly transient networks of neurons, 'neuronal assemblies', operate over a sub-second time frame. Such assemblies of brain cells need not necessarily respect well-defined anatomical compartmentalisation, but represent an intermediate level of brain organisation"

Functional clusters or neuronal assemblies can be assumed to implement short-lived changes in the organisms behavioral dispositions. In that case, the nature of the postulated bias *P* with which behavior dispositions are sampled could be experimentally investigated by studying the stochastic dependencies between different possible functional clusterings over time. We may observe that of the many different ways in which neural systems may cluster in principle, only relatively few cluster types happen with high frequency in practice over long periods of time under constant conditions. If so, this would suggests that *P* is relatively high peaked, implementing a stronger bias than when *P* would be flat throughout. Also, the hypothesis of non-constancy of *P* could be investigated by trying to fit a constant model to the observed stochastic dependencies and see if it fails to account for the observations. Following this, different *P*s, each a different hypothesized function of internal conditions and environmental factors, can be formulated and tested for their ability to explain observed stochastic dependencies of clustering over time and under variable conditions. Of particular interest and relevance for the latter type of experimental investigation would be to consider internal homeostatic states as variables for the function *P*, since by analogy with the *E coli* we hypothesize that much (if not all) of the bias in our sampling of behavioral dispositions is a function of such states.

Our proposal of a 'lazy' traffic facilitator control system also raises a question that we think may be answered using robotic simulation: How can humans, or any other complex organism, come to have a bias *P* that works well enough for the organism to get around the world on 'auto-pilot', without giving the selection of behaviors much thought, most of the time? We think that the answer lies in the type of co-evolution of control systems (in this case the bias *P*) with the life world of the organism, as described in Section 3. This explanation may be tested, or at least a proof of concept may be given, using robotic simulation. For example, a robotic simulation could start by endowing robotic systems with a 'lazy' control system P_0 and letting it evolve for n generations through $P_1, P_2, \ldots P_n$ in interaction with its life world. By systematically manipulating (i) the set of layers available to the robot, (ii) the nature of the initial P_0, (iii) the

way each P_i depends on internal and external conditions of life world *i*, (iv) aspects of the environment, and (v) the nature of the evolution process, one could get a better understanding of how these factors (i)–(v) interrelate. Hypotheses about the interrelation generated in the simulation process may serve as hypotheses for how these factors relate for (higher) organisms. To the extent that such hypotheses pertain to factors (i) – (iv) for animal and human brains, bodies and environments they can again be subjected to cognitive neuroscience testing.

Although we realize that our suggestions for experimentation in cognitive neuroscience and robotics need to be worked out more concretely in order to result in actual simulations and experiments, we do feel that they indicate that the traffic facilitation metaphor and the general view of EEC underlying it are not too far removed from empirical investigations.

5. Conclusion

Compared to for instance the *E. coli*, humans have an exceptionally rich behavioral repertoire that gets applied with great flexibility and sensitivity to environmental conditions. We argued against the received view in cognitive neuroscience, i.e. that cognitive systems can display this behavior only by maintaining mental representations of the world on the basis of which plans are made in order to achieve specific goals. We explained how such a position leads to the problem of computational intractability. We proposed that effective control may be possible for a more tractable, even 'lazy', control system that does not maintain any internal models of the world, assuming that such 'lazy' control systems co-evolve with the bodies and environments of organisms. This co-evolution ensures a certain degree of "fit" between the control system of an organism and its life world. The 'lazy' control mechanism that we postulated raises several interesting questions, each of which we think is amenable to experimental investigation using brain measuring methods. Also, our claim that 'lazy' control systems can plausible evolve, even for quite complex organisms in quite complex environments, can be directly investigated using the methods of robotic simulation. In all, we hope to have shown that an EEC view on the higher-level control functions of the brain is not only possible, but that it can be made precise enough to suggest experimental investigation in cognitive neuroscience, as well as robotics.

References

Abdelbar, A.M. & Hedetniemi S.M. (1998). Approximating MAPs on belief networks is NP-hard and other theorems. *Artificial Intelligence, 102*, 21–38.

Archibald S., Mateer C., & Kerns K. Utilization behavior. Clinical manifestations and neurological mechanisms. *Neuropsychology Review, 11:* 117-130, 2001.

Arora, S. (1998). The approximability of NP-hard problems. Survey based upon a plenary lecture at the *ACM Symposium on Theory of Computation*, 1998.

Brooks, A. (1997). *Cambrian intelligence: The early history of the new AI*. Cambridge, MA: MIT Press.

Bruck J. & Goodman, J.W. (1990). On the power of neural networks for solving hard problems. *Journal of Complexity, 6*(2), 129-135.

Bylander, T. (1994). The computational complexity of propositional STRIPS planning. *Artificial Intelligence, 69*, 165-204.

Bylander, T., Allemang, D., Tanner, M. C., & Josephson J. R. (1991). The computational complexity of abduction. *Artificial intelligence, 49*, 25-60.

Cairns-Smith, A. G. (1996). *Evolving the mind: On the nature of matter and the origin of consciousness*. Cambridge: Cambridge University Press.

Carruthers, P. (2003a). On Fodor's problem. *Mind & Language, 18*(5), 502–523.
Carruthers, P. (2003b). Is the mind a system of modules shaped by natural selection? In C. Hitchcock (Ed.), *Contemporary Debates in the Philosophy of Science*. Blackwell, Oxford.
Chakraborty, S., Sandberg, S., & Greenfield, S. A. (2007). Differential dynamics of transient neuronal assemblies in visual compared to auditory cortex. *Exp. Brain Res., 192*: 491-498.
Chater, N., Oaksford, M., Nakisa, R. & Redington, M. (2003). Fast, frugal and rational: How rational norms explain behavior. *Organizational Behavior & Human Decision Processes, 90*, 63-86.
Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences, 10*(7), 287-291.
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.
Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.
Clark, A. & Toribio, J. (1994). Doing without representing?. *Synthese, 101*, 401-431.
Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence, 42*(2-3), 393-405.
Cosmides, L. & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.
Deacon, T. (1997). *The symbolic species: The co-evolution of language and the human brain*. Penguin Press.
Edelman, Gerald M. (1992). *Brilliant Air, Brilliant Fire: On the Matter of Mind*. New York: Basic Books.
Edelman, Gerald M and Tononi, Giulio (2001). *Consciousness: How Matter Becomes Imagination*. London: Penguin Books.
Eslinger P. (2002). The anatomic basis of utilisation behaviour: a shift from frontal-parietal to intra-frontal mechanisms. *Cortex, 38*, 273-276.
Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: The MIT Press.
Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: The MIT Press.
Ford, K. M. & Pylyshyn, Z.W. (Eds.) (1996). *The robots dilemma revisited: The frame problem in artificial intelligence*. Ablex Publishing.
Garey, M. R. & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. New York: Freeman.
Lock, A. (2003) Book Review of The Evolution and Function of Cognition by Felix Goodson Mahwah, N.J.: Lawrence Erlbaum Associates, Inc. 2003. *Human Nature Review*, 3: 104-107.
Gould, S. J., & Vrba, E. S. (1982). Exaptation - a missing term in the science of form. *Paleobiology*, 8(1), 4-15.
Gould, S. J. (1991). Exaptation: A crucial tool for evolutionary psychology. *Journal of Social Issues*, 47, 43-65.

Hamilton, M., Müller, M., van Rooij, I., & Wareham, T. (2007). Approximating solution structure. In E. Demaine, G. Z. Gutin, D. Marx, and U. Stege (Eds.), *Structure Theory and FPT Algorithmics for Graphs, Digraphs and Hypergraphs*. Dagstuhl Seminar Proceedings (Nr. 07281). Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

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.

Jarvhilehto, T. (1999) The theory of the organism-environment system: III. Role of efferent influences on receptors in the formation of knowledge. *Integrative Physiological and Behavioral Science*, 34, 90-100

Jonker, Snoep, Treur, Westerhoff, & Wijngaards (2001). Putting intentions into cell biochemistry: An artificial intelligence perspective. *Journal of Theoretical Biology*, 214, 105--134.

Joseph, D. A. & Plantinga, W. H. (1985). On the complexity of reachability and motion planning problems. *Proceedings of the First ACM Symposium on Computational Geometry* (pp. 62-66). ACM Press, New York.

Kording and Wolpert, D. (2006) Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10*(7), 320-326.

Love, B. C. (2000). A computational level theory of similarity. *Proceedings of the Cognitive Science Society*, 22, 316-321.

Oaksford, M. & Chater, N. (1998). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning*. Hove, UK: Psychology Press.

Mead, G.H. (1934). *Mind, self, and society*. Chicago: Chicago Univ. Press.

Pylyshyn, Z.W. (1980). Computation and cognition. *Behavioral and Brain Sciences, 3*, 111-169.

Pylyshyn, Z. W. (1984). *Computation and cognition: Towards a foundation for cognitive science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (Ed.) (1987). The *robot's dilemma: The frame problem in artificial intelligence*. Ablex Publishing

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press

Roth, D. (1996). On the hardness of approximate reasoning. *Artificial Intelligence, 82*, 273-302.

Shanahan, M.P.. (2005). Perception as abduction: Turning sensor data into meaningful representation. *Cognitive Science, 29*, 103-134.

Sperber, D. (2002). In defense of massive modularity. In E. Dupoux (Ed.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 47-57). Cambridge, MA: The MIT Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Thagard, P. & Verbeurgt, K. (1998) Coherence as constraint satisfaction. *Cognitive Science, 22*, 1–24.

Todd, P.M. and Gigerenzer, G. (2000) Precis of simpleheuristics that make us smart. *Behavioral and Brain Sciences, 23*, 727–780.

Varela, F.J., Thompson, E., & Rosch, E. (1991) *The embodied mind: Cognitive science and human experience.* Cambridge, MA: The MIT Press).

van Dijk, J., Kerkhofs, R., van Rooij, I., & Haselager, P. (in press). Can there be such a thing as

embodied embedded cognitive neuroscience? *Theory & Psychology*.
van Rooij (in press). The tractable cognition thesis. *Cognitive Science*.
van Rooij, I., Bongers, R. M., & Haselager, W. F. G. (2002). A non-representational approach to imagined action. *Cognitive Science, 26*(3), 345-375.
Von Uexkull, J. (1934). A stroll through the worlds of animals and men. In C. Schiller (ed.), *Instinctive Behavior*. New York: International Universities Press, 1957.
Wolpert, D.M. & Ghahramani, Z. (2000) Computational principles of movement neuroscience. *Nature Neuroscience,* 3 supp, 1212--1217.
Yoa (1992). Finding approximate solutions to NP-hard problems by neural networks is hard. *Information Processing Letters, 41*, 93–98.
Ziemke, T. & Sharkey, N. E. (2001) A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica, 134*(1-4), 653-694