

Representations: The Hard Problem for Artificial Life

Harry Halpin¹

¹School of Informatics, University of Edinburgh, Edinburgh, EH8 9LW
H.Halpin@ed.ac.uk

Abstract

Representationalism, the original driving theory behind artificial intelligence and cognitive science, has fallen upon hard times. Representationalism identifies cognitive processes with the manipulation of explicit symbolic structures according to well-defined rules. Rocha and Hordijk put forward three requirements for something to qualify as a representation: dynamically incoherent memory, construction code, and self-organization; they also forsake the traditional idea of representation as a “standing-in” relationship between a representation and its content. While we note that Rocha and Hordijk’s three requirements are an improvement over the definition of representation as a style of explanation, we argue that this “standing-in” relationship is the defining characteristic of representations. We define representations as physically implemented structures that can create local effects that are effective by virtue of their correspondence with non-local parts of the world, and argue that this definition, while being perfectly natural, does present a difficult, although not insurmountable, challenge for artificial life.

Introduction

Representationalism, the idea that cognition involves the manipulation of symbolic structures, originally was at the heart of artificial intelligence and cognitive science. At first it was taken to be a revolutionary force capable of explaining previously inexplicable phenomena. Yet its proponents could not explain much of physical behavior by virtue of representations alone, and artificial life and other disciplines that did not rely on explicit representations have now risen to the forefront. Despite the success of anti-representationalism, there has been recent broad recognition that some problems are “representation hungry,” and that an adequate theory of representations is needed (Clark, 1997). Several researchers in artificial life have suggested that work in cellular automata (CAs) can either evolve representations as explanations (Mitchell, 1998) or provide a simulation of the evolution of physically implemented representations (Rocha and Hordijk, 2005). We argue representations are not just explanations, but physically realized symbol systems, and that current efforts in artificial life ignore a crucial component of representations, a correspondence condition

with a non-local process. Many researchers in artificial life are aware of these complexities and may even be aiming at a “lower” level of representation or radical re-interpretation of representation. If so, there is still much work to go if cognitive science is to believe the story that artificial life is capable of evolving genuine representations and the very concept of representation needs theoretical clarification first.

Hard Problems

Representations are tricky for a number of reasons, and we will follow the lead of Chalmers (Chalmers, 1995) by classifying representation as either a “hard” and “easy” problem (Chalmers, 1995). An *easy problem* is one “that seem directly susceptible to the standard methods of cognitive science,” while a *hard problem* is one that “is intractable, and for which no good explanation can be given” (Chalmers, 1995). The problems with representations are manifold: How can representations be accessed and maintained? Where are they located? What kinds of representations exist and why are they useful? These are all difficult in their own right, but are fundamentally dependent on the larger question: What is a representation? In other words, what are the identifying *requirements of the representational* that allow us to separate all those parts of the world that are representations from those the rest? If some criteria can not be found, either everything is a representation or representations do not exist.

Haugeland states that representation can be defined by their “standing-in relationship” (Haugeland, 1991). A representation “stands-in” for something, so that when that thing is not present it may “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 in for is its *content*; and its standing in for that content is *representing it*.” (Haugeland, 1991). This “standing-in” relationship is called *interpretation* in logic. This relationship is also known as the *substitutivity* property and *meaning* in philosophy of language. It is this relationship that makes the problem *hard*. In physics, as commonly conceived, for one thing to *affect*

another thing (change the “physical configuration” of), both things must be physically concrete and in direct contact with each other. So properties that are effective are also *local*, such that there is “no action at a distance” (Smith, 1995). If cognitive science wishes to be at least coherent with physics, and if physics allows cause and effect only to be local, then the “standing-in” relationship is physically *spooky*, since it by definition does work by virtue of standing in for something that is not local. Note that we are not requiring context-free representations, or some sort of representational system that is independent of content, but aiming for a weaker and more acceptable notion of representation that does not posit such strong requirements. However, without any requirements the concept of representation would be vacuous.

Unnatural Representations

Representations have been a bugbear of naturalizing the mind (Dretske, 1995). While much of contemporary philosophy of the mind is concerned with the other bugbear of consciousness, there are still depths to be plumbed to understand the fundamental notion of representation. The relationship between a representation and its content has been called *reference*, and reference seems to violate physics. Alonzo Church is attributed as saying that “reference outstrips the speed of light” (Smith, 1995). If I am thinking about the Eiffel Tower, my thought refers to an actual Eiffel Tower in Paris instantly, without any direct physical connection between myself and the Eiffel Tower at the time of thinking. I can imagine the Eiffel Tower even I am not in Paris looking at it, and even if I have never been to Paris but only seen pictures of the Eiffel Tower. It is this access to the content of the representation that is spooky, since the content is removed from the representation of the content in both time and space. Indeed, for this to work the representation itself, but not necessarily its content, must be physical and local. Therein lies the core of implementing representations in artificial systems.

My referential access to the content of a representation *does work*, which makes it even more spooky. While in Scotland, I can think of the Notre Dame and remember that the Notre Dame is in Paris. If I desire to go, I can use my knowledge of the Eiffel Tower to book a ticket to Paris, a very physical and effective action, when my plane lands in Paris. Now, one could hypothesize that my representation of the Notre Dame is physically epiphenomenal. This hypothesis seems improbable at best, as the correlations between my representations “line up” with my behavior, and these correlations can be externalized and so observed. The main problem is that while representations seem almost necessary to explain human language (Chomsky, 1980), and also certainly needed in any theory of the mind, their relationship with their content is non-effective at the time they are doing work, and so one suspects, this relationship is possibly non-physical. This would prevent the mind from being nat-

uralized with the physical (Dretske, 1995).

To further elucidate our terminology by an appeal to physics, there are generally two kinds of separate events in a relativistically invariant theory (a theory that obeys the rules of special relativity, in that the theory looks the same for any constant velocity observer). There are those that are separated by space, which we call “non-local,” and those that are separated by time, which when causally connected we call “local” and “effective.” The word “spooky” is used to describe the fact that at their given moment of use, a representation and its content are space-like separated, and until an explanation capable of implementation is given to show how a representation and its content are actually local, or merely time-like separated, then the relationship of reference will remain physically spooky.

Problems that do not fit within an explanatory framework but that de facto either exist or seem to exist are considered “hard” problems. These problems can be banished by two routes: either an expansion, refactoring, or revolution of the explanatory framework such that a solution for the “hard problem” is forthcoming, or proof that the problem does not exist. In the case of representations, neither type of solution has been forthcoming. However, two routes have been tried by researchers in artificial life. The first is to explain representations as modes of explanations (Mitchell, 1998), while the second is define representations in terms of physical properties and show how these physical properties could naturally evolve (Rocha and Hordijk, 2005). While the two routes may be compatible, they are not necessarily coupled and so can be dealt with separately.

Representations as Explanations

Philosophical disputes are often disguised as methodological clashes, and one case in point is the contrast in cognitive science between computer simulations rich with explicit representation (“computation”) and the use of differential equations without any obvious explicit representations (“dynamics,” or more broadly construed, *anti-representationalism*). Increasingly, more researchers argue that we need both methodologies (Mitchell, 1998). I am inclined to believe this is true, but it dodges the simple question: Do representations even exist?

Representationalist cognitive scientists like Bechtel make a crucial point that representations can be thought of as explanations (Bechtel, 1998). Avoiding the spookiness of representations, Bechtel focuses on what a representational account must entail. He believes that while dynamic accounts tend to be “covering-law” accounts (mathematical equations that predict quantitative data), representational accounts are “mechanistic accounts” that have employ two qualitative heuristics: decomposition and localization. “Decomposition is the assumption that the overall activity results from the execution of component tasks” while “localization is the assumption that there are components in the system that per-

form these tasks” (Bechtel, 1998). Now there are two distinct things going by the term “representation.” The first is a *physical representation*, which is a physically implemented system capable of the “standing-in” relationship that is *at least partially internal to the observed system*. The second is a *representation as explanation*, in which the decomposition and localization is observed in the system by the observer, and to explain this *the observer uses their own representations*, which are *external to the observed system*. It is often assumed that these are the same thing.

To assume so would make the classic mistake of projecting one’s own internal representations unto the subject matter being studied and then assuming those internal representations to be true, a classic mistake in cognitive science as noted by Maturana and Varela (Maturana and Varela, 1973). Let us assume there is a neural network under investigation that has been trained, unknown to our observer, to perfectly model whether or not it has been exposed to the presence of a duck via a video camera hooked up to the neural network. To even a clever cognitive scientist, if he is unaware of how the duck has been coded in the input and how this duck affects output, inspection of the neural network weights in search of an explanation is non-trivial at best, futile at worst. Our scientist will likely sensibly pursue a dynamic systems description of the network. Is this explanation incorrect? The answer would seem to be no. I would further argue that since the internal states of the network preserve whether or not a duck has been seen and can serve as a “stand-in” for the presence of a duck, the neural network physically has a representation of the duck, independent of the observer’s explanation. Likewise, one can explain the rhythmic tapping of fingers as a *for loop* over the symbols *fingerUp* and *fingerDown*, but that explanation does not make rhythmic finger tapping representational. In contrast to explanations, we define *representations* as physically implemented structures that can create local effects that are effective by virtue of their correspondence with non-local parts of the world

The Problem with Particles

It has been shown by a number of researchers in artificial intelligence that it is possible to evolve cellular automata that are also decomposable into macro-level structures called “particles” with localizable behavior (Crutchfield and Hanson, 1993). It has been claimed that these particles are amendable to a representationalist explanation since each particle has its own higher-order behavior (Mitchell, 1998). This “proves” representations can be evolved by artificial life, and so a *computational dynamics* can bridge the representationalist versus anti-representationalist debate. We argue that particles show that representations can be great explanations of evolved higher-order structure, but that particles are not physical representations. Particles are just emergent structures.

Particles show at least localization (the global behavior

of the CAs is localized to particles), so they can be sensibly explained in representational terms (Bechtel, 1998). Yet particles in of themselves do not represent anything. Particles are just “localized boundaries” between “simple to describe” domains (Mitchell, 1998). The content of a particle is not non-local, since the particle is said to represent the boundary between one group of CAs and another, and this boundary is precisely where the particle is to be found. The particle is a description of the boundary, not a physical representation of it, since the particle *is* the boundary. The particle do not have a *non-local* “stand-in” relationship with the boundary. Particles are going in the right direction of representation since they do have structure, but are not quite there, since at least for some period of time their content needs to be non-local for them to qualify as representations.

Physical Symbol Systems

Do physical representations exist? The answer, as given by Newell and Simon in their Physical Symbol System Hypothesis, is a resounding “yes.” A symbol is defined by its referential operation, which they call *designation* “though we might have used any of several other terms, e.g., reference” (Newell, 1980). They then define designation as when:

“An entity *X* designates an entity *Y* relative to a process *P*, if, when *P* takes *X* as input, its behavior depends on *Y*. There are two keys to this definition: First, the concept is grounded in the behavior of a process. Thus, the implications of designation will depend on the nature of this process. Second, there is action at a distance . . . This is the symbolic aspect, that having *X* (the symbol) is tantamount to having *Y* (the thing designated) for the purposes of process *P*” (Newell, 1980)

The “action at a distance” spookiness remains, as does the important “standing-in” relationship of having content that affects the behavior of the system. They then prove such systems exist by noting that such systems are “realizable in our physical universe” since classic digital computers exist. In an example, a symbolic system consists of “a *memory*, a set of *operators*, a *control*, an *input* and an *output*” (Newell, 1980).

We do not need explanations of how computers use representations because we humans designed them to deal with representations, and we program them in terms of representations. Computers are clearly effective, since they land airplanes, fire missiles, monitor heart-rates, help us calculate taxes, and so on: Their behavior can have very real results. However, Newell goes further to state that “humans are physical systems” and that “a biological based architecture could have evolved” to be a physical symbol system (Newell, 1980).

Representational Failure

The development of representations was brought to the forefront in Chomsky’s victory over behaviorism, where he

hypothesized only something like “rules” and “representations” could account for language where behaviorism failed (Chomsky, 1980). Chomsky delimited the study of linguistic syntax from the conceptual system that lurked in the background. Work by Newell and Simon on matching human protocols to computer output attempted to directly study the conceptual system (Newell and Simon, 1976). In philosophy, “the language of thought” or “mentalese” was given as the rules and representations of the conceptual system (Fodor, 1975). Yet ambitious knowledge representation languages like KRL, which combined explicit behavioral procedures attached to representations, ultimately failed (Bobrow and Winograd, 1977). There was widespread disagreement on what the components of the typical knowledge representation language actually mean (Woods, 1975). The knowledge representation languages also failed to *do anything* like inference successfully with the knowledge in tractable amounts of time. Lastly, there was the question of whether there was any empirical evidence that could determine the content of a representation given by the knowledge representation system. While it appeared that KRL could represent many things to a human observer, it was unclear if the matter of representation was anything more than opinion.

Yet if one removed the representational “labels” from the structure of the KRL semantic network, one was left with something resembling non-symbolic processing, giving rise to neural networks (Rumelhart and J. McClelland, 1986). This “connectionist” framework proved to be successful in simulating many cognitive and motor functions, while knowledge representation languages focused on inference and divorced from cognitivism (Levensque and Brachman, 1987). Without representationalism, artificial intelligence fractured.

Now the anti-representationalist trend seems to not be making progress against “representation-hungry” problems in planning and language (Clark, 1997). Representations are now found to be much more domain-specific than previously thought, curiously reminiscent of the “micro-worlds” of AI (Winograd, 1972) and are argued to provide a crucial evolutionary advantage, since it allows creatures to simulate potentially dangerous situations (Grush, 2003). If representations are important again, how did we humans evolve such structures from the physical world?

Evolving Representations

In one ambitious work, Rocha and Hordijk attempt to show how artificial life can support both dynamic processes and representations by *evolving representations* from cellular automata (Rocha and Hordijk, 2005). They begin by observing a problem we call *dynamical determinism*: “The attractor landscape of a given dynamical system, while providing several possible outcomes for the system’s dynamics, does not truly yield alternatives until there is control of the initial

conditions” (Rocha and Hordijk, 2005). They go on to state that this dynamic determinism can only be overcome through the use of *memory* to store and retrieve those initial conditions. This memory must be *dynamically incoherent*, which means it is not changed by dynamic process it initiates and so must be built on an *inert* physical structure.

The words “inert” and “incoherent” are misleading to traditional representationalists, as Rocha and Hordijk would note. First, memory is only *relatively inert* in lieu of a particular representation that creates one particular dynamic process. Since the representation sets the initial conditions, it *causes the action*, and so can be thought of as providing the “coherence” for the dynamic processes whose conditions it sets. This is similar to KRL’s representations being attached to processes (Bobrow and Winograd, 1977). As the results of one dynamic process can set the initial conditions of another ad infinitum, the “inert” structure can *change* due to the results of process it began. Also, the maintenance of the *dynamically incoherent memory* requires considerable work, as exemplified in how voltages must be carefully regulated and maintained to maintain computer memory. In other words, the memory should be *digital* to “make sense as a practical means to cope with the vagaries and vicissitudes, the noise and drift, of earthy existence” (Haugeland, 1981).

In characterizing a representation, Rocha and Hordijk develop the three following requirements (Rocha and Hordijk, 2005):

1. **Dynamically incoherent memory:** Material representations demand inert physical structures which can effectively be seen as dynamically incoherent regarding the overall machinery in which they are utilized as representations. The role of these physical structures is not defined by their dynamic characteristics but rather by their informational value.
2. **Construction Code.** There needs to exist machinery to construct arrangements of building blocks (which subsequently self-organize to produce some dynamic behavior) from the representations of Requirement 1 - an arbitrary construction code.
3. **Self-organization and selection.** The encoded building blocks, after construction via the code of Requirement 2, self-organize at different levels.

Although it could be claimed that this is mixing phenomenological levels, there is an underlying analogy to Turing machines. The construction code seems to be the equivalent of the Turing machines ability to produce behavior from the output of a symbol. The dynamically incoherent memory qualifies as the tape of the Turing machine, and the self-organization and selection provide whatever behavior the symbol it outputs entails. What is left out is the ability of the representation to be input into the system, the *in-*

put code that serves as the *construction code for representations as opposed to dynamic behavior*. This would be equivalent to a Turing machine reading an input symbol from the tape. While an output procedure without an input procedure is odd, it seems that it is assumed that evolution will provide the input. Given that DNA seems to be the type of representation that has elicited the most interest from artificial life, given its roots in theoretical biology, this should not be surprising. For with DNA there is not a distinct input process besides evolution.

Rocha and Hordijk surprisingly reject Haugeland's central tenet of representation: the "standing-in" relationship: "Semantics is about construction of dynamics and behavior. This construction is not about correspondence between internal and external elements, but about a material process of controlling the behavior of embodied agents in an environment" (Rocha and Hordijk, 2005). This reflects certain strands of neo-Fregean philosophy of language that seek to define content in terms of agency and rational explanation of behavior. By this clever move that disposes of the "standing-in" relationship, the physical spookiness of representations is banished, since the content of the representation is the behavior the representation engenders. The behavior only has to be effective, but it does not have to be effective by virtue of a correspondence with some non-local part of the world.

This seriously weakens the case for these to be representations, since the only requirement for a representation is that it cause behavior and can be stored in a memory not effected by the behavior itself. So in their view representation *may have no content*, so that a representation may be "meaningless to the overall system that uses them" (Rocha and Hordijk, 2005). The concept of representations without content seems to sap representations of their intuitive appeal and rational power in explaining behavior.

Something seems wrong with rejecting the "standing-in" relationship. First, let us use a classic example of representation: You are out in town and thinking of your cat. You remember that your cat is hungry so you buy her a tin of food. The cat you invoke from memory while shopping is a representation that is classically considered **implemented** on neurons. However, you do not think of "neurons" or "nothing" while shopping for your cat; you think of "your cat" and neurons are firing. Although this is an appeal to introspection, there is something deeper going on here. There is an effective relationship between some configuration of some neurons in memory that **constructs** a "mentalese cat" and your **behavior** of buying the cat food. This all works because your "mentalese cat" **represents** your cat at home so your behavior **corresponds** to your cat.

One could set a re-occurring alarm on your computer that causes a noise reminding you to feed your cat. A representation of the cat inside the computer causes the noise to be emitted. This representation is **implemented** in the electrical charges in the memory of the computer. When invoked,

memory can **construct** a "computerese cat." This representation in turn causes a **behavior**, the noise. This behavior is hopefully effective in virtue of the "computerese cat" having a **correspondence** with the hunger of a very real cat.

Abstracting from our examples, there is a **representational cycle**. Symbols that are physically implemented in memory **construct** a dynamic process that causes **behavior**. This behavior is effective in virtue of **corresponding** to some non-local part of the world.

Rocha and Hordijk base their theory of material representation on DNA, so the representational cycle above may not hold. DNA causes the the **construction** of protein sequences, that then cause the **behavior** of the phenotype in combination with environmental factors. Does the DNA *represent* the protein? Does the protein *represent* the behavior? Where's the content? It is simply unclear. The relationship almost seems too close to be non-local, as every step is effective, and nowhere is a correspondence to be found. While DNA, or "genic" representations, may be some sort of representations, they are far removed from the clear cases of representation used by Newell and others that serve as the assumption of so much work in cognitive science.

Their experiments with cellular automata are exciting yet inconclusive as regards representations. There are some cellular automata, given an initial configuration and update rules, that **construct** particles and these particles **behave** like certain logical rules. Can an abstract logical rule become local, then disengage, and become non-local? The matter is confusing and depends on one's belief in Platonism. It seems like while logical rules may be the right sort of thing to prove evolution of non-trivial *computation*, they are the wrong sort of thing to prove evolution of *representation*. Rocha and Hordijk notes when observing the inability of the cellular automata to create dynamically incoherent memory, "our stumbling block was in obtaining the necessary separation between representations and content" (Rocha and Hordijk, 2005). So perhaps the "standing-in" relationship was too hastily rejected?

The Correspondence Condition

To resolve this problem we propose a new requirement to be added to those given earlier, which we call the *correspondence condition*. While this is in no way a new requirement, it is a rephrasing of the "standing-in" relationship in terms that are more compatible to artificial life. This condition has been stated earlier as having effective behavior due to some correspondence with a non-local part of the world. To rephrase it in terms of artificial life, the **correspondence condition** is the dynamic behavior caused by operation of the construction code that corresponds to a non-local process.

By **non-local** we assert that the dynamic behavior caused by the representation in memory must be dynamically incoherent in terms of the current state of the non-local dynamic

process. In other words, when I imagine my cat getting run over by a car, my actual cat does not get run over by a car.

Second, by **correspondence** we mean the dynamic behavior caused by the construction code should have some covariance with the state of the non-local dynamic process, likely the state of the non-local dynamic process at the time of its *storage* into memory. I may be wrong or right about my cat being hungry, but I get the food for the cat because of my actual cat, not because of any other reason.

Finally, it must avoid *representational leakage*, in which everything can potentially be considered representational. There must be a separate input code that inputs, via some set of possibly evolved relevance criteria, those things to be stored in memory.

The Representational Cycle

At this stage representationalism is still a “hard” problem because the correspondence is physically non-local. However, I believe that the process can be naturalized, since the process is only non-local at the time of the deployment of the construction code. If the process is viewed from a larger time scale, everything can be explained in terms of local cause and effect. This **representational cycle** is in a large part a rephrasing of Brian Smith’s concept of registration (Smith, 1995).

1. **Presentation:** Process *A* is in effective local contact with process *B*.
2. **Input:** The process *A* possesses dynamically incoherent memory M_a . An input procedure of *A* puts M_a in correspondence with process *B*. This is entirely non-spooky since *A* and *B* are in effective local contact.
3. **Separation:** Processes *B* and *A* change in such a way that the processes are non-local.
4. **Output:** Due to some local effect in process *A*, it activates a construction code that uses M_a to initiate the local dynamic behavior.

Does the representational cycle presented here pose a “hard” problem or an easy problem for artificial life? Does it allow any more progress on the evolution of representations that, unlike those genic representations based on DNA, are recognized across disciplinary boundaries as representational? While it could be argued that introducing the correspondence condition a priori defeats the purpose of Rocha and Hordijk to evolve exactly such as requirement, this would be insincere since unless the requirements of an simulation are known, then it is unclear how to judge its success. Also, the correspondence condition was explicitly jettisoned by Rocha and Hordijk from their experiment. Furthermore, evidence from their experiment and our arguments show a

radical re-interpretation of representations without the correspondence condition to be unfaithful to the understanding of representations used in much of cognitive science. If Rocha and Hordijk are satisfied with evolving mere DNA-like representations, then these genic representations should be acknowledged as far removed from the symbolic systems of Newell. Lastly, if the representational cycle or something like it is already well-known, then it is time for artificial life to tackle the challenge of evolving it.

There is nothing physically spooky in the representational cycle, once the cycle is viewed in the proper time-frame, so it does not challenge the philosophical framework of naturalizing the mind under physics. It merely shows how the reference relationship between a representation and its content is time-like separated, and so not a case of action at a distance between space-like separated objects. This cycle may challenge the methodological frameworks employed by artificial life. Like Rocha and Hordijk, the cycle explicitly rejects the fact that all processes are *dynamic* all the time. These processes are both dynamic and stable to a mixed extent - in other words, more like *objects* (Smith, 1995). The proposal is far from perfect, for the notions of “effective physical contact” and “locality” need to be explicated further and simulated.

This cycle requires the evolution not just of memory and a construction code, but of “non-locality” and an input code. Given a limited amount of memory, only a finite amount of information can be represented, so the input mechanism must rely on some notion of *relevance* that also must be evolved. At the current time cellular automata may be too simple of a model to suffice. After all, the creation of dynamically incoherent memory is difficult by itself (Rocha and Hordijk, 2005). Furthermore, the construction code and input code correspond to the input and output of a Turing machine upon a tape. The existence of input and output procedures were *assumed* by Turing when he defined Turing machines. Turing himself never successfully evolved these procedures from simpler components. No one ever said evolving representations was going to be easy.

References

- Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist’s challenge in cognitive science. *Cognitive Science*, (22):295–318.
- Bobrow, D. and Winograd, T. (1977). Experience with KRL-0: One cycle of a knowledge representation language. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 213–222.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3):200–219.
- Chomsky, N. (1980). *Rules and Representations*. Columbia University Press, New York City, NY.

- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. MIT Press, Cambridge, MA.
- Crutchfield, J. and Hanson, J. (1993). Turbulent pattern bases for cellular automata. *Physica D*, (69):279–301.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press, Cambridge, MA.
- Fodor, J. (1975). *The Language of Thought*. MIT Press, Cambridge, MA.
- Grush, R. (2003). In defence of some Cartesian assumptions concerning the brain and its operation. *Biology and Philosophy*, (18):53–93.
- Haugeland, J. (1981). Analog and analog. In *Mind, Brain, and Function*. Harvester Press, New York, NY.
- Haugeland, J. (1991). Representational genera. In *Philosophy and Connectionist Theory*, pages 61–89. Erlbaum, Mahwah, NJ.
- Levensque, H. and Brachman, R. (1987). Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, (3):78–103.
- Maturana, H. and Varela, F. (1973). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing, Dordrecht.
- Mitchell, M. (1998). A complex-systems perspective on the computation vs. dynamics debate in cognitive science. In *Proceedings of the Conference of the Cognitive Science Society*, pages 710–715.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, (4):135–183.
- Newell, A. and Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, 19(3):113–112.
- Rocha, L. and Hordijk, W. (2005). Material representations: from the genetic code to the evolution of cellular automata. *Artificial Life*, (11):189–214.
- Rumelhart, D. and J. McClelland (1986). *Parallel Distributed Processing*. MIT Press, Cambridge, MA.
- Smith, B. C. (1995). *The Origin of Objects*. MIT Press, Cambridge, MA.
- Winograd, T. (1972). *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Woods, W. (1975). What's in a link: the foundation for semantic networks. In *Readings in Knowledge Representation*. Morgan Kaufmann.