PHILOSOPHICAL PSYCHOLOGY, VOL. 16, NO. 1, 2003

Representationalism vs. anti-representationalism: a debate for the sake of appearance

PIM HASELAGER, ANDRÉ DE GROOT & HANS VAN RAPPARD

ABSTRACT *In recent years the cognitive science community has witnessed the rise of a new, dynamical approach to cognition. This approach entails a framework in which cognition and behavior are taken to result from complex dynamical interactions between brain, body, and environment. The advent of the dynamical approach is grounded in a dissatisfaction with the classical computational view of cognition. A particularly strong claim has been that cognitive systems do not rely on internal representations and computations. Focusing on this claim, we take as a starting point a question recently raised by Cliff and Noble: "… if evolution did produce a design that used internal representations, how would we recognize it?" (Knowledge-based vision and simple visual machines,* Philosophical Transactions of the Royal Society: Biological Sciences, 352, *1165–1175, 1997). We will argue that cognitive science lacks a proper operationalization of the notion of representation, and therefore is unable to fruitfully discuss whether a particular system has representations or not. A basic method to detect representations in a physical system, grounded in isomorphism, turns out to be quite unconstrained. We will look at a practical example of this problem by examining the debate on whether or not van Gelder's (What might cognition be, if not computation?* Journal of Philosophy, 92, *345–381, 1995) controversial example of the Watt Governor is representational. We will conclude that cognitive science, as of yet, has no empirically applicable means to answer Cliff and Noble's question unequivocally. This makes the recent representationalism vs. anti-representationalism debate a debate for the sake of appearance.*

1. Introduction

The dynamical approach to cognition has gained a growing popularity within cognitive science over the last decade (e.g. Beer, 1995a,b; Bickhard & Terveen, 1995; Brooks, 1991, 1995; Clancey, 1997; Clark, 1997a; Hendrik-Jansen, 1996; Kelso, 1995; Keijzer, 2001; Port & van Gelder, 1995; Thelen & Smith, 1994; van Gelder, 1995, 1998; van Gelder & Port, 1995; van Rooij *et al.*, 2002; Varela *et al.*,

Pim Haselager, Department of Cognitive Science, NICI, Katholieke Universiteit Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands, email: w.haselager@nici.kun.nl; André de Groot, Theoretical Psychology, Vrije Universiteit de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, email: ad.de.groot@psy.vu.nl; Hans van Rappard, Theoretical Psychology, Vrije Universiteit de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, email: jfh.van.rappard@psy.vu.nl

ISSN 0951-5089/print/ISSN 1465-394X/online/03/010005–20 © 2003 Taylor & Francis Ltd
DOI: 10.1080/0951508032000067761

1991). Proponents of this approach claim that cognitive systems should be studied as brain processes in continuous dynamical interaction with the body and the immediate environment.

First, this claim highlights the fact that our cognitive system is essentially grounded in the brain as it is integrated with our body. Much of the nature of our cognitive processes is determined by the specific action possibilities afforded by our body. That is, our cognitive system is not an independent information-processing system, but a system that is essentially *embodied*. Second, the dynamical approach emphasizes the fact that cognition and behavior are the result of the interaction between our cognitive system and its immediate environment. As such, it stresses the fact that cognition and behavior are *embedded* in a causal web that encompasses brain, body, and environment. In all, the dynamical approach entails a framework in which cognition and behavior are taken to result from complex dynamical interactions between brain, body, and environment.

The advent of the dynamical approach coincides with a growing discontent with the prevailing approaches to cognition. In particular, there is serious doubt about, and sometimes strong dissatisfaction with, the traditional view that cognition should be defined in terms of computational transformations over internal representations [1]. In this paper, we do *not* want to formulate a position on whether representationalists or anti-representationalists are correct. Instead, we focus on the question whether the notion of representation is operationalized well enough in order to allow a fruitful debate between the two sides.

The starting point of our investigation is a question recently raised by Cliff and Noble (1997, p. 1170): "if evolution did produce a design that used internal representations, how would we recognize it?" In our view this is an important question, because the assumption that cognition consists of the computational manipulation of representations is a basic one in cognitive science. This presupposes that there must be a principled answer to Cliff and Noble's question. The problem, we suspect, has become more urgent because of the developments in cognitive modeling over the last decades. In the "early days," one would write a program, i.e. one specified the representational structures and their potential transformations, and then let it run on a machine that was designed to faithfully mirror the specified state-transitions on a physical level. A direct mapping between the computational and representational description and the physical-causal processes taking place within the system is a given under those circumstances. One *starts* with representations and computations and assures (or is assured) that the machine implements them faithfully. But the situation confronted by Cliff and Noble is a radically reversed one. Cliff and Noble, as many researchers working with genetic algorithms or autonomous robots, find themselves in the situation of having to examine a physical system that has learned to perform certain tasks more or less independently from a precise representational pre-programming. As they (1997, p. 1165) say:

> We discuss simple machine vision systems developed by artificial volution rather than traditional engineering design techniques, and note that the task of identifying internal representations is made difficult by the lack of an operational definition of representation at the causal mechanistic level.

> Consequently, we question the nature and indeed the existence of representations posited to be used within natural vision systems (i.e. animals).

The question we take Cliff and Noble to raise is how, without a proper operationalization, the notion of representation is going to help cognitive scientists such as themselves in adequately capturing the inner complexities that supposedly enable the system to behave as it does.

In Section 2, we will examine the standard practice within cognitive science to interpret representation in terms of a "stand-in" that is used by a system to guide its behavior. In Section 3, we will turn to a notion that plays an important role in the attempt to determine representations within a system: that of isomorphism. Isomorphism implies a structure-preserving relation between the physical-causal make-up of the system and the formal structure of the computational model supposedly instantiated by the system (e.g. Block, 1995; Chalmers, 1995; Chrisley, 1995; Kim, 1996). We will find that isomorphism is a relatively unconstrained notion, allowing almost any physical system to qualify as a representational system. In Section 4, we will examine the discussion between van Gelder (1995) and Bechtel (1998) about whether or not a relatively simple system, the Watt Governor, is a representational system. Our aim here is not to take sides, but to focus on the role of isomorphism in the debate. Finally, in Sections 5–6, we will conclude that cognitive science lacks the "firm, preferably operational, definition of representation" Cliff and Noble (1997, p. 1170) need in order to be able to decide whether a particular physical system actually operates on the basis of representations or not. We will argue that this situation is problematic for cognitive science in general and precludes a fruitful debate between representationalists and anti-representationalists.

2. The dynamical approach and the notion of representation

According to some authors, the dynamical approach should not be viewed as a mere extension of the traditional computational approach to cognition, but as viable replacement of it. Specifically, these authors object to the central roles representation and computation play in the study of cognition. As van Gelder (1998, p. 662) says:

> A small but influential contingent of dynamicists have found the notion of representation to be dispensable or even a hindrance for their particular purposes.

The interpretation of representation referred to here (as in most other cases) is that of an internal state standing in for some other state. An oft-cited (e.g. Clark, 1997a; Clark & Toribio, 1994; van Gelder, 1995) and more detailed characterization in this respect is the one formulated by Haugeland (1991, p. 62):

> A sophisticated system (organism) designed (evolved) to maximize some end (e.g., survival) must in general adjust its behavior to specific features, structures, or configurations of its environment in ways that could not have been fully prearranged in its design ... But if the relevant features are not

> always present (detectable), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*; that which it stands for is its *content*; and its standing in for that content is *representing* it. (italics in original) [2]

Two central features of this interpretation are that an internal representation is a *stand-in* for another (external) state, and that a representation has the *causal power* to guide behavior, so that an organism or system can use representations in order to adjust its behavior.

We consider the issue of representation from a *realist* perspective. That is, "representation" refers to an identifiable physical state within a system that stands in for another (internal or external) state and that *as such* plays a causal role in (or is used by) the system generating its behavior. Thus, the question we take Cliff and Noble to be raising is whether the notion of representation is operationalized well enough to allow them to establish whether representations are really there within the system, doing their causal work. This perspective has to be distinguished from views according to which explanatory or predictive utility is sufficient to warrant representational claims about systems (e.g. Dennett, 1991). We are concerned with the methods cognitive science has at its disposal to establish the *actual existence* of representations within a system, not with the question of whether or not it is useful to *talk* about representations.

3. Isomorphism

The standard method to detect representation and computation within cognitive science is to find an isomorphism between entities and processes at functional and physical levels of analysis (e.g. Block, 1995, p. 399; Chalmers, 1995, pp. 393–394; Kim, 1996, p. 89; Searle, 1992, pp. 207–212). A physical system is thought to realize a particular computational model when there is a one-to-one mapping between the computational states and their formal structure on the one hand and the physical states and their physical-causal structure on the other hand. As Kim (1996, p. 89) says:

> [T]here are real physical (perhaps neurobiological) states in [physical subject] S, $<Q_0, Q_1, \ldots, Q_n>$, corresponding to [computational system] M1's internal states $<q_0, q_1, \ldots, q_n>$, and these Q's are causally hooked up to each other … in a way that ensures that, for all computational processes generated by M1, isomorphic causal processes occur in S. As we may say, S is a "causal isomorph" of M1.

Isomorphism in this sense can be interpreted as answering Cliff and Noble's (1997) question regarding the identification of representations by providing a criterion for ascribing representational status to a system. Representations can be identified as those physical states that correspond one-to-one with the content carrying states outlined on the computational level and that have the causal connections to match the computational transformations. So, if at the computational level representational states *A*, *B*, and *C* are described, and the overall formal structure of the model is isomorphic with the physical-causal structure of physical system *P*, then the physical states (let's say *a*, *b*, and *c*) that correspond one-to-one with the formal states *A*, *B*, and *C* are the physical realizations of *A*, *B*, and *C*, and hence the physical states *a*, *b*, and *c are* the representations at the physical level.

3.1. Isomorphism and universal realizability

Putnam (1988) has severely criticized the usefulness of isomorphism. According to his universal realizability thesis, every open physical system realizes every computation (see also Searle, 1992, pp. 207–212, who makes a similar claim). Therefore, Putnam concludes, the notions of computation and representation are of no use as a foundation for the study of cognition.

The proof Putnam presents for his thesis consists in showing that for any arbitrary physical system and any arbitrary computational model it is possible to give a description of the physical system (in terms of its physical-causal structure) that maps one-to-one with the computational transitions specified at the computational level of description. The particular case Putnam presents as proof for his theorem is a finite-state automaton (FSA), which, during a given time period, goes through the consecutive transitional (computational) states "*ABABABA*." Given this FSA, he shows that during the same time period the overall state of an arbitrary physical system can be described in terms of seven consecutive physical states numbered 1–7. To see how these physical states map one-to-one with the seven computational states the FSA goes through, Putnam sets up two disjunctive sets of physical states which each correspond with the two different types of states the FSA passes through (one set for state *A* and one set for state *B*). Next, he defines two mapping relations such that (1) if the physical system is in state 1 or 3 or 5 or 7 (call this disjunctive set *a*) there is a direct mapping with computational state *A* ($f(a) = A$), and (2) if the physical system is in state 2 or 4 or 6 (call this disjunctive set *b*) there is a direct mapping with computational state *B* ($f(b) = B$). The overall point Putnam wants to make is that under a particular description of a physical system it is possible to generate a situation in which this description corresponds one-to-one with the consecutive states described at the computational level. In addition, Putnam claims that since the example he has offered holds for every arbitrary physical system and every arbitrary computational model, the conclusion can be drawn that computation is universally realizable. Any statement to the effect that cognitive processes are computational and representational (and that our brain realizes particular computational models) is vacuous since such a statement holds for every arbitrary physical system. As a result, Putnam claims, the notions of computation and representation cannot play a foundational role in the study of cognition.

3.2. Criticisms of the universal realizability claim

Chalmers (1994, 1995, 1996) and Chrisley (1995) have rejected Putnam's (1988) claim that computation is universally realizable. The two main arguments against Putnam are that he relies on an interpretation of causality that is too weak, and that he employs a computational formalism that is not complex enough to characterize the structure of cognition.

Regarding the first point of criticism, Chalmers (1995) claims that the physical realization of a computation depends not only on a one-to-one mapping between computational states and physical states. What is needed, in addition, is that the relations between the individuated physical states satisfy "strong" conditionals (i.e. the relations must be reliable or lawful). That is, given a formal transition from computational state *A* to computational state *B*, it must be the case that *if* the system were in physical state *a*, state *b would have to* follow. This modal (counterfactual supporting) characteristic of state-transitions at the physical level of description should hold for all transitions specified at the computational level. Both Chalmers (1995) and Chrisley (1996) claim that the state-transition conditionals satisfied in Putnam's example are not strong enough. As Chalmers (1995, p. 313) says:

> if environmental conditions had been slightly different, the system's behavior would have been quite different. Putnam's construction establishes nothing about the system's behavior under such circumstances. The construction is entirely specific to the environmental conditions as they were during the time-period in question. It follows that his construction does not satisfy the relevant strong conditionals.

Chrisley (1996, p. 410; see also Searle, 1992, p. 209) adds to this that Putnam's interpretation of causality does not distinguish causally related events from contiguous events. It is only the former interpretation that incorporates modal force and the support of counterfactuals.

Regarding the second issue—the adequacy of the computational formalism employed by Putnam—Chalmers argues that this formalism is inappropriate for the study of cognition. In particular, the formalism of a finite-state automaton lacks the internal structural complexity needed in order to describe and specify the human mind. In Chalmers' view, a much more suitable formalism is one that has combinatorial structure—such as the combinatorial state automaton (CSA)—which, as Chalmers (1995, p. 394) explains:

> [only] differs from an FSA in that an internal state is specified not by a monadic, [single] label S, but by a vector (S^1, S^2, S^3, ...). The elements of this vector can be thought of as the components of the overall state.

A FSA lacks the type of combinatorial structure that is part of a CSA. Due to this, Putnam's employment of the FSA is claimed to be not constrained enough for capturing the kind of complex structure that computation and cognition involve. A CSA (due to the non-monadic nature of its states) does satisfy the need for complex combinatorial structure, and is as such much more suitable for the study of cognition. It is the complex structure of the CSA that puts much stronger constraints on implementation. According to Chalmers, the structural complexity of the CSA model, and the complex dependencies that exist among different parts of the model, precludes many physical systems from realizing this type of computational formalism. So, not only is there a much better fit between the complexity of cognition and the formal structure of a CSA (as compared to that of an FSA), but this formalism also puts much stronger constraints on its realization by a physical system.

Both Chalmers and Chrisley argue that Putnam has failed to prove the validity of the universal realizability thesis because he based his proof on an a weak conception of causality and on an FSA formalism that is inadequate to characterize cognition. Therefore, Putnam's claim that the notion of computation is of no use for the study of cognition can be rejected.

However, we would like to stress two points that remain. First, isomorphism continues to be the basis for determining that a physical system realizes computations and representational states. That is, despite the constraints put on the type of causality involved and the type of formalism thought necessary, the interpretation remains that computations and representations are realized by a physical system when the representational-computational description is isomorphic with the physical-causal make-up of the system. Second, and more importantly, these more stringent constraints on the application of isomorphism do not exclude the possibility that *any* physical system realizes *certain* computations; or, as Chalmers (1995, pp. 396–397) puts it himself:

> Does every system implement some computation? Yes. For example every physical system will implement the simple FSA with a single internal state; most physical systems will implement the 2-state cyclic FSA and so on. This is no problem, and certainly does not render the account vacuous. That would only be the case if every system implemented *every* computation, and that is not the case ... In general, there is no canonical mapping from a physical object to "the" computation it is performing. We might say that within every physical system, there are numerous computational systems ... however, there is no threat of vacuity. The question of whether a given system implements a given computation is still entirely objective.

We do not agree with Chalmers that there is not a serious problem for the computational approach to cognition. Though the notion of isomorphism is more constrained (or rather, can be made to be more constrained) than Putnam suggests, it still is relatively unconstrained in the sense that every physical system automatically instantiates *several* computational and representational systems. Of course, it is common practice within cognitive science to use additional constraints while trying to find out which computational and representational process a system is actually performing. Relevant in this regard are the suggestions by Pylyshyn (1980, 1984) concerning time complexity and cognitive penetrability. Other obvious sources of constraints are the details of the analyses at the physical level (e.g. neurophysiological data), and ideas concerning the nature of the task which the system is performing (Marr's 1981 computational level). All these constraints help to narrow down the class of interesting computational and representational processes that a system can be supposed to perform. However, in relation to the question whether a particular system is in a realistic sense representational or not, these constraints play a minor role, if any at all. The basis for claims concerning the *presence* of representations within a system remains the establishment of an isomorphism between the computational description and the causal processes taking place at the physical level. The problem is that isomorphism, even with the additional constraints indicated by Chalmers and Chrisley, still allows many systems to qualify as representational while the actual presence and use of representations within these systems is, at least intuitively, highly debatable. We will examine this issue more fully in the next two sections, when we examine the case of the Watt Governor.

4. The Watt Governor: representational or non-representational?

In his paper "What might cognition be, if not computation?" van Gelder (1995) presents the example of the Watt Governor. This device, invented by James Watt in the late 18th century, was designed to regulate the irregular output of steam engines into a smooth motion of a flywheel needed for high quality performance in industrial applications of steam power. This result was accomplished by connecting two moveable arms (with weights attached to them) by means of hinges to a vertical spindle geared into the flywheel. Depending on the speed of the rotation of the flywheel the angle between the arms and the spindle of the flywheel increases or decreases. The upward or downward movement of the arms forces the engine's throttle valve to close or to open.

According to van Gelder (1995), the Watt Governor is not only a paradigmatic case of a dynamical system (p. 367), but it also constitutes a new landmark for models of cognition (p. 381). In the way it operates, the Watt Governor does not rely on internal representations and computational manipulations of them. Therefore, its operation should not be explained in these terms.

Van Gelder argues against the interpretation of the Watt Governor as a representational and computational system. He claims that this view is grounded in the intuition that the angle between the rotating arms of the Watt Governor and the spindle of the flywheel around which they rotate is a *representation* for engine speed. It is against this intuition that van Gelder presents the following four arguments.

First of all, the central feature that gives rise to the idea that arm angle is a representation of engine speed is the apparent *correlation* between these two elements. However, this idea is unwarranted since mere correlation is not enough for assigning representational status. As argued by various authors (e.g. Bechtel, 1998; Clark & Toribio, 1994; Palmer & Kimchi, 1986), what is missing for calling a correlational relation like this a representational relation is the *use* of the correlational information. That is, there must be an inner state that has the function of carrying the information embedded in the causal correlation, whereby, in addition, certain processes use this information in the course of the system's generation of behavior. Only when such an inner state can be identified and the correlational information is used in a systematic way are the necessary conditions met for calling the relation representational.

Second, van Gelder points out that the correlational relation between arm angle and engine speed only exists during a state of equilibrium, that is, when both engine speed and arm angle are not changing. Whenever the engine speed changes (for example, when the workload increases) the correlation between the two elements is disrupted. Because of that, van Gelder argues that there is not even a full correlation to be exploited in claiming that the Watt Governor is representational. So, not only is correlation alone insufficient for grounding the representational claim, in the case of the Watt Governor the correlation itself is not even present at all times.

Next, van Gelder argues that a device or system is only representational when the notion of representation has a clear explanatory utility for understanding the system. However, since, for the Watt Governor, a description in representational terms does "explain nothing over and above what one could explain before," van Gelder (1995, p. 352) argues that there is no ground for the claim that the governor is representational.

Fourth and finally, van Gelder (1995, p. 353) argues that "the notion of representation is just the wrong sort of conceptual tool to apply" in our effort to fully understand the relation between engine speed and arm angle. The reason for this is his contention that the notion of a representational relation (realized by the correlational relation) is too simple to capture the actual interaction between the Watt Governor and the engine. Only the framework of the dynamics can cope with this, since this is the only framework that offers the tools for a correct description that can deal with the fact that "arm angle and engine speed are at all times both determined by, and determining each other's behavior" (p. 353).

Following these arguments, van Gelder concludes that (1) in its operation the Watt Governor does not rely on internal representations, and (2) that it is better not to rely on the notions of representation and computation in an attempt to understand and explain its operation (in this respect the language of dynamics is the right alternative). The device is, in van Gelder's view, non-representational and non-computational.

Several authors have responded to van Gelder's conception of the dynamical approach to cognition and his claim that the Watt Governor is a paradigmatic example of a non-representational and non-computational device (e.g. Bechtel, 1998; Clark, 1997b; Clark & Toribio, 1994; Eliasmith, 1996, 1997). Despite the fact that most commentators are quite critical regarding van Gelder's dynamical approach to cognition, there is no clear consensus regarding his claim that the Watt Governor is a non-representational system. Eliasmith (1996, 1997), for one, argues that there is indeed no reason to doubt the non-representational and non-computational status of the Watt Governor. Clark (1997b; Clark & Toribio, 1994) comes to a similar conclusion. A main reason why he believes that the Watt Governor is a non-representational device is that the type of problem solved by the Watt Governor is not "representation-hungry." That is, since in the case of the Watt Governor there is a continuous coupling between the engine and the governor, there is simply no need for representations. The need for genuine internal representations is limited to cases in which the coupling between the system and its immediate environment is discontinuous.

So far, then, the responses support van Gelder. However, Bechtel (1998) has

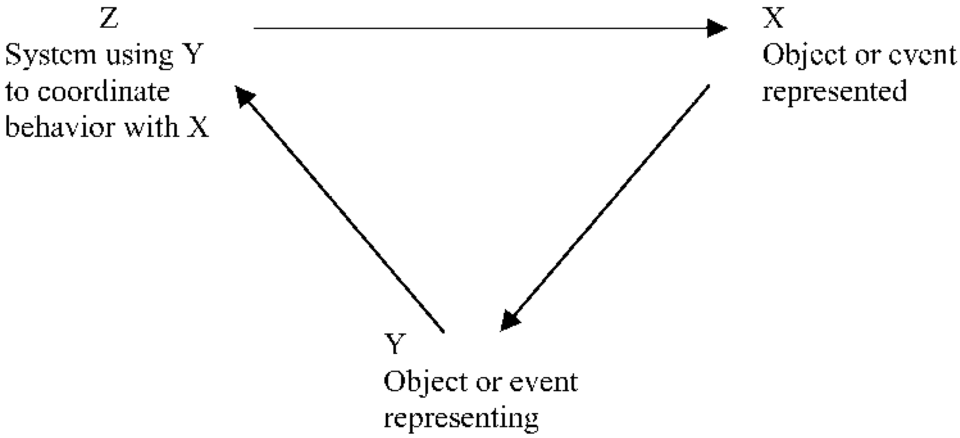

FIG. 1: Bechtel's (1998, p. 299) analysis of the notion of representation.

argued that van Gelder is wrong regarding his claim that the Watt Governor is a non-representational and non-computational system. Since Bechtel's argument can be considered as an illustration of the prominence of isomorphism in cognitivistic thinking about representations, we will examine his analysis of the Watt Governor in more detail. Thus, our aim in the remainder of this section is not to argue for or against either van Gelder or Bechtel, but to examine the role of isomorphism in their debate and to indicate why the unconstrainedness of this notion is problematic for the debate between representationalists and anti-representationalists.

According to Bechtel (1998; see also Bechtel & Richardson, 1993), the standard definition of a representation can be systematically displayed as in Figure 1. Following a functional analysis it must be demonstrated that a system's behavior depends on there being: (1) an (external) object or event *X* internally represented by the system; (2) an internal state *Y* which acts as a stand-in represents the object or event; and (3) a system *Z*, defined either as the overall system or as part of the overall system, that coordinates its behavior towards the represented object or event on the basis of the internal representation of the object or event. Representational status can only be assigned to a system when this requirement is met.

According to Bechtel, the Watt Governor can be interpreted such that there is an object *Y* that carries information about *X* and a system *Z* that uses *Y* in coordinating its behavior with respect to *X*. For the Watt Governor, *X* is the speed of the flywheel (*the represented*), *Y* is the angle of the arms with the spindle (*the representation*), and *Z* is the valve mechanism that *uses Y* (by opening or closing the throttle valve) in regulating the speed of the flywheel (*X*). Support for this view lies in the match between Bechtel's schematic representation of the Watt Governor (see Figure 2) and the components in his analysis of a representation shown in Figure 1.

With Figure 2, Bechtel indicates how the arm angle can be interpreted as a representation for engine speed that is used by the throttle mechanism in a computational way to regulate the flywheel's speed.

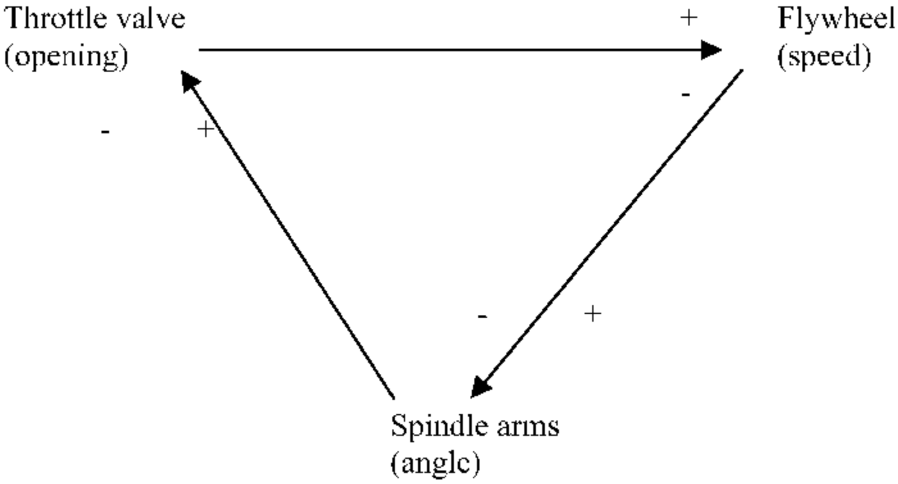

FIG. 2: Bechtel's (1998, p. 302) schematic rendition of the Watt Governor.

The importance of the isomorphism between Bechtel's Figure 2 and the Watt Governor becomes clear when we contrast Bechtel's interpretation with van Gelder's (1995) hypothetical computational solution to the problem facing the Watt Governor. According to van Gelder (1995, p. 348), a possible algorithm for a hypothetical steam-regulating computational device could be as follows:

1. Measure the speed of the flywheel.
2. Compare the actual speed against the desired speed.
3. If there is no discrepancy, return to step 1. Otherwise,
    a. measure the current steam pressure;
    b. calculate the desired alteration in steam pressure;
    c. calculate the necessary throttle valve adjustment.
4. Make the throttle valve adjustment.
5. Return to step 1.

Although a computational device operating according to this algorithm would exhibit the same behavior as the Watt Governor, the Watt Governor does *not* instantiate this algorithm. No part of the Watt Governor can be interpreted as somehow measuring the current steam pressure, nor is this information used to calculate the desired alteration in steam pressure or to calculate the necessary throttle valve adjustment. Following this, van Gelder claims that the Watt Governor cannot be interpreted as instantiating *any* computational model whatsoever, since the principles underlying the operation of the Watt Governor are in no way expressible in the type of algorithm he has presented. The crucial difference with Bechtel's model is that it *can* be viewed as being instantiated by the actual Watt Governor, precisely because of the isomorphism, whereas this is not possible for van Gelder's algorithm. As Bechtel (1998, p. 302) puts it, Figure 2 is "in the same format" as what the Watt Governor actually looks like. It is isomorphism, then, that ultimately supports Bechtel's claim that the Watt Governor actually instantiates a computational model.

In all, Bechtel has presented a computational and representational model that, given the isomorphism between the model and the physical system in question, can be claimed to be instantiated by the Watt Governor. We have no wish to quarrel with Bechtel's claim that the Watt Governor is, despite van Gelder's arguments to the contrary, computational as well as representational. Rather, we will argue that his analysis provides an example of the problematic operationalization of representation indicated earlier.

5. The ubiquity of representations

Earlier, in Section 3, we discussed Putnam's universal realizability thesis. According to Putnam, every open physical system realizes every computation. Chalmers and Chrisley have rejected this thesis because Putnam relies on an interpretation of causality that is too weak, and because the formalism he employs is not of the right type to characterize the underlying structure of cognition. However, we have pointed out that the additional constraints put on isomorphism do not exclude *any* physical system from realizing *certain* representational states and computations. Isomorphism, therefore, is not a severe enough constraint on the notion of representation in the sense that it simply does *not* allow a system to be non-representational. Because for every system certain computations can be found with which the physical-causal make-up is isomorphic, the conclusion that every physical system is computational and representational is unavoidable. For the Watt Governor too, one or more computational interpretations can be found with which the physical-causal make-up is isomorphic, as Bechtel's interpretation clearly illustrates. Of course, Bechtel's interpretation is interesting because it is relevant to *our understanding* of the rationale behind Watt's mechanism. That is, we can understand how the Watt Governor performs its task by understanding the relationship (as outlined in Figure 2) between the speed of the flywheel, the arm angle, and the throttle valve. However, the question here is whether this justifies the assumption that the Watt Governor actually *has* content-bearing vehicles that are used to guide its functioning. At one point, Bechtel (1998, p. 303) clearly seems to think so:

> The fact that the angle of the spindle arms represents the speed of the flywheel becomes more clear when we consider why it was inserted into the mechanism to begin with. The flywheel itself has a speed, but there is no way to use this directly to open and to close the valve. The spindle and arms were inserted so as to encode information about the speed in a format that could be used by the valve opening mechanism. The reason no one has to comment explicitly on the fact that the arm angles stand in for and thus represent the speed of the flywheel is that this system is very simple, and most people *see* the connection directly.

We suggest, however, that there is another reason why (at least some) people are reluctant to ascribe the use of representations to the Watt Governor. Claiming that the throttle valve actually uses the arm angle as a representation seems *excessive*. Even though there is an isomorphism between Bechtel's representational description of the Watt Governor and the actual mechanism, and even though Bechtel's description provides insight into the functioning of the mechanism, *all* that seems to be happening in the Watt Governor is a transformation of forces. If a transformation of forces is enough to warrant a representational (in a realistic sense) interpretation, it would follow that even a *bike* is a representational system. One could say, after all, that the force that a person exerts on the pedals is represented by the chain, which is used by the chain ring in order to determine the speed of the wheel.

Of course, as Bechtel himself has pointed out [3], a difference between our example of an ordinary bike and the Watt Governor is that in the Watt Governor there is a feedback loop in which the representation is used to affect that which is represented, whereas in the case of the pedals, chain and chain ring there is not, because the chain ring is not influencing the pedals. But, first of all, one may point to the existence of certain bikes (one-wheeled ones used for balancing acts in circuses, or two-wheeled ones used for bike races, held in stadiums, in which the "sur place" plays a prominent role) that have the connection between the chain and the chain ring fixed in both directions, which creates a feedback loop between the chain ring and the pedals, thus fulfilling even this condition. Calling bikes such as these representational systems would not seem to be less excessive than in the case of ordinary bikes.

Second, the presence of a feedback loop is not an essential requirement for there being representations. In many cases, representations could guide behavior in relation to entirely different objects or events than the represented object or event itself. For instance, a person's (representational) thought about a person *P* might influence her behavior regarding a person *Q*, without ever influencing her behavior in relation to *P* himself (e.g. she even never might act in relation to person *P*). Note also that this feedback loop is not part of Haugeland's definition referred to earlier, as that definition refers to "guiding behavior" in general, without specifying that the behavior has to be guided in relation to the object or event represented. In this sense, a strict application of Bechtel's third condition on representations (that there should be a system *Z* that coordinates its behavior *towards the represented object or event* on the basis of the internal representation of the object or event) is not just ruling out our example of the ordinary bike, but turns out to be overly restrictive in excluding many cases of representation use that normally would count as genuine. If one refrains from a strict application of the feedback criterion, this problem would be evaded, but then ordinary bikes would qualify as representational systems as well.

In any case, the point of our example of the bike is that if the Watt Governor would qualify as a representational system on the basis of Bechtel's scheme, there does not seem to be a principled way to distinguish mere transformation of forces from representation use, and representations would be ubiquitous. This, we submit, would seriously undermine the status of representations as explanatory constructs.

Bechtel (1998, p. 305) acknowledges this problem at least implicitly, for he states that the kind of representations that he attributes to the Watt Governor are low-level, rather like the notion of a "presentation" as discussed by Grush (1997; Bechtel, 1998, p. 314, note 4; see also Section 6 below), and not of the kind used in planning and decision making. He is the first to admit that the Watt Governor "is not a particularly interesting case of a representational system" (p. 301) and moreover says that the notion of representation that he has been using here "is a very minimal one, and one that admittedly makes representations fairly ubiquitous" (p. 313) [4].

In our view, qualifications such as these reflect the lack of a precise operationalization of representation. When "weak" or "low-level" forms of representation do, ultimately, count as representations, it is unclear exactly when or how a system can be shown *not* to be representational. A notion that encompasses almost everything from low-level force transmission to high-level model construction and planning is too unconstrained to be useful as a fundament for research in cognitive science. Moreover, it does not allow even the start of a debate on the validity and necessity of representations in models of cognition.

Indeed, from the anti-representationalist perspective, the easy applicability of the notion of representation has not gone unnoticed and it has worried several theorists. For example, Brooks (1991, p. 149), after introducing his non-representational approach to robotics, ventures the worry that it will always be possible to call his robots representational:

> An extremist might say that we really do have representations, but that they are just implicit. With an appropriate mapping of the complete system and its state to another domain, we could define a representation that these numbers and topological connections between processes somehow encode. However, we are not happy with calling such things a representation. They differ from standard representations in too many ways.

More recently, Beer (2000, p. 97), after describing several dynamical approaches to cognition, says:

> More fundamentally, there is a great deal of controversy over exactly what a representation or a computation is. Some have argued that, although objectivist, symbolic notions of representation and computation might be inapplicable to dynamical models, suitably distributed, analog, context-dependent and action-oriented notions will be applicable. However, great care must be taken in generalizing these notions. If any internal state is a representation and any systematic process is a computation, then a computational theory of mind loses its empirical force.

Interestingly, several years ago, Churchland and Sejnowski (1992, pp. 61–69) anticipated the question of whether it is too easy to count a particular mechanical device as computational. Their discussion did not concern the Watt Governor, but something that perhaps just as easily could have become an anti-representational landmark: Garrett's improved threshing machine from 1851. Churchland and Sejnowski subscribe to the importance of isomorphism by holding that a systematic mapping between a functional description and the physical states of a system warrants, in principle, a computational interpretation [5], claiming that Garrett's machine may be considered as a computational and representational system. Importantly, they do discuss the problem of a characterization of representation and computation that is too general; as they (1992, p. 66; second emphasis added) put it:

> It may be suggested as a criticism of this very general characterization of computation that it is *too* general. For in this very wide sense, even a sieve or a threshing machine could be considered a computer, since they sort their inputs into types, and *if one wanted to spend time at it*, one could discover a function that describes the input–output behavior. While this observation is correct, it is not so much a criticism as an apt appreciation of the breadth of the notion.

The phrase "if one wanted to spend time at it" captures precisely the upshot of our examination of the debate between Putnam, Chalmers, and Chrisley on isomorphism. Churchland and Sejnowski's positive view on the breadth of computation can perhaps be understood if one takes into account that they took the very notions of computation and representation as largely undisputed within cognitive science. In the debate between symbolic cognitive science and connectionism, the focus was on the nature and format of computation and representation; not their very role or existence. However, with the rise of the dynamical approach to cognition things have changed. Precisely in a debate with anti-representationalists it cannot be taken as self-evident that cognition consists of computational transitions of representational states. Given the change in the kind of dispute, the appreciative attitude as expressed by Churchland and Sejnowski concerning the broadness of the notions of representation and computation can no longer be sustained without argument.

6. Conclusion

Ultimately, Cliff and Noble's question of how to establish (on the causal mechanistic level, as they would have it) whether a system is representational or not remains unanswered. Importantly, there must be something more than the mere operation of causal forces in order to warrant talk of representations. That is, the standing-in, the content, must play a crucial role in the system's functioning. But specifying exactly what this crucial role of the content is, and *how* it can be discovered that the behavior of the system *relies* on this crucial role of content when examining a system's causal-physical operations, turns out to be surprisingly difficult. One aspect of the difficulty is that the standard answer (involving isomorphism) to this problem seems to imply the omnipresence of representations. As things stand, then, cognitive science is lacking a proper (principled) distinction between a "mere" causal force and a representation. Without such a principled distinction, a useful operational definition cannot be formulated.

Although this concludes the main intent of our paper, we would like to examine a recent suggestion about how to proceed towards a more constrained interpretation of representation, indicated by Grush (1997; see also Clark & Grush, 1999; Wheeler & Clark, 1999). For a start, it may help to look back at the time of the debates between cognitive scientists and behaviorists, and the reasons why representations were thought to be so useful. Ultimately, representations were considered necessary because without them systems appeared to be incapable of performing *higher* cognitive processes (as occurring in e.g. language, planning and problem solving). Perhaps, then, it is wise to restrict the label "representational system" for systems capable of engaging in high-level cognitive processes. Of course, the phrase "high-level cognitive process" is not very precisely defined within cognitive science. Still, formulating a clear distinction between cognitive and non-cognitive systems might be a first step towards restricting, in a productive sense, the applicability of the notion of representation.

An example of such a distinction is given by Grush (1997, p. 5), who offers a theory of representation that "allows us to sharply distinguish those complex behaviors which are genuinely cognitive from those which are merely complex and adaptive." The basic idea is that genuinely cognitive behavior depends on a emulation or simulation of some external system that can be used off-line (without the external system being perceivably present), in order to provide information about what the external system would do under diverse conditions (Grush, 1997, p. 12; see also Clark & Grush, 1999; Wheeler & Clark, 1999, pp. 123–124). Thus, the emulator represents the external system, and we can engage, for instance, in counterfactual reasoning about the external system. Only in the case of a representation is there an inner state that stands in for some other state, and that is used by a system in coordinating its behavior towards the object or event represented. In the case of a presentation there is no off-line use of certain information. Presentations, like representations, play a role in the production of behavior, but this role is embedded in the direct and continuous coupling between the presentation and the elements presented by it. The angle of the arms of the Watt Governor is a presentation of the engine speed. Since the angle of the arms and the engine speed are in a continuous coupling relation, the angle does not qualify as a genuine "full-blooded" representation. Therefore, on this suggestion, the Watt Governor is a non-representational device that operates on the basis of presentations. When restricted to capacities such as these (i.e. the capacity to decouple from the environment and use an internal emulation instead; Grush, 1997, p. 17), the notion of representation no longer harbors the danger of being omnipresent. It is fair to say though that this strong interpretation of representation does not coincide with the more general interpretation currently used within cognitive science. Thus, "some degree of revisionism" (Clark & Grush, 1999) regarding the operationalization of the notion of representation within cognitive science seems called for [6].

Whatever the value of this particular suggestion, it appears to us that cognitive science can no longer tolerate a situation in which its core concepts allow extremely conflicting positions on whether or not a relatively simple system like the Watt Governor is representational. In our view, the value of these recent debates consists in emphasizing the need for a more stringent definition of representation and computation. Without a properly constrained notion of representation, the debate between representationalists and anti-representationalists is bound to remain a debate for the sake of appearance.

Acknowledgements

We would like to thank two anonymous reviewers, William Bechtel, Maria Eunice Quilici Gonzalez, and especially Jelle van Dijk for their constructive comments on earlier versions of this paper.

Notes

[1] Although the notions of computation and representation are not to be identified with each other, it is clear that within cognitive science there is an intimate relation between the two notions. See, for instance, Fodor (1975, p. 31): "No representations, no computations. No computations, no model"; and Churchland and Sejnowski (1992, p. 62): "we can consider a physical system as a computational system when its physical states can be seen as representing states of some other systems, where transitions between its states can be explained as operations on the representations." Given our aims in this paper, we will allow ourselves sometimes to be speaking of representations and sometimes of computations.

[2] Although Haugeland's interpretation is not the only one, it is the one most often referred to by both proponents and opponents of the anti-representational claims coming from the dynamical approach. Two other interpretations sometimes referred to (e.g. Bechtel, 1998; Cliff & Noble, 1997) are the ones offered by Newell (1980) and Harvey (1996). According to Newell (1980, p. 156): "The most fundamental concept for a symbol system is that which gives symbols their symbolic character, i.e. which lets them stand for some entity. We call this concept *designation* (…) An entity X designates an entity Y relative to a process P, if, when P takes X as input, its behavior depends on Y" (Newell, 1980, p. 156); entity X is in this case represented by Y. Cliff and Noble (1997), on the other hand, rely on Harvey's (1996) characterization according to which "… P is used by Q to represent R to S," whereby R is the object or state of affairs represented by P. Despite some differences, the import of these interpretations is the same as Haugeland's.

[3] Personal communication. Bechtel also indicated a problem with the requirement of the feedback loop in relation to a proper definition of representation, as discussed later on in the text.

[4] Bechtel (2001) stresses that, on his account, representations are relatively ubiquitous, but not completely promiscuous. His safeguard against rampant representations lies in the "designed use" of representations. As he says, "Representations are found only in those systems in which there is another process (a consumer) designed to use the representation in generating its behavior (in the simplest cases, to coordinate its behavior with respect to what is represented) and where the fact that the representation carried that information was the reason the down line process was designed to rely on the representation" (p. 4). However, the bikes we mentioned above would still count as representational, since just as in the case of the Watt Governor, these bikes (specifically the chain rings) could be described as being designed to use the representations (the chains) in order to perform their function. In this respect the teleological component in Bechtel's approach to representation does not make a difference.

[5] Churchland and Sejnowksi (1992, p. 65) definition again illustrates the importance of isomorphism: "a physical system computes some function *f* when (1) there is a systematic mapping from states of the system onto the arguments and values of *f*, and (2) the sequence of intermediate states executes an algorithm of that function … We count something as a computer because, and only when, its inputs and outputs can usefully and systematically be interpreted as representing the ordered pairs of some function that interests us."

[6] Actually, speaking about "some degree of revisionism" may be a considerable understatement. The tendency to use the label "representation" for, in the terminology of Grush, presentations is widespread not only in cognitive science but also in neuroscience and biochemistry. Bechtel (2001) indicates that this entails that representations would appear, for instance, in physiological systems such as fermentation. As he says, "But this, I would contend, is as it should be. Representational or intentional vocabulary is in fact regularly used in the sciences dealing with such systems. Without it, scientists would be hard pressed to explain how these systems perform the tasks for which they appear to have evolved" (p. 5). However, the distinction between realism and instrumentalism (see Section 2) is relevant here. The fact that in several disciplines scientists use representational vocabulary when talking about the systems they analyze is not enough to establish the actual presence of representations that as such and in virtue of their content do causal work. Speaking about representations in relation to fermentation may be useful for analysis, but it need not constitute an ontological claim. Second, the point that anti-representationalists raise is that the widespread representational talk might not do justice to the actual internal processes operating within a system. Indeed, exactly this was one of the main messages of van Gelder (1995). So, the actual ubiquity of representational talk should not be taken as an argument for its aptness, although it does indicate that "the degree of revisionism" would have to be substantial.

References

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

BECHTEL, W. (2001). Representations: from neural systems to cognitive systems. In W. BECHTEL, P. MANDIK, J. MUNDALE & R.S. STUFFLEBEAM (Eds) *Philosophy and the neurosciences: a reader*. Oxford: Basil Blackwell.

BECHTEL, W. & RICHARDSON, R.C. (1993). *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

BEER, R. (1995a). Computational and dynamical languages for autonomous agents. In R.F. PORT & T. VAN GELDER (Eds) *Mind as motion: explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

BEER, R. (1995b). A dynamical systems perspective on agent–environment interaction. *Artificial Intelligence*, *72*, 173–215.

BEER, R. (2000). Dynamical approaches to cognitive science. *Trends in cognitive science*, *4*, 91–99.

BICKHARD, M.H. & TERVEEN, L. (1995). *Foundational issues in artificial intelligence and cognitive science: impasse and solutions*. Amsterdam: Elsevier.

BLOCK, N. (1995). The mind as the software of the brain. In E.E. SMITH & D.N. OSHERSON (Eds) *Thinking: an invitation to cognitive science*. Cambridge, MA: MIT Press.

BROOKS, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

BROOKS, R. (1995). Intelligence without reason. In L. STEELS & R. BROOKS (Eds) *The artificial life route to artificial intelligence: building embodied, situated agents*. Hillsdale, NJ: Erlbaum.

CHALMERS, D.J. (1994). *A computational foundation for the study of cognition: Philosophy–Neuroscience–Psychology Technical Report 94–03*. St. Louis, MO: Washington University.

CHALMERS, D.J. (1995). On implementing a computation. *Minds and Machines*, *4*, 391–402.

CHALMERS, D.J. (1996). Does a rock implement every finite-state automaton? *Synthese*, *108*, 309–333.

CHRISLEY, R.L. (1995). Why everything doesn't realize every computation. *Minds and Machines*, *4*, 403–420.

CHURCHLAND, P.S. & SEJNOWSKI, T.J. (1992). *The computational brain.* Cambridge, MA: MIT Press.

CLANCEY, W.J. (1997). *Situated cognition: on human knowledge and computer representation.* Cambridge: Cambridge University Press.

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

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

CLARK, A. & GRUSH, R. (1999). Towards a cognitive robotics. *Adaptive Behavior*, *7*, 5–16.

CLARK, A. & TORIBIO, J. (1994). Doing without representing? *Synthese*, *101*, 401–431.

CLIFF, D. & NOBLE, J. (1997). Knowledge-based vision and simple visual machines. *Philosophical Transactions of the Royal Society: Biological Sciences*, *352*, 1165–1175.

DENNETT, D. (1991). Real patterns. *Journal of Philosophy*, *88*, 27–51.

ELIASMITH, C. (1996). The third contender: a critical examination of the dynamicist theory of cognition. *Philosophical Psychology*, *9*, 441–463.

ELIASMITH, C. (1997). Computation and dynamical models of mind. *Minds and Machines*, *7*, 531–541.

FODOR, J.A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

GRUSH, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*, 5–23.

HARVEY, I. (1996). Untimed and misrepresented: connectionism and the computer metaphor. *AISB Quarterly*, *96*, 20–27.

HAUGELAND, J. (1991). Representational genera. In W.M. RAMSEY, S.P. STICH & D.E. RUMELHART (Eds) *Philosophy and connectionist theory*. Hillsdale, NJ: Erlbaum.

HENDRIK-JANSEN, H. (1996). *Catching ourselves in the act: situated activity, interactive emergence, evolution, and human thought.* Cambridge, MA: MIT Press.

KEIJZER, F. (2001). *Representation and behavior.* Cambridge, MA: MIT Press.

KELSO, J.A.S. (1995). *Dynamic patterns: the self-organization of brain and behavior.* Cambridge, MA: MIT Press.

KIM, J. (1996). *Philosophy of mind.* Boulder, CO: Westview Press.

MARR, D. (1981). *Vision.* San Francisco, CA: W.H. Freeman.

NEWELL, A. (1980). Physical symbol system. *Cognitive Science*, *4*, 135–183.

PALMER, S.E. & KIMCHI, R. (1986). The information processing approach to cognition. In T.J. KNAPP & L.C. ROBERTSON (Eds) *Approaches to cognition: contrasts and controversies.* Hillsdale, NJ: Erlbaum.

PORT, R.F. & VAN GELDER, T. (1995). *Mind as motion: explorations in the dynamics of cognition.* Cambridge, MA: MIT Press.

PUTNAM, H. (1988). *Representation and reality.* Cambridge, MA: MIT Press.

PYLYSHYN, Z.W. (1980). Cognition and computation: issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*, 111–132.

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

SEARLE, J. (1992). *The rediscovery of mind.* Cambridge, MA: MIT 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*, *92*, 345–381.

VAN GELDER, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, 615–665.

VAN GELDER, T. & PORT, R.F. (1995). It's about time: an overview of the dynamical approach to cognition. In R.F. PORT & T. VAN GELDER (Eds) *Mind as motion: explorations in the dynamics of cognition.* Cambridge, MA: MIT Press.

VAN ROOIJ, I., BONGERS, R. & HASELAGER, W.F.G. (2002). A non-representational approach to imagined action. *Cognitive Science*, *26*, 345–375.

VARELA, F.J., THOMPSON, E. & ROSCH, E. (1991). *The embodied mind: cognitive science and human experience.* Cambridge, MA: MIT Press.

WHEELER, M. & CLARK, A. (1999). Genic representation: reconciling content and causal complexity. *British Journal for the Philosophy of Science*, *50*, 103–135.