To appear in *Cognitive Systems Research Journal*

Representation in dynamical and embodied cognition

Fred Keijzer

Faculty of Philosophy
University of Groningen
A-weg 30
9718 CW Groningen
The Netherlands
E-mail: f.keijzer@philos.rug.nl

Abstract

The move toward a dynamical and embodied understanding of cognitive processes initiated a debate about the usefulness of the notion of representation for cognitive science. The debate started when some proponents of a dynamical and embodied approach argued that the use of representations could be discarded in many circumstances. This remained a minority view however and there is now a tendency to shove this critique of the usefulness of representations aside as a non-issue for a dynamical and situated approach to cognition (Clark, 1997b; Bechtel, 1998). In opposition, I will argue that the representation issue is far from settled, and instead forms the kernel of an important conceptual shift between traditional cognitive science and a dynamical and embodied approach. This will be done by making explicit the key features of representation in traditional cognitive science and by arguing that the representation-like entities that come to the fore in a dynamical and embodied approach are significantly different from the traditional notion of representation. This difference warrants a change of terminology to signal an important change in meaning.

1 A criticism of the use of representations

The last few years, new themes are acknowledged as important factors of cognition. Cognition is now increasingly seen as a phenomenon which is intrinsically embodied. Cognition is not restricted to happenings within the head but involves also bodily and environmental happenings. Both the cognitive agent’s body, as well as the environment in which the agent operates, provides a lot of structure that the cognitive agent’s internal cognitive apparatus uses to solve the agent’s problems. As a result, the internal cognitive apparatus can remain a lot simpler than was previously thought. For example, moving about in a city does not require an elaborate representation of that city’s lay out. It usually suffices that a cognitive agent recognizes a limited number of decision points where it has to take a specific turn. In between, just following the street will lead it to a next decision point until a destination is reached.

Awareness that cognition is essentially embodied is connected with an appreciation of the central role played by the dynamical, time-dependent, relations involved in the interactions between a neural system, a body and an environment (Beer, 1995; Van Gelder, 1995, 1998). In this perspective, cognitive agents are dynamical systems and, more specifically, cognitive functioning has to be understood as a dynamical process for which the tools provided by dynamical systems theory are best suited to make sense of them (Van Gelder, 1995, 1998). This perspective on cognition will here be called the Dynamical and Embodied view on Cognition, or DEC for short.

At the onset of DEC, several authors downplayed the importance of the notion of representation for explaining cognition, and sometimes even hinted that one could do fully without this notion (Beer, 1995; Brooks, 1991; Kelso, 1995; Thelen & Smith, 1994; Van Gelder, 1995). They presented case studies (e.g. robots, developmental processes, and motor coordination) where the embodiedness and/or the dynamical characteristics of the agent could account for the observed behavior without invoking the notion of representation. Most famous are probably Brooks' insect robots of the late eighties and early nineties (Brooks, 1986, 1999). Compared to previous attempts, these robots became better in dealing with a real world environment by skipping the idea that there must be some central representational mechanism.

If DEC does away with the need to invoke any use of representations in cognitive science, then it would be a sure sign of an enormous conceptual change in cognitive science. Together with the notion of computation, no concept is more central to existing cognitive science than representation. The issue here is of course the "if". The examples presented by the authors mentioned above are impressive and, consequently, the strong no-representation claim had a large impact. However, only a very small minority became actually convinced that one could do without representations in cognitive science. In contrast, there have been many attempts at rebutting the strong no-representation claim (Kirsh, 1991; Clark & Toribio, 1994; Clark, 1997a, 1997b; Bechtel 1998). The counter arguments tend to converge on two main points. (a) The criticisms of representation focus on a much too narrow interpretation of representation, namely on representations as complete world-models. Subsequently, it is argued that the use of representations in cognitive science is not restricted to this particular kind of representation. (b) The critics of representation focus on lower-level processes, mainly those involving motor capacities. It may be that in some particular low-level cases the use of representations can be disbanded. However, when it comes to real, higher-level, cognitive processes—those involving thought, language and memory—representations are essential.

As a result of these often-repeated counterclaims, there is a strong tendency to conclude that the critics of representation overstate their case. Cognitive science cannot and should not do without representations. In this view, associating DEC with non-representationalism merely detracts from the genuine contributions of DEC to cognitive science. These genuine contributions do not contradict existing cognitive science ideas, but rather extend and supplement those ideas. For example, Clark states that "the real challenge lies not in the supposed implications for notions of representation and computation but in the ideas concerning the dense spatial and temporal interplay between neural, bodily and environmental factors" (1997b, p.475). Bechtel also argues that the issue of representations does not set DEC apart from other views in cognitive science: DEC "is not challenging the use of representations but is a collaborator in understanding the format of representations" (1998, p.305).

These authors subsequently conclude that DEC does not instigate an important conceptual change in cognitive science. Rather, DEC is acknowledged to be an important addition to older insights, but it does not overthrow or severely criticize cognitive science's basic assumptions. Instead, DEC and computational approaches are held to *complement* one another, without any antagonism between the two. In this view, the proclamation that a dynamic and embodied view on cognition inaugurates a major reorientation in cognitive science is too strong. The suggestion, made by both Bechtel and Clark, is that it is time to disband the revolutionary talk and to separate the scientific DEC-wheat from the rhetorical DEC-chaff.

Indeed, sober-mindedly separating wheat from chaff is an important task in any science at any time. However, I will argue that at least some of the revolutionary talk is part of the scientific DEC-wheat. More specifically, I argue that DEC definitely involves an important conceptual switch with respect to the notion of representation in cognitive science. An all out DEC-view will still contain representation-like entities. It will be possible to interpret those entities as representations. However, these representation-like entities have their origins in a very different conceptual background and they play a different role compared to cognitive science's ordinary usage of representations. The changes are important enough to warrant the claim that these entities are not representations at all but entities that require a different terminology.

The remainder of this paper consists of the following sections. First I will discuss what kind of representation cognitive science is committed to, and give a short sketch of DEC in relation to representations. In section 4, I describe how self-organization provides a general perspective of the dynamic and embodied view on cognition and how representation-like entities feature in this context. In section 5, I argue that these entities are so different from the traditional use of representation that they warrant the claim of being part of a non-representational account.

2 Traditional representation in traditional cognitive science

Representation is the central notion for traditional cognitive science. Representation is so central that large parts of cognitive science only deal with representational issues while little interest is shown into the context where cognitive systems ultimately ought to operate. DEC reacts to this negligence by stressing the importance of the dynamic bodily and environmental context, and in this way draws attention to the behavioral—the perception-action—context of cognition. The question of interest then becomes what the notion of representation actually contributes to cognitive explanations of the perception-action couplings of a cognitive system. Additionally, it becomes possible to pinpoint in more detail which characteristics of representation are essential for performing this explanatory function.

What does representation contribute to the cognitive explanation of behavior? First it brings into play our common sense, folkpsychological explanations for behavior. The notion of representation is normally used in a context involving agents. Agents, or persons, use symbols (words, pictures, signs, ciphers) to re-present things which are not immediately at hand. For example, I may use a sketch to represent the lay out of a city in order to help me going from place A to B in that city. Thus representations add a first-hand familiarity to cognitive explanations that make them insightful in an intuitive way. At the same time cognitive science is committed to developing computational models capable of generating the phenomena of the mind, such as behavioral regularities. A distinction that is important here is between personal and subpersonal levels of organization.

Dennett introduced the term subpersonal into philosophy in 1969 when he made the distinction between “the explanatory level of people and their sensations and activities and the subpersonal level of brains and events in the nervous system” (quoted in Hornsby, 2000, p.7). Of course, there are many subpersonal levels when it comes to happenings in the brain. One might look at the activity of individual neurons, small neural circuits or even more global brain processes. However, the point of using the label subpersonal is to remain close to a personal level of description. It is not meant to refer to brain processes as brain processes. A subpersonal level of organization is that level (or those levels) of organization which provides the underlying mechanism for personal level characteristics. It is the level which describes brain (and bodily and environmental) processes as the means by which personal level properties are generated. In this sense, the subpersonal level is a core topic for psychological investigation.

There is nothing new here, except maybe for the wording. Cognitive science and cognitive neuroscience have never been understood as an investigation of persons, but always as an investigation of the principles that enable us to be human. There is however something which obscures the subpersonal focus of a large part of psychology. The intriguing and often confusing move that is made within traditional cognitive science is that the personal level description that we use in our everyday explanations is used as the general blueprint for explanations at a subpersonal level of explanation. I call this use of personal-level representation-based explanations at a subpersonal level of behavioral explanation the Agent Theory, or AT for short (Keijzer, 2001).

AT consists of the hypothesis that—when viewed at a general, behavior-related level of operation—a brain (the agent’s control structure) operates basically the same as a complete agent operates (as interpreted in terms of folk psychology). This hypothesis involves the idea that the notion of representation, as used at a personal level, can also be used as a part of a description of the functioning of the brain. So the first step in making cognitive science's use of representations more explicit is by marking it as a form of *subpersonal representation*. This form of representation may not derive its representationality, its meaningfulness, from full-blown agents who use something as a representation. Whereas, I may use pieces of string, syntactically structured symbol systems or whatever as representations, at a subpersonal level representation must be formulated independently of any such ascription. To use the phrase in the philosophy of mind for this problem, meaning must be naturalized, and this has been a core problem for many years that has still not been solved in a satisfactory way (Fodor, 1987; Cummins, 1996). However, there is a way to come up with a fully subpersonal account of representation that can be used within cognitive science, even when it does not solve all the philosophical problems related with meaning. The important point to note here is that subpersonal representation as used within cognitive science becomes dependent on this particular way of acquiring meaning. Subpersonal representation is not as selfevidently available for cognitive science as it is at a personal level where the presence of intelligent agents can be assumed from the start.

The key to cognitive science's solution to the problem of subpersonal representation is to think of it as a form of internal modeling. This modeling idea came to the fore when one tried to account for the fact that a lot of behavior is oriented toward the achievement of future goals and is relatively independent from the immediate sensory environment. Something else is necessary to explain how a cognitive system achieves its goals despite ambiguous and insufficient sensor-readings. The idea of internal modeling was fundamental for cognitive science's answer to this question: By what factor can behavior be guided when this guidance is not provided by the current state of the sensors?

Someone who deserves a lot of credit in coming up with the current— artificial intelligence and cognitive science—way of dealing with this problem is Kenneth Craik. Craik stated that one of the most fundamental properties of thought is its power to predict events (Craik, 1943, p.50). He mentioned three essential processes:

> (1) 'Translation' of external process into words, numbers or other symbols,
> (2) Arrival of other symbols by a process of 'reasoning', deduction, inference, etc., and
> (3) 'Retranslation' of these symbols into external processes (as in building a bridge to design) or at least recognition of the correspondence between these symbols and external event (as in realising that a prediction is fulfilled). (ibid.)

The process of reasoning produces a final result similar to that which might have been reached by causing the actual physical processes to occur: The thought process mimics the external events and so can be used to predict these external events (on the condition that there is a time difference between the two). Thus, according to Craik, the essence of thought is that it provides a *model* of the external world.

Craik postulated that the capacity of modeling is not restricted to minds: "Surely, however, this process of prediction is not unique to minds, though no doubt it is hard to imitate the flexibility and versatility of mental prediction." (p.51). As examples he mentioned—it was 1943—a calculating machine, an anti-aircraft predictor and Kelvin's tidal predictor as mechanical parallels to the three stages of thought. Johnson-Laird described a simple robot, made by Christopher Longuet-Higgins. In the absence of direct sensor information the robot's behavior was guided by a model. Johnson-Laird called this a *Craikian automaton*.

> The robot moves freely on the surface of a table, but whenever it reaches an edge it rings an alarm bell to summon its human keeper to rescue it from disaster. It possesses neither pressure sensors for detecting edges, nor any sort of electronics. How then does it respond to the edge of the table? The answer turns— literally—on a model. As the robot travels around the table, its main wheels drive two small wheels that hold a piece of sandpaper up underneath its baseplate. The paper is the same shape as the table, and as the small wheels turn they move the paper around beneath the baseplate. In fact, their

> position on the paper at any moment corresponds exactly to the robot's position on the table. There is a ridge at the end of the paper (corresponding to the edge of the table) so that whenever one of the small wheels is deflected by it, a simple circuit is closed to ring the alarm. The large wheels in effect act as perceptual organs (as well as a means of transport) that register the robot's movement through its environment. The small wheels and the piece of sandpaper are not intrinsically a model, but they become a model in the robot because of their function as an arbitrarily selected symbolic notation that is used to register the position of the robot in the world. (Johnson-Laird, 1983, p.404)

Of course, this sandpaper robot is extremely simple. Still, it incorporates the three essential properties that Craik ascribed to thought: There is a translation process from the world to an internal representational medium, subsequently a process of prediction takes place, and finally the result is retranslated into the world where it modifies the behavior of the automaton in an adaptive way.

The example illustrates how a mind-derived conceptual framework can be realized in a fully mechanical way and turned into a subpersonal theory of the generation of behavior. The externally visible behavior of a cognitive agent is thus explained by postulating the existence of an internal model that is isomorphic to the behaviorally relevant processes happening outside the organism (Rosen, 1979, 1987). The model can subsequently act as a program that instructs the motor system so that specific, distal goals are achieved. This formulation is identical to what Brooks terms the SMPA (sensing, modeling, planning, and acting) framework (Brooks, 1999), except that Craik's middle stage is here differentiated into two separate stages. These components are thought to operate in a globally sequential way, one after the other, even when many local deviations from strict sequentiality occur.

All this ought to be familiar. The reason for going into this matter here is that it highlights the specific contribution of the notion of representation to cognitive science. Representations are often thought of as an internal covariant of an external situation or object. The rough idea is that some internal state A is considered to be a representation of external state X when the presence of X systematically leads to the presence of A through a perceptual process. Disregarding the problems that come to the fore when one tries to work out this rough idea systematically as is done within philosophy (Fodor, 1987; Cummins, 1996), the point is that this is not the way that cognitive science makes use of representations. Representation is used as part of an account that is capable of explaining behavior under the assumption that sensory information is usually insufficient and ambiguous. To actually serve this explanatory purpose, representation cannot be defined as being dependent on external circumstances. Representation needs to be derived from something else and that is being an internal model of the external situation (Cummins, 1996; Keijzer, 2001). Of course this doesn't preclude the possibility that a causal dependence between X and A may occur—on the contrary, this is what one expects when the model is adequate—however this is not sufficient for turning A into a representation.

Casting representation strictly as a model-based phenomenon disqualifies many cases that we habitually think of as representations from being actually representational as far as the explanatory purposes of cognitive science are concerned. It might be helpful for us to view covariating internal states as representations. However, this does not turn these states into subpersonal representations *for the so described system*, as long as they are not part of some internal form of modeling. Any argument into the direction of a non-representational approach in cognitive science is thus not contradicted by the obvious fact that there are many systematic relations between specific environmental states and specific internal cognitive states. Representation in cognitive science is a more specific and stronger claim than is often thought.

Another important point about the role played by representations within AT is that they are conceived of as the source of order that is present within the outward behavior of a cognitive system. The motor system is traditionally seen as a peripheral part of the total cognitive system. In this view, the motor system is no more than an executive, a translating function or mechanism, passive in itself unless brought into play to perform some planned action. The 'real' action—that is deciding and planning what is to be done—takes place within the cognitive system and leads to sequences of sets of instructions that enable the motor system to turn these instructions into actual movements. It is precisely this specific, passive interpretation of the motor system that spawned the dynamical and embodied approach.

To summarize, I will take the two aspects of being a form of internal modeling and being the source of behavioral regularities as the defining conceptual criteria of traditional representation in traditional cognitive science. When the notion of representation is criticized in the following it will be this particular interpretation of representation.

3 The organism-environment conceptualization

In contrast to AT's primary focus on the internal cognitive apparatus, a dynamical and embodied approach proposes to explain intelligent behavior as a result of the fine-grained interactions between a neural system, the particular embodiment of an agent and the characteristics of the environment. To explain the occurrence of intelligent behavior one no longer postulates that everything of interest happens in an internal cognitive system which subsequently imposes its plans on the body and consequently on the environment. The body and the environment are now themselves seen as important players in an account of intelligence, rather than as the passive conveyers of decisions made by a central internal intelligence. As a result, the internal cognitive system can be offloaded to a large extent as it doesn't need to incorporate all intelligent characteristics beforehand. These intelligent characteristics arise as a result of concrete organism-environment interactions.

The general idea of a dynamical organism-environment coupling has been described many times (for introductions of the main ideas see, for example: Brooks, 1999; Clark, 1997a; Keijzer, 2001). The specific point of interest at present is the extent to which the neural system can be offloaded from its representational characteristics within this way of conceptualizing intelligent systems. As said, there was an initial claim that the ongoing organism-environment interaction is sufficient for almost all of human behavior. However, a strong argument against this claim remains the observation that at least some aspects of human behavior require more long-term guidance than the short-term couplings between an agent and its immediate environment provide. In particular anticipatory behavior, which consists of achieving long-term goals, is insufficiently explained in terms of immediate organism-environment interaction (Clark, 1997a). Clark describes such behavior as a representation-hungry problem. In this case, behavior consists of regular sequences that are not guided from moment to moment by the ongoing organism-environment couplings. Thus, some form of internal guidance remains necessary, in addition to such short-term coupling. At this point, there seems to be a strong reason for combining a DEC approach with some form of representation. The notion of representation then helps to explain how short-term ongoing organism-environment couplings succeed in producing behavior that is oriented toward long-term goals even when the environment provides insufficient guidance.

I will not go into the reasons for being dissatisfied with this so-called hybrid approach. A discussion of such reasons can be found in Keijzer (2001). As a general observation, I just want to point out that a hybrid approach leaves the awkward personal-level origin of representation intact, while it remains unclear to what extent representational motor instructions can hand over control to dynamical coupling. Leaving aside the issue whether model-based representation, and a dynamical and embodied approach could make a good match, the point I want to make in this paper is that it is *not necessary* to make this step from representation-hungry problems to the postulation of model-based internal representations, as described in the previous section. A different kind of internal state will suffice.

4 Self-organization and 'representational steering': the two-component view

Casting intelligence, or at least intelligent behavior, as a phenomenon that arises out of a large collection of small-scaled organism-environment interactions brings it into the domain of self-organization. The notion of self-organization is a familiar one nowadays. It has become common place that certain kinds of physical, chemical and biological systems have this self-ordering capacity in which a collection of initially disordered components become mutually coordinated and come to show patterns as a collective.

It should be stressed at this point that the pattern-formation exhibited by self-organizing systems is a general physical phenomenon. Haken (1987), for example, originally formulated his account of self-organization in the context of laser-light, while the development of convection rolls in a fluid heated from below provides one of the standard cases to introduce the idea of self-organization (Haken, 1995; Kelso, 1995). Self-organization then is a general phenomenon that is not necessarily related to biological or psychological phenomena. The patterns generated simply follow general mathematical rules (Stewart, 1998).

When intelligence is to be interpreted as a form of self-organization, in what way is it different from self-organization as a general phenomenon? Intuitively there seems to be a big difference between laser-light and convection rolls on the one hand, and thought and action on the other. Something seems to be missing so far. Actually, this problem already occurs at a strict biological level as the pattern-formation occurring in living systems are usually much more complex than anything seen in non-living systems. What makes the difference between general physical forms of self-organization and self-organization as it occurs in biological and cognitive systems?

The general solution to the problem is to add an extra ingredient to the self-organizing process. This ingredient consists of a factor that 'steers' the pattern forming processes into specific directions. In biology, DNA plausibly performs this role, while in psychology and cognitive science notions like internal representations and intentions come to mind. Kelso calls this addition that makes biological and cognitive self-organization different from self-organization in general "the second cornerstone of biological self-organization" (1995, p.137-138), self-organization itself being the first one. This "second cornerstone" is most easily introduced for the biological case, which will be done in the remainder of this section. In the following section, I will turn to the issue how this way of steering bears on the issue of representations in cognitive systems.

At first sight, invoking DNA as a steering factor in biological systems seems to contradict the importance of self-organization there. Particularly in the popular media, DNA is often staged as a complete set of master instructions that specify the organization of a living system (Keller, 2000). The latter is then seen as a direct derivative of the instructions present in the DNA. In this image, DNA acts as a controlling agent that actively forces an essentially passive and compliant medium (consisting of intra and extra-cellular processes) into specific patterns. However, this general view of DNA as the sole origin of biological order is increasingly challenged by a view that stresses the additional importance of self-organizing processes in biological systems (Kaufmann, 1993; Goodwin, 1994; Maynard Smith & Szathmáry, 1999; Stewart, 1998). Stewart puts it as follows:

> The complexity of an adult organism, such as a tiger, exceeds that of its DNA by virtually any sensible measure. The wiring diagram for a tiger's nervous system alone is more complex, by several orders of magnitude, than the tiger's entire DNA sequence. That sequence, complex as it is, versatile as the contingency plans may be does not contain enough information to specify how to build a tiger's brain, let alone an entire tiger. … The missing information is supplied by the mathematical rules (the laws of physics) that govern the behavior of matter. (1998, p. 13)

Although there is definitely no agreement about the relative importance of the two components—self-organization and genetic instructions—it is now increasingly clear that both are essential. Most important for present purposes is the shift in interpretation this brings about when it comes to the functioning of DNA in biological systems.

The notion of steering takes on a different meaning when it involves a self-organizing process. Steering is not a matter of imposing order onto a system where there is no order to begin with. The system that is being steered is not passive matter waiting to be molded into shape. On the contrary, the steered medium provides itself the pattern forming process. Steering is much less direct and rather takes the form of modulating a set of ongoing processes. When a system is in the position to develop into two different directions at a certain moment, steering consists of nothing more than to give a little push into one direction rather than another.

There are two concepts that are important to make this idea of steering a self-organizing process more concrete: the order parameter and the control parameter. The order parameter represents the large-scale (macroscopic) order brought about by the interactions between a number of parts at a smaller (microscopic) scale (Haken, 1987). In turn, the macroscopic order parameter constrains or "enslaves" the behavior of the microscopic parts. Whether an order parameter arises or what form it takes depends on the state of the system as described by a set of parameters. Some of these parameters act as control parameters. That is, by varying these parameters, the system as a whole can be shifted from a disordered state into an ordered state, or shifted from one macroscopic pattern into another. Going back to the example of convection rolls above, changing the temperature difference between the bottom and top of the fluid layer is a way to shift the system from a disordered macroscopic state into one which exhibits large-scale convection rolls (and back again). This makes the temperature difference a control parameter for this particular form of self-organization. Manipulating the control parameter offers a way to manipulate the order parameter.

In the general physical occurrence of self-organization, such as in the example of the fluid rolls, the presence and the value of the order parameter is a matter of chance as far as the order producing system is concerned. Whether or not someone puts on the fire beneath a pan filled with oil is not for the oil to decide. However, the situation is different in biological forms of self-organization where the genetic factor comes into play. Using the language of order and control parameters, one can hold that a biological system is capable of manipulating its own control parameters, turning them on and off in order to achieve specific macroscopic states that can be described by low-dimensional order parameters. As Kelso puts it, in biological self-organization there are "specific parametric influences" are at work (Kelso, 1995, p.138).

Genes can then be interpreted as a set of control parameters that are stored inside the biological system on a long-term basis. Genes provide the specific parametric influences that are required to modulate self-organizing processes in such a way that they develop into the many different structures and shapes of living systems. Following Meijer and Bongaardt (1996, see also Keijzer, 2001), I will call such a control parameter an *internal control parameter*, or ICP for short. Thus, physical and biological self-organization can be said to differ in the sense that only biological systems manipulate their own control parameters, and in this way the order parameters that make up the large-scale structure of biological systems.

The link between genes and ICPs brings to the fore another aspect of ICPs before they can act as the steering factor of biological self-organization. These ICPs must themselves be steered by other factors. In simple cases, an inactive ICP can be activated by some other external and essentially arbitrary parameter. For example, Meijer and Bongaardt (1996) use the effect of cornstarch on the viscosity of soup as an example of an internal control parameter. In this example the temperature of the soup triggers the internal parametric influence of the cornstarch. When the soup starts to boil, it becomes viscous. The drawback of this setup for biological self-organization is that this process is still dependent on an unregulated external force; someone who wants to make soup, who puts in the cornstarch, and who turns on the fire beneath the pan. In biological systems the triggering of genes that could act as ICPs is itself under the control of the living system. There is a complicated loop-like system in which organismal factors influence in many ways the transcription of structural genes that actually code for specific proteins (Keller, 2000). The whole takes the form of a so-called genetic regulatory network (Kauffman, 1993) consisting of widely dispersed and reciprocal influences between genes, proteins and metabolic processes. As Kauffman argues, such networks are themselves to be interpreted as a self-organizing process that are very robust to change and very flexible in overcoming disturbances.

There are two aspects to this way of interpreting biological cases of self-organization that are particularly important when this interpretation is also applied to psychological systems. First, it must be stressed that ICPs are a part of the microscopy. Self-organization implies a distinction between a microscopy and a macroscopy as two different levels of aggregation and description. The microscopy self-organizes into a macroscopic pattern as described by an order parameter. ICPs are factors that happen to initiate and maintain particular macroscopic patterns and in this sense are closely linked to the macroscopy. However, to act as control parameters ICPs must have an existence prior to and independent from the macroscopic patterns they help to bring about. Without such independence, new instances of a particular biological form could never be generated in this way. Normal control parameters remain outside the range of the microscopy-macroscopy system and are thus by default independent from the macroscopic order. For example, the heat gradient that results in convection rolls in a fluid exists independent from these rolls. ICPs however, must be part of this two-level system, and the only place where they can function as a parameter independent from the macroscopy is within the microscopy.

In addition, these internal control parameters must also maintain a certain independence from the microscopic processes to which they belong. When ICPs are to act as a switching device, initiating and stopping specific biological self-organizing processes, they cannot continuously take part in the ongoing interactions between the microscopic elements. Their continuous physical presence within the microscopy must not lead to a continuous specific parametric influence. This influence has to be a variable, sometimes present sometimes not. An ICP must in this sense be buffered from the microscopy at large, while its parametric influence has to be triggered by some other influence.

All of this corresponds in a general way with current knowledge about the operation of genes in metabolic processes and morphogenesis. DNA, the material embodiment of genes, situates the genetic factor firmly within the biomolecular operation of living systems. At the same time DNA remains relatively sheltered from the immediate ongoing biomolecular processes and influences these processes indirectly by coding for particular proteins that do the actual work. Also, the specific codings present within the DNA are derived from long-term evolutionary selection and in this sense are independent from the short-term ongoing, biomolecular processes.

To summarize, bringing self-organization to bear on biological systems thus involves thinking in terms of and teasing out complex regulatory networks with their self-organizing properties. There is ‘plain’ physical self-organization such as in convection rolls. And there is the biological steering of this self-organization by specific parametric influences in the form of ICPs. This self-steering involves a mesh-like system of multiple, parallel and reciprocal influences, involving regulatory and structural genes, gene products, and factors deriving from the large-scale build up produced by all this activity. How to make sense of such a mesh-like organization? This brings us back to the question whether the notion of representation plays a necessary role in this context.

5 Non-representational internal states as a conceptual option for DEC

The existence of representation-hungry problems for cognitive systems is a forceful reason for thinking that it is necessary to combine a dynamic and embodied view on cognition with representations. How does the discussion on biological self-organization bear on this claim for the necessity of representations within DEC? In this final section, I will argue that the biological case shows that it is *possible* to envision a long-term steering mechanism for cognition that is not based on the notion of subpersonal representations. Internal states—such as biological internal control parameters—could guide short-term organism-environment couplings over longer time scales into specific directions. These internal states would perform a function that is equivalent to the one usually performed by representations in ordinary cognitive accounts. However, at the same time, these internal states are so different from the cognitive science’s traditional notion of representation that it is perfectly reasonable to hold that they are not representations. Thus the seemingly inescapable connection between representation-hungry problems and the invocation of representations within DEC can be cut.

Before presenting my reasons for thinking that ICPs are not representations in a *cognitive* context, I will give a short sketch how the ICP story falls in with a tendency in *biology* away from a symbolical interpretation of genes and DNA.[1] Assuming for the present that genes can be cast in the role of ICPs, what message can be drawn from biology concerning the question whether genes (and in this roundabout way ICPs) are to be understood as the biological equivalent of representations? Since the rise of genetics in the first half of the twentieth century, the dominant metaphor for the genome has been that of a blueprint that provides all the necessary information for the construction of an organism (Savageau, 1998). In this metaphor, the rest of living systems becomes a mere derivative of the genetic information.

However, living in a time that the Human Genome Project is finished and we have a preliminary list of all human genetic information stored within our DNA, it has also become clearer than ever before that we hardly have any general understanding how genes are supposed to produce their large-scale effects within the human body. There is no unidirectional flow of control, starting with the genes and then outward into the large-scale structures of the body. Genes act as components within a complex network of regulatory factors, involving initiation, modulation and feedback loops at and between different levels of organization. Obviously, genes remain important here, but they definitely lose their larger than life character as the sole controllers of living systems (Keller, 2000). The challenge now is to decipher the operation of such regulatory networks in which genes partake.

How does the old notion or metaphor that genes are a kind of symbols fit in with this challenge? Opinions are divided here. The idea that genes are a symbolic description is still widely dispersed. At the same time, there is an important move away from this interpretation (Goodwin, 1994; Keller, 2000; Schaffner, 1998). In this view, a symbolic interpretation of genes does not help to think about genetic regulatory circuits in terms of genetic regulatory circuits. Rather, it imposes a traditional frame of mind onto this kind of organization. As Savageau puts it: “The common metaphor of the genome as a blueprint for construction of the organism masks the difficult task of relating structure and function of the intact organism to its underlying genetic determinants.” (1998, p.55). A symbolic interpretation of genes highlights the genes and their presumed phenotypic effects and downplays the contribution of the intermediate network of regulatory processes that connects the two. This interpretation can easily become counterproductive as it obscures the kind of interaction involved in relating a genome to a whole organism. Under these circumstances it seems unhelpful to see the genome as an instance of a symbol system as we know them from personal-level experience and from computer science. Rather than seeing the problem as one of filling a gap between genes and their effects, new concepts and tools seem to be required that deal directly with the complexity of the complete regulatory networks involved (Bell, 1999; Gibbs, 2001; Kauffman, 1993). It is exactly in this context that the notion of ICPs and biological self-organization become relevant.

So far, this discussion within biology is far from settled. Nevertheless, for present purposes, it suffices to point out that here is a context where a seemingly self-evident interpretation as a symbol system is actively challenged, while new conceptualizations of genetic functioning are sought capable of overcoming the shortcomings of the blueprint idea of genes. Genes are not necessarily a biological equivalent of cognitive representations even though their importance within the genetic regulatory networks is not denied. There is an openendedness to the situation in which it is unsettled whether genes are to be so interpreted or not. The central message here for DEC is that even when it is deemed useful to insert internal states like ICPs, this insertion implies in no way a direct commitment to internal representation.

The biological case makes it plausible in a general sense that accepting ICPs does not force a commitment to representations. How is the situation in the cognitive context? Presupposing a commitment to DEC, does adding ICPs to this framework amount to combining DEC with a representational framework? Before I will argue that the answer is not necessarily yes but plausibly no, it will prove beneficial to reiterate which particular kind of representation is at stake here. The traditional cognitive science way of having representation at a subpersonal level is by casting it as a form of internal modeling. Because some internal states are in an abstract and systematic way isomorphic to certain aspects of the external environment, they act as a model and thus count as representations. In addition, ‘real’ representation is dissociated from the uncountable spurious instantiations of isomorphism within the natural world by adding the requirement of a user—a cognitive system—that uses the isomorphic states as a guide for its behavior. However, despite invoking a user, the *representationality* of model-based representations does not derive from this use, but from the isomorphism that makes it useful for the user. The user is here only brought into play to select instantiations of representations proper from an overabundance of potential ones. In this view, representations have their own prior existence, while the user acts as an intermediate between the representation and the represented.

This is exactly the general intuition that lies behind traditional cognitive science with its stress on knowledge and knowledge representation and the view on motor control and behavior as a by-product. Given this particular interpretation of (modeling-based and subpersonal) representation, do the ICPs of steered self-organization fit into its mold?

I hold that the obvious answer at this point is no. The material setup that enables internal control parameters to take over the long-term guiding role of ordinary representations differs significantly from that of traditional cognitive representations. ICPs do not have their own independent representational status. The starting point consists of a set of self-organizing processes. ICPs come only into play in a secondary role as parameters that modulate self-organizing processes. They have no status, leave alone meaning, outside this context. Their only significance lies in a capacity to influence the self-organizing processes in repeatable ways. Compared to a representational account, the whole set up of the explanation is reversed. In a representational account, representation forms the kernel of the explanation, the starting place for outward intelligent behavior. Anything which happens in between is an obstacle that has to be overcome in order to produce intelligent behavior. In DEC the obstacles are turned into the resources—dynamical organism-environment couplings—that enable intelligent behavior to arise. ICPs modulate this inherent order-producing capacity, but they are not the source of this order. This sets up an essential difference between the ways in which representations and ICPs acquire their cognitive relevance. I take this difference to be sufficient to claim that an account that deals with representation-hungry problems by incorporating internal states like ICPs is plausible cast as a non-representational account. The seemingly inevitable combination of DEC with representation can be avoided in this way.

Before finalizing that last conclusion, it is proper to return to the pronouncement that it is obvious that ICPs are non-representational. At that point, many people will probably react with something like: “Even waving that improper use of the word ‘obvious’, I do not see that sufficient reasons have been given for such a momentous claim as a denial of representations!” My reply to this line of thought consists of providing more context which should make clear that the present claim is hardly as momentous as one might think, but nevertheless sufficiently important to make an issue out of it. The discussion will be centered around two questions. First, has sufficient proof been given to warrant the claim that ICPs are non-representational? I argue that the point made here is only that a non-representational interpretation is possible and plausible, not that a representational one is impossible. It is important to consider where the burden of proof lies in this context. The second question falls in with this answer: Why would one go to all the trouble of a non-representational interpretation if a representational one remains an option as well?

Has sufficient proof been given to warrant a non-representational interpretation of ICPs? If the issue had not been about representations but a more innocuous set of concepts, such as batteries and petrol tanks, I don’t think that many eyebrows would have been raised. Batteries and petrol tanks are equivalent in the general sense that both can be used as a car’s energy store. However, that identical function is performed in a different way, being based on different physical and chemical properties and related to a different kind of car engine. When it comes to making sense of the functioning of either petrol tanks or batteries in the operation of a car, it is obviously a good procedure to make a distinction between the two. The argument given above to claim that ICPs are not representations is closely analogous to the one behind the battery and petrol tank distinction. There are a number of structural characteristics connected to both concepts and these diverge sufficiently to claim a difference between the two, despite the general functional similarities. When the reasoning in the car engine case can be considered a straightforward pragmatic decision, why would the decision be different for the representational case?

My guess, for what it is worth, is that the issue whether or not to insert representations in some cognitive explanations is usually not treated as a relatively neutral theoretical decision, strictly aimed to improve our understanding of some particular aspects of cognitive functioning. Representation is also an important concept within our current, general view of the mind. Any criticism of its possible usefulness in cognitive science then easily becomes interpreted as an attack on the mind itself, a step towards the reduction or even elimination of mental concepts from a mature cognitive science. In the face of such a reduction or elimination, many philosophers and scientists would put up a fierce battle to try to avoid such a portentous conclusion. When seen in this light, a denial of the representationality of ICPs is to be treated with the utmost suspicion, and only taken into consideration if there was no single loophole left to avoid that conclusion. Seen from this perspective, no doubt one finds the claim that ICPs are non-representational is heavily under-defended and not to be accepted.

However, this is certainly not a perspective that I can share. As this topic is huge in its own right, I will give only a listing of reasons for a more light-hearted attitude in the case for a non-representational interpretation of internal states like ICPs. First, the argument is about subpersonal representation. The discussion does not address the personal-level issues of the mind, and there is no denial involved of our own, personal-level use of representations. Second, the argument does not deny the possibility of subpersonal representations in *addition* to non-representational ICPs. It only claims the possibility of the latter, not their universality at all relevant, cognitive subpersonal levels. Third, the argument targets specifically modeling-based representation, as this is arguably the basic concept behind most of current cognitive science's explanations. The argument can be confounded by shifting the concept of representation to one of its many other interpretations, but then it is no longer the cornerstone of current cognitive explanations that is at stake. Fourth, ICPs may turn out to be representations for other reasons besides being a particular ICP—even being a modeling-based representation[2]—without interfering with the non-representational interpretation of ICPs. Their potential representationality does not bear on their operation as an ICP. That operation is strictly formulated in terms of an influence on a process of self-organization, and only the consequences of that influence—whether or not it results in anticipatory behavior—decide on the appropriateness of an ICP. Fifth and final, the argument sets up only a *conceptual* possibility that enables thinking about the empirical investigation of the non-representational steering of cognitive systems in order to deal with representation-hungry problems. A representational interpretation of ICPs amounts to the extra claim of a 'direct', symbolizing connection between an ICP and some aspect of the non-local (distal) environment. A non-representational interpretation of ICPs only claims their influence on self-organizing processes occurring within a dynamic and embodied cognitive system, which results in behavior that is oriented toward non-local features of the environment. Empirically, it may turn out to be that performing this influencing function requires of ICPs that they have also an additional representational 'shortcut' to the non-local environment. However, this is an empirical issue, it should not be decided on beforehand by equating ICPs with representations.

In conclusion, the present discussion about ICPs must be dissociated from the philosophical discussion about the status of mental concepts in the natural world. The discussion about ICPs is not about minds but about subpersonal principles that are at work behind mental phenomena. The presence of representations is not a default assumption that must be actively disproved. It suffices to note that ICPs and modeling-based representations are different concepts to dissociate them without much ado. Reifying representations in the form of ICPs would only make representation an arbitrary plastic and empty feel-good concept. Also, while accepting this conceptual distinction, an additional representational interpretation of ICPs may become empirically necessary in the end, but only if it turns out to be explanatory useful, and not because we do not allow ourselves to even consider other options.

These considerations lead to the answer to the second question posed above: Why argue explicitly for a non-representational interpretation of ICPs if a representational one remains possible and is generally held to be much more acceptable? Why not argue that ICPs form a different kind of representation? For example, Wheeler and Clark (1999) talk about *genic representation* in a very similar context, while Bickhard (1998) invokes the notion of *interactive representations*. These theorists also share a general commitment to the principles of DEC but have no problems with casting their ideas as a different kind of representation. Why antagonize the cognitive science community by targeting the word representation itself? In addition to the reasons stated above, I want to mention a final heuristic issue that is at stake.

The concept of representation invites thinking in terms of long-distance connections between internal states and non-local environmental events. It is a concept that instigates a general, abstracted understanding of a the relation between a cognitive system and the actions that stem from it by skipping the intermediate details. DEC however is based on the hard won lesson that the intermediate details are the key factor in the processes that make us intelligent: Intelligence is based on the many, reciprocal, dynamical and physical couplings between organism and environment. Representation is not a neutral word, but exemplifies a way of thinking that is in direct opposition to DEC. Naming ICPs a form of representation would highlight their long-term behavioral relevance rather than their immediate control parameter functioning within the messy self-organized details out of which behavioral patterns arise. This is exactly the wrong message from the perspective of DEC. The notion of representation acts as a disturbing conceptual attractor that draws one's thinking into paths that DEC explicitly tries to avoid. To withstand the pull, it is wise to maintain a safe distance.

Notes

1 For much more thorough discussions of this topic see e.g. Keller (2000), Schaffner (1998), and Sterelny and Griffiths (1999).

2 I thank Virgil Whitmyer for making me think about the implications of this possibility.

References

Bechtel, W. (1998). Representations and cognitive explanations: assessing the dynamicist's challenge in cognitive science. *Cognitive Science, 22*, 295-318.

Beer, R.D. (1995b). Computational and dynamical languages for autonomous agents. In R.E. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 121-147). Cambridge, MA: MIT Press.

Bell, A.J. (1999). Levels and loops: the future of artificial intelligence and neuroscience. *Philosophical Transaction of the Royal Society in London B, 354*, 2013-2020

Bickhard, M.H. (1998). Robots and representations. In R. Pfeifer, B. Blumberg, J-A Meyer & S.W. Wilson (Eds.), *From animals to animats 5* (pp. 58-63). Cambridge, MA: MIT Press.

Brooks, R.A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, 1*, 14-23.

Brooks, R.A. (1991). Intelligence without reason. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 569-595). San Mateo: Morgan-Kaufman.

Brooks, R.A. (1999). *Cambrian intelligence*. Cambridge, MA: MIT Press.

Clark, A. (1997a). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press

Clark, A. (1997b). The dynamical challenge. *Cognitive Science, 21*, 461-481.

Clark, A. & Toribio, J. (1994) Doing without representing? *Synthese, 101*, 401-431.

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Cummins, R. (1996). *Representations, targets and attitudes*. Cambridge, MA: MIT Press.

Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Gibbs, W.W. (2001). Cybernetic cells. *Scientific American, 265(2)*, 42-47.

Goodwin, B.C. (1994). *How the leopard changed its spots: The evolution of complexity*. New York: Scribner's Sons.

Haken, H. (1987). Synergetics: an approach to self-organization. In F.E. Yates (Ed.), *Self-organizing systems: The emergence of order* (pp. 417-437). New York: Plenum Press.

Haken, H. (1995). Some basic concepts of synergetics with respect to multistability in perception, phase transitions and formation of meaning. In M. Stadler & P. Kruse (Eds.), *Ambiguity in mind and nature*. Berlin: Springer Verlag.

Hornsby, J. (2000). Personal and subpersonal: a defence of Dennett's early distinction. *Philosophical Explorations, 3*, 6-24.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Kauffman, S.A. (1993). *The origins of order: Self-organization and selection in evolution*. New York: Oxford University Press.

Keijzer, F.A. (2001). *Representation and behavior*. Cambridge, MA: MIT Press.
Keller, E.F. (2000). *The century of the gene*. Cambridge, MA: Harvard University Press.
Kelso, J.A.S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
Kirsh, D. (1991) Today the earwig, tomorrow man? *Artificial Intelligence, 47,* 161-184.
Maynard Smith, J. & Szathmáry, E. (1999). *The origins of life*. Oxford: Oxford University Press.
Meijer, O.G. & Bongaardt, R. (1996). The maizena argument: a reaction. *Journal for Ecological Psychology, 7*, 285-290.
Rosen, R. (1979). Anticipatory systems in retrospect and prospect. *General Systems Yearbook, 24*, 11-23.
Rosen, R, (1987). On complex systems. *European Journal of Operational Research, 30,* 129-134.
Schaffner, K.F. (1998). Genes, behavior, and developmental emergentism: One process, indivisible? *Philosophy of Science, 65*, 209-252.
Savageau, M.A. (1998). Rules for the evolution of gene circuitry. *Pacific Symposium on Biocomputing, 3*, 54-65.
Sterelny, K. & Griffiths, P.E. (1999). Sex and death: An introduction to philosophy of biology. Chicago: Chicago University Press.
Stewart, I. (1998). *Life's other secret*. Harmondsworth: Penguin Press.
Thelen, E. & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
Van Gelder, T. (1995) What might cognition be if not computation? *Journal of Philosophy, 91*, 345-381.
Van Gelder, T. (1998) The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21*, 615-665.
Wheeler, M. & Clark, A. (1999). Genic representation: reconciling content and causal complexity. *British Journal for the Philosophy of Science, 50*, 103-135.