



Contents lists available at ScienceDirect

## New Ideas in Psychology

journal homepage: [www.elsevier.com/locate/newideapsych](http://www.elsevier.com/locate/newideapsych)



# The biological foundations of cognitive science<sup>☆</sup>

Mark H. Bickhard<sup>\*</sup>

17, Memorial Drive East, Lehigh University, Bethlehem, PA 18015, USA

### ARTICLE INFO

*Article history:*

Available online 27 May 2008

*Keywords:*

Psychological theory  
Neuropsychology  
Emergence  
Evolutionary psychology  
Normativity  
Epistemology  
Dynamic systems  
Interactivism

### ABSTRACT

Cognitive Science originated in reactions against behaviorism that were motivated in significant part by the example of the computer. The computer raised the exciting possibility that mind could be understood almost entirely independently of brain: if the operations of the mind are akin to the execution of a program, then almost all the relevant aspects of mind would be captured by that program, independently of whatever was running it, be it transistors or neurons. This presumed independence of cognitive science from biology has waned considerably in recent decades, but in this paper, I argue that there is at least one crucial aspect of biology that has yet to be appreciated for its relevance to mental and other normative processes—the thermodynamics of living systems. In particular, I argue that the emergence of normativity in general—and normative function and representation in particular—depends on special systems that are far from thermodynamic equilibrium; these form the interface between the factual world of atoms and molecules and the normative world of mind. The nature of that emergence, in turn, imposes strong constraints on how the central nervous system functions, and, therefore, on how cognition is realized.

© 2007 Elsevier Ltd. All rights reserved.

## 1. Introduction

One of the founding assumptions that Cognitive Science drew from computer science was that mental phenomena can be modeled relatively independently of the brain. Mental phenomena are assumed to be akin to the execution of a computer program, and program properties are independent of the

<sup>☆</sup> Thanks to Robert Campbell for very helpful comments on an earlier draft.

<sup>\*</sup> Tel.: +1 610 758 3633.

E-mail address: [mark@bickhard.name](mailto:mark@bickhard.name)

URL: <http://bickhard.ws>

specifics of the computer that is running that program. Similarly, mental properties are assumed to be relatively independent of the brain in which they are realized. Furthermore, the body is “just” a robot body housing that computer, and, perhaps, incorporating the sensory transducers that feed inputs into the computer. Recognizing that the distinction between program and computer might provide a model for the relationship between mind and brain was, in fact, one of the early sources of excitement in the field (Franklin, 1995; Gardner, 1987).

The re-emergence of connectionist models in the 1980s was accompanied by claims of greater faithfulness to actual brain organization and process compared to what is obtainable with computational models that manipulate symbols (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). But while symbolic computationalism and connectionism differ in architecture, they share the underlying assumption that, while the brain may realize the relevant processes to constitute mind, the critical properties of those processes can be still be modeled independently of neural level details in the brain.

More recently, dynamic systems and autonomous agent approaches have argued for the importance not only of specifics about the brain, but also of details of the body. Cognition, in this view, is inherently embodied, and therefore cannot be understood in the disembodied forms provided by either computationalism or connectionism (Beer, 1995; Maes, 1990).

Dynamic and agent approaches, however, have also reintroduced basic questions about what the aim of modeling in Cognitive Science should be. Many have argued that mental phenomena such as representation are simply not relevant to either the design or the modeling of embodied autonomous agents. Representation, they claim, is a vestigial legacy from computationalism that should be rejected (Brooks, 1991). Dynamics are everything, so it is held, and getting them right, whether in design or in modeling, is the only relevant criterion (van Gelder, 1995).

It is clear that the relationships between cognition and the brain, and cognition and the body more broadly, are far from settled. I wish to argue that there are, in fact, close relationships between brain and bodily processes, on the one hand, and mental phenomena, on the other. Mind is not strongly independent of brain or body, as computationalism and connectionism would have it, nor is mind to be eliminated in favor of brain or body, as championed by some dynamicists. Instead, mental processes emerge in particular kinds of brain and bodily processes—and in biological processes more broadly.

## 2. The emergence of function and representation

The biological foundations of cognition are those biological properties and processes that are essential for its emergence. I will explore the implications for such foundations of an approach to Cognitive Science called *interactivism*. Interactivism is, in the first instance, a model of the nature and emergence of representation, and, therefore, of cognition. The emergence of representation, in turn, depends on the emergence of normative function. So, the model to be outlined is that of the emergence of normative function, followed by the emergence of representation as a particular kind of function; the biological foundations are those crucial to each step of emergence. Interactivism is in important respects an instance of dynamic and autonomous agent models, but differs in basic ways from standard positions in both camps. Correspondingly, it offers a novel range of implications for the biological foundations of cognition.

My focus here is on those biological foundations. The interactive model itself has been presented elsewhere (Bickhard, 1993, 1996–1998, 2001–2003, 2004a, 2004b, 2005; Bickhard & Richie, 1983; Bickhard & Terveen, 1995; Campbell & Bickhard, 1986), so I will present a highly condensed version of the model and the related arguments—hopefully, just enough to support the exploration of the implications for biological foundations.

Consider a far-from-equilibrium system, such as a candle flame or Bénard cells in a pan of water. Bénard convection cells form when there is sufficient heat differential between the top and the bottom of a layer of water. The cells will persist so long as the heat differential is maintained. The maintenance of the heat differential, in turn, is dependent on whatever external processes are creating it, such as a fire or electrical heat source underneath the pan holding the water. The crucial point for my current purposes is that the maintenance of this far-from-equilibrium system is dependent on processes external to it.

The candle flame is an example of a different kind of system property. The flame makes contributions to its own maintenance through time. It maintains temperatures above the combustion threshold; in a standard gravitational field and atmosphere, it induces convection, which both brings in oxygen and eliminates combustion products, and it vaporizes wax into fuel. A candle flame, like the Bénard cells, is a far-from-equilibrium system. It also depends for its continued existence on the maintenance of particular properties and processes that support its far-from-equilibrium conditions. Unlike the Bénard cells, however, the flame is *self-maintenant*—it helps to maintain its own far-from-equilibrium maintaining conditions (Bickhard, 1993).

The candle flame, however, can do nothing if it is running out of candle. The flame cannot adopt relevantly differing self-maintaining interactions with its environment when that environment changes. But some far-from-equilibrium systems can. A science-fiction candle flame could do that if it could seek new fuel when it detected that the candle was almost gone. A paramecium, to take a non-science-fiction example, is capable of swimming, and continuing to swim, so long as it is swimming up a sugar gradient, but will tumble for a moment if it “finds itself” swimming down the sugar gradient (Campbell, 1974, 1990). Such systems can alter their methods of self-maintenance in ways appropriate to their current environments. They tend to maintain their own property of being self-maintenant: they are *recursively self-maintenant*.

Self-maintenant and recursively self-maintenant systems are, I argue, the key forms of far-from-equilibrium systems in which normative function and representation emerge, respectively (Bickhard, 1993). I will outline that model and some of the arguments in its favor, then turn to examining its implications for biological foundations.

### 2.1. Normative function

Simply put, the heat of a candle flame serves a function for the flame insofar as it contributes to the flame’s maintenance. Function, in this view, is a contribution, or a tendency to contribute, to the maintenance of a far-from-equilibrium system—so function is always relative to some such system (Bickhard, 1993; Christensen & Bickhard, 2002). The heart of a parasite, for example, serves a function for the parasite, while being dysfunctional for the parasitized host.

This model of function stands in sharp contrast to the dominant conception of function in terms of evolutionary selection history (Godfrey-Smith, 1994; Millikan, 1984). According to such history-oriented conceptions, a heart has the function of pumping blood, and not, say, that of making heart beat sounds, because it is only in virtue of evolutionary predecessors having pumped blood that this heart exists in the first place. Having a function consists in having the right kind of etiology, the right kind of evolutionary history.

I will not focus on the details of such etiological models of function. Here I wish to simply point out that an etiological model entails that nothing can have a function unless it has the right history—whether it contributes to self-maintenance or not. Millikan (1984) accepts the consequence that if a lion were to magically pop into existence that was molecule-for-molecule identical to a lion in the zoo, the science fiction lion’s heart would not have a function because it would not have the proper history. Accepting such examples that will never happen might be a price worth paying if the etiological approach satisfied all other desiderata. Unfortunately, however, what the case of the science-fiction lion demonstrates is that the etiological approach cannot model function in terms of current state of a system. For only the current state can be causally efficacious. The lions are dynamically and causally identical, but one has organs with functions while the other doesn’t.

For a case that is not science fiction, consider the first time that a contribution is made by some organ to the persistence of an organism. The etiological theory maintains that if that contribution has been selected for over enough generations, the organ may come to have that contribution as a function (Godfrey-Smith, 1994). So the organ in a later generation may be causally identical to the ancestral organ from the first generation—yet the first-generation organ will *not* have that contribution as a function. Again, whether something has an etiological function or not makes no causal difference.

The etiological approach, therefore, yields a model of function that is unavoidably epiphenomenal. Nothing causal in the world depends on the presence or absence of having a function; that is, on the presence or absence of having the proper evolutionary history. What does make a causal difference is

current state: Does the heart pump blood, or not? Yet current state is not sufficient for function on the etiological view.

Clearly, the model of function as contribution to self-maintenance is definable in terms of current state—that is how it has *been* defined—so it is in no danger of epiphenomenality.

## 2.2. Representation

Function emerges in self-maintenant systems. Representation emerges in recursively self-maintenant systems. Here is the basic idea. *Selecting* some process of interaction between the system and its environment has the *function* of contributing to the self-maintenance of the system, but *the selection could be in error*. That selection involves a kind of anticipation about the environment: that the current environment is of the kind for which the selected processes are appropriate in helping to maintain relevant far-from-equilibrium conditions. The paramecium “anticipates” that swimming is the right thing to do, and it could be wrong—it could be swimming up a saccharin gradient, for example, not a sugar gradient.

This, I claim, is the most primitive level at which representational truth value emerges. Recursively self-maintenant selections involve anticipations that can be false. Selecting implicitly involves predicating—e.g., predicating of this environment that it is of a type appropriate for swimming—and the predicating can be false. Representational truth value emerges quite naturally in the solution to the evolutionary problem of action selection.

The forms of representation that I have discussed so far are quite primitive. They do not look much like paradigm cases, such as representations of objects or numbers, which poses a challenge: how can the interactive model account for representations beyond the simple cases of predicating types of interactive appropriateness to current environments? That question, and many others, are addressed elsewhere (Bickhard, 1993, 1999, 2000, 2001, 2004a, 2005, 2006, 2007a, in press, in preparation; Bickhard & Terveen, 1995; Campbell & Bickhard, 1986). Here I will merely indicate the direction of response.

The selections of a recursively self-maintenant system necessarily involve some sort of sensitivity to the environment. The paramecium must detect that the current environment is of a type usually appropriate for swimming. Selecting, then, is conditional on prior detecting—on outcomes of earlier engagements with that environment. Such engagements may be relatively simple and passive, but they may also be potentially complex and fully interactive themselves: the internal outcome of an *interaction* can serve to detect a relevant environmental type as well as passive input processing can—if not better.

The relevant implication here is that interactions are not restricted to being selected and being indicated as potential selections. Interactions can also differentiate environmental types for subsequent indications of the appropriateness of interactions. Selections, then, can be based on indications of potentiality, and indications of potentiality can be *conditional*: if *this* interactive outcome has occurred, then *that* interaction is now indicated as appropriate.

Conditional indications of interactive potentiality are the key to more complex forms of representation. Such conditional relationships can branch—a single interaction outcome can function to indicate multiple further interactive potentialities. They can iterate—completing interaction A may indicate the potentiality of B, which, if completed, would indicate the potentiality of C, and so on. That is, they can form potentially quite complex webs of conditional interactive potentialities, and particular kinds of such webs, I argue, yield the emergence of higher kinds of representation, such as of manipulable objects (Bickhard, 1980, 1993, 2001, 2004a; Piaget, 1954). Representing abstractions, such as of numbers, requires still further development (Campbell & Bickhard, 1986).

## 2.3. Correspondence models of representation

For current purposes, I will take the interactive model of representation to be a viable contender. This model of representation, like the interactive model of function, disagrees sharply with dominant approaches. A standard information-semantic approach to representation will construe interactive differentiations or detections as representations of the properties that they differentiate or detect. When

a recursively self-maintaining system accomplishes such differentiations, which are necessary if its subsequent selections are to be appropriately sensitive to the environment, the system does create informational, and perhaps causal, correspondences with whatever properties it has detected. But the interactive model just needs detections *in fact* of properties that are relevant to further indications of interactive potentiality; it does *not* require that those properties be *represented*. Whereas it is precisely such detections or differentiations that are standardly construed as being representations of whatever properties they have differentiated, particularly when the differentiations result from passive processing of inputs of the sort that allegedly occurs in perceptual systems (Bickhard & Richie, 1983).

Correspondence models construe representation as looking backward in time, down the input stream, toward some privileged locus in that input stream, such as the surface of a table reflecting light into the eye. The interactive model construes representation as future oriented, toward the potentialities for further interaction that are afforded by the current environment, and the conditional webs of potential interactions that can extend indefinitely in space and time.

Correspondence models of representation suffer from a host of fatal problems (Bickhard, 1993, 1996, 2001, 2004a, 2006, 2007a, *in press*, *in preparation*; Bickhard & Terveen, 1995). Only some need outlining here. The first is that correspondence-based models cannot account for their own foundations. It is clear that not all informational or causal correspondences are representations. In fact, hardly any are: the universe is full of such correspondences, because every single instance of every causal law yields one. A correspondence model has to specify what special *kind* of correspondence qualifies as representational—what makes the purportedly *representational* correspondences so different from all the others. There are multiple candidates for these extra criteria to be imposed on correspondences (Bickhard, 1993, 2004a; Bickhard & Terveen, 1995; Fodor, 1990, 1998; Stich & Warfield, 1994), but none of them provide a naturalized model of the *content* for these alleged representations.

Representational content is what a representation is *supposed* to represent. It is the normative aspect of representation. It is what makes it possible for a representation to be in error: if a content of “cow” is predicated of what is in reality a horse, then that predication is false.

Correspondence models are in a serious bind. If the special kind of correspondence exists, then the representation exists, and that representation is correct. But if the special correspondence does not exist, then the representation does not exist—and it cannot be false. Information semantics, and other correspondence-based models, have grave difficulties accounting for representational error (Fodor, 1990; Levine & Bickhard, 1999; Loewer & Rey, 1991). Which means they cannot account for representational content.

Such models can explain complex representations as the result of combining more basic representations. The more basic representations might, in their turn, be constructed by combining still more basic representations. But this composition and decomposition process must have a ground or foundation, consisting of representations that are neither constructed out of any other representations, nor defined in terms of them. Yet this foundation cannot be accounted for. The models permit new representations to be created only from already available representations. There can be no account of newly emergent representation. (Even if a new *correspondence*—a new differentiation—were created, perhaps in a connectionist network, there is no account of what those correspondences are *supposed* to represent—hence no way of getting any representational content.)

One move has been to posit that all basic representations must be innate, and all representations that, say, adult human beings are capable of are built as complexes out of those innate atoms (Fodor, 1981). But the problem here is a logical one: there is no model of how representations by correspondence can come into existence, or of how they could get any content once they came into being, and evolution is just as helpless in the face of that problem as individual level learning and development (Campbell & Bickhard, 1987). Correspondence models, in other words, presuppose a foundational level of representation that they are incapable of modeling, because in their own terms it is impossible. They involve an internal contradiction—an incoherence (Bickhard, 1993).

Correspondence models render the emergence of new representation impossible. If they were correct, neither evolution nor development nor learning could ever get representation to emerge. But there were no representations at the moment of the Big Bang and there are representations now. Therefore, representation has emerged. Therefore, representational emergence is possible. Therefore,

any model that renders representational emergence impossible is false. Therefore, correspondence-based models are false.

By contrast, the interactive model makes representational emergence so natural as to border on trivial: any new construction of system process organization that involves the right kinds of indications of interactive potentiality will constitute an emergent representation.

### 3. Biological foundations

#### 3.1. Far-from-equilibrium systems

The clearest implication of the interactive model for biological foundations is that genuine representation and genuine cognition can emerge only in far-from-equilibrium systems—more precisely, recursively self-maintaining systems. It is only in far-from-equilibrium systems that function can emerge, and only a certain kind of interactive indicational function that can constitute emergent representation.

This is already a major shift from computer models or connectionist conceptions. If the interactive model is correct, various properties of cognition can be *simulated* in computers and in connectionist nets, but not genuinely *instantiated*. Such models suffer from key deficiencies. First, they are not self-maintaining systems. By most standards, they are not even far-from-equilibrium systems, though their dependence on externally supplied power does keep them some distance from thermodynamic equilibrium.

#### 3.2. Open systems

A second is that computers are not in any relevant sense open systems. They process inputs or data, but do not interact with their environments. Recursive self-maintenance requires environmental interactions that achieve closure in the sense of circling back to support the far-from-equilibrium conditions that made those interactions possible in the first place (Christensen & Hooker, 2000; Ruiz-Mirazo & Moreno, 1998).

#### 3.3. Interactive timing

A third issue is timing. Successful interactions require appropriate timing. Mere speed is not sufficient; an interaction can fail from being too fast just as easily as from being too slow. Computers do have timing information in their internal clocks, but there is no timing in the Turing machine formalisms (and equivalents) for computers (Bickhard & Richie, 1983; Bickhard & Terveen, 1995), and the way timing is engineered into digital computers, by wiring in a central clock, is not an evolutionary possibility. If the brain involved clock-driven processes akin to those in a computer, every evolutionary change in the brain would have to have involved simultaneous adaptive changes in the interactive circuitry *and* precisely coordinated changes in the timing circuitry. That is vanishingly improbable for any instance, and simply impossible for any extended sequence of evolution (Bickhard & Terveen, 1995).

The brain solves the timing problem in a very different way: It puts clocks everywhere, and constructs all of its functional relationships out of relationships among those clocks. Since clocks are “just” oscillators, this translates into building the central nervous system out of oscillators, and constructing all functional relationships out of modulatory relationships among oscillatory subsystems. This is, in fact, the basic architecture of neural functioning.

Neurons and neural circuits are oscillatory; each has baseline levels of oscillation that are modulated by influences from other neurons and neural circuits. Some kinds of neurons never fire at all, but do modulate the activities of others. And evolution has created a virtual tool kit of modulatory influences ranging over time and space, from tiny and very fast gap junctions, to classical synapses, to volume transmitters that diffuse throughout a local population of neurons, to transmitters released in a graded instead of all-or-nothing fashion, to control of the local neural environment by astrocytes, and so on (Bickhard, in preparation; Bickhard & Terveen, 1995; Nedergaard, 1994; Nedergaard, Ransom, & Goldman, 2003; Newman, 2003; Nieuwenhuys, 2000; Ransom, Behar, & Nedergaard, 2003).

Such oscillatory and modulatory principles are at least as powerful as classical conceptions of central nervous system architecture: a limit case of one system modulating another consists of the first system switching the second on and off, and switches are enough to build computers out of. Oscillatory and modulatory principles are more powerful in that they inherently provide timing, while Turing machines do not, and the timing in computers takes a biologically impossible form.

Idealizing neurons into threshold switches as is so often done in computer modeling, or into simple activation level transformers as in connectionist modeling, is seriously unfaithful to what actually occurs in the brain (Bickhard, 2007b, in preparation; Bickhard & Terveen, 1995). The oscillatory and modulatory architecture proposed here is not logically forced by interactivist principles alone, but it is forced by evolutionary considerations, and it is consonant with actual brain processes.

### 3.4. *In reverse: From CNS functioning to representation*

Action and interaction based models of representation require timing, not just sequence. The need for timing converges with multiple facts about the way the central nervous system functions: with oscillatory processes that exert mutual modulatory influences at multiple spatial and temporal scales. Here I will focus on those modulatory influences and show how, in their turn, they generate a future-oriented, anticipatory, model of representation. If so, then the theoretical considerations about representation yield strong constraints on CNS functioning, and the facts of CNS functioning, in turn, yield the same kind of anticipatory model of representation that we began with. Each implies the other.

Modulatory influences in the CNS range from very small and very fast gap junctions between two neurons to much larger and slower volume transmitters and astrocyte influences on local populations of neurons; classical synapses are in between, though somewhat closer to the gap junctions. Processes in the CNS have dynamics going on simultaneously at all these scales, influencing one another across all the scales. Compared to the fast processes, the slower processes are relatively constant: they change at a rate that sets the parameters for the fast processes, and the parameters change over multiple intervals of the fast processes.

Such parameter setting for dynamic systems is the equivalent of programming for discrete systems. The slow processes are, in this sense, setting up the local programming, the dynamic constraints, for the fast processes. They are *microgenetically* constructing the framework for how the fast processes will proceed—ongoingly, in real time (Bickhard, 2006, 2007b, in preparation).

The larger scale modulatory influences are ideally situated to create and change local correlations among the populations of faster processes, and such local correlations are one way to induce attractor landscapes—topologies of dynamic attractors, their shapes, and their basins—for the faster processes. Microgenesis, then, will induce and modulate the dynamics of the attractor landscapes that frame the faster processes. It can be argued that the fundamental mode of neural functioning is in terms of such microgenetically constructed attractor landscapes (Bickhard, 2007b, in preparation; Freeman, 2000)—and a central theme in the macro-evolution of the nervous system (Bickhard, 2007c).

What is crucial for my current purposes is that microgenetic processes set parameters in an *anticipatory* fashion. They set the CNS up to function in one manner rather than another, constituting a functional anticipation that this manner of processing will be appropriate to the actual flow of interaction between the organism and its environment. Microgenetic anticipations, therefore, can be *false*: the environment may not cooperate with the anticipation that the ensuing interactions will proceed within the framework that has been constructed.

The multiple scales of modulatory influence in the nervous system, thus, yield a dynamic version of microgenetic programming, which, in turn, anticipates where the interactive flow will proceed in the near future. The anticipations, in their turn, carry primitive truth value: they are representational. We have now completed a full circle from representational theory to central nervous system dynamics, and back.

### 3.5. *Irreversibility and normative scaling*

One further biological foundation that I would like to consider arises from looking more carefully at the case of a computer controlling a robot. A robot can interact with the world, and so would seem to

satisfy the interactive condition. To be successful in its interactions, the computer would have to appropriately handle timing in some way, even if not in a biologically plausible way. If we suppose, for example, that among the capabilities of the robot are detecting when its batteries are running low, and seeking out power sources to replenish its batteries under those conditions, then we would seem to have, in some minimal sense, a far-from-equilibrium system. What's more, it is at least minimally self-maintenant, and even recursively self-maintenant (because it can switch into and out of "power source seeking" just as a paramecium can switch into and out of "swimming").

Could such a robot have emergent cognition? There is an evident contrast with the biological case, because most of the robot's body is *not* far-from-equilibrium, it cannot be self-maintained, and it certainly cannot be recursively self-maintained. Biological organisms have an elaborate infrastructure—organs, bones, and so on—that are themselves (mostly) far-from-equilibrium and self-maintained through a constant turnover of molecular constituents. What confers the status of *infrastructure* on some parts of the overall far-from-equilibrium system is the slower time scales, and often-higher energy levels, than are characteristic of the system's other interactions. Such infrastructure is required in order to control and enable the multiplicity of metabolic processes necessary to self-maintenance and to creating, selecting, and executing recursively self-maintenant interactions (Ruiz-Mirazo & Moreno, 1998).

The robot is far-from-equilibrium only in the sense that its operations require power. The basic existence of the robot is not far-from-equilibrium, and does not require self-maintenance or recursive self-maintenance. The robot therefore does not have much at stake in its self-maintenant activities, not at least by biological standards. What, then, is the metaphysical significance of those differences for the emergence of function and representation?

Any kind of emergent phenomenon is likely to have its marginal cases, and the robot example is clearly marginal. Its marginality is important, I contend, because it illustrates how profoundly far-from-equilibrium conditions, along with the processes that maintain them and the processes that maintain the system's ability to maintain them, permeate the emergence of function and representation in the paradigmatic biological cases. Robots can be marginally far-from-equilibrium by design, so they can be marginally self-maintenant and recursively self-maintenant. They do not, however, form an *evolutionarily* marginal case. Neither function nor representation could have evolved via such marginality.

In particular, living beings are *irreversible* far-from-equilibrium systems, while robots, in general, exhibit *reversibility*. A robot can be turned off, then on again; if its battery runs down, nothing in particular need be lost if that battery is later recharged. The thermodynamics of living systems are not reversible: If the far-from-equilibrium processes that constitute them are halted, they cannot be restarted. So it is not just the *activities* of living systems that require far-from-equilibrium conditions. Their very ontological constitution is far-from-equilibrium, and it begins to move irreversibly toward equilibrium once that far-from-equilibrium process is sufficiently interrupted or disrupted (even by the accumulation of errors during normal functioning that we call aging).

What makes far-from-equilibrium infrastructure and metabolism significant is that the normative aspect of representation emerges in functional contributions to the maintenance of far-from-equilibrium conditions. If the far-from-equilibrium conditions are of minimal importance then the normative aspects of function and representation will be, too. The degree of far-from-equilibrium status for the "body" of the system involved indirectly affects the degree of normativity. Function and representation are normative to the extent that there is something that is ontologically at stake in their "successful" functioning (Bickhard, 1993, 2006, 2007a; Christensen & Bickhard, 2002).

### 3.6. *Self-reproducing and living systems?*

There is one final point that I wish to make about biological foundations. Cognition, if the conception that I have presented is correct, requires far-from-equilibrium systems that are recursively self-maintenant, perhaps in very complex ways. There is nothing in the model, however, that requires that these systems be *self-reproducing*. It is a matter of definition, then, whether they must be *living*. (Not all animals are capable of self-reproduction, not even all animals that are capable of clear cognition; any sterile organism, be it insect, such as a bee, or mammal, or human, is a counterexample.) The

interactive model, then, allows that designed systems could be fully capable of function, representation, and cognition. But they will not attain any of these with the architectural and process resources available in computer or connectionist models.

#### 4. Conclusion

The interactive model is a version of the dynamic systems approach, and of autonomous agent approaches (Bickhard & Terveen, 1995). Unlike the other variants currently in circulation, however, interactivism neither argues against the usefulness of correspondence-based representations (Port & van Gelder, 1995) nor for their purported necessity (Clark & Toribio, 1995). Instead, it rejects the shared assumption that representation is correspondence. It argues instead that function and representation emerge naturally in the evolution of dynamic autonomous biological agents—in ways that correspondence approaches cannot account for. Interactivism reconfigures our assumptions about the biological foundations for cognition, and, therefore, for Cognitive Science.

#### References

- Beer, R. D. (1995). A dynamical systems perspective on agent–environment interaction. *Artificial Intelligence*, 73(1/2), 173–215.
- Bickhard, M. H. (1980). *Cognition, convention, and communication*. New York: Praeger.
- Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, 5, 285–333.
- Bickhard, M. H. (1996). Troubles with computationalism. In W. O'Donohue, & R. F. Kitchener (Eds.), *The philosophy of psychology* (pp. 173–183). London: Sage.
- Bickhard, M. H. (1997). Is cognition an autonomous subsystem? In S. Ó Nualláin, P. McKeivitt, & E. MacAogáin (Eds.), *Two sciences of mind* (pp. 115–131). Philadelphia: John Benjamins.
- Bickhard, M.H. (1998). A process model of the emergence of representation. In G.L. Farre & T. Oksala (Eds.), *Emergence, Complexity, Hierarchy, Organization: Selected and Edited Papers from the ECHO III Conference. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series*, 91, 263–270
- Bickhard, M. H. (1999). Interaction and representation. *Theory & Psychology*, 9(4), 435–458.
- Bickhard, M. H. (2000). Motivation and emotion: an interactive process model. In R. D. Ellis, & N. Newton (Eds.), *The caldron of consciousness* (pp. 161–178). Philadelphia: John Benjamins.
- Bickhard, M. H. (2001). Why children don't have to solve the frame problems: cognitive representations are not encodings. *Developmental Review*, 21, 224–262.
- Bickhard, M. H. (2002). The biological emergence of representation. In T. Brown, & L. Smith (Eds.), *Emergence and reduction* (pp. 105–131). Mahwah, NJ: Erlbaum.
- Bickhard, M. H. (2003). An integration of motivation and cognition. In L. Smith, C. Rogers, & P. Tomlinson (Eds.), *Development and motivation: joint perspectives* (pp. 41–56). Leicester: British Psychological Society.
- Bickhard, M. H. (2004a). Process and emergence: normative function and representation. *Axiomathes*, 14, 135–169, Reprinted from. In J. Seibt (Ed.), *Process theories: crossdisciplinary studies in dynamic categories* (pp. 121–155). Dordrecht: Kluwer Academic, 2003.
- Bickhard, M. H. (2004b). The social ontology of persons. In J. I. M. Carpendale, & U. Müller (Eds.), *Social interaction and the development of knowledge* (pp. 111–132). Mahwah, NJ: Erlbaum.
- Bickhard, M. H. (2005). Consciousness and reflective consciousness. *Philosophical Psychology*, 18(2), 205–218.
- Bickhard, M. H. (2006). Developmental normativity and normative development. In L. Smith, & J. Vonèche (Eds.), *Norms in human development* (pp. 57–76). Cambridge: Cambridge University Press.
- Bickhard, M. H. (2007). Mechanism is not enough. *Pragmatics and Cognition*, 15(3), 573–585.
- Bickhard, M. H. (2007). *The brain doesn't work that way*. Paper presented at the 4th Interactivist Summer Institute, May 28, 2007. The American University in Paris.
- Bickhard, M. H. (2007). *The evolutionary exploitation of microgenesis*. Paper presented at the 4th Interactivist Summer Institute, May 28, 2007. The American University in Paris.
- Bickhard, M. H. The interactivist model. *Synthese*, in press.
- Bickhard, M. H. The whole person: toward a naturalism of persons—contributions to an ontological psychology, in preparation.
- Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: a case study of James Gibson's theory of perception*. New York: Praeger.
- Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: impasse and solution*. Amsterdam: North-Holland.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1–3), 139–159.
- Campbell, D. T. (1974). "Downward causation" in hierarchically organized biological systems. In F. J. Ayala, & T. Dobzhansky (Eds.), *Studies in the philosophy of biology* (pp. 179–186). Berkeley: University of California Press.
- Campbell, D. T. (1990). Levels of organization, downward causation, and the selection-theory approach to evolutionary epistemology. In G. Greenberg, & E. Tobach (Eds.), *Theories of the evolution of knowing* (pp. 1–17). Hillsdale, NJ: Erlbaum.
- Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages*. Basel, Switzerland: Karger.
- Campbell, R. L., & Bickhard, M. H. (1987). A deconstruction of Fodor's anticonstructivism. *Human Development*, 30, 48–59.
- Christensen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *The Monist*, 85(1), 3–28.

- Christensen, W. D., & Hooker, C. A. (2000). An interactivist–constructivist approach to intelligence: self-directed anticipative learning. *Philosophical Psychology*, *13*(1), 5–45.
- Clark, A., & Toribio, J. (1995). Doing without representing? *Synthèse*, *101*, 401–431.
- Fodor, J. A. (1981). The present status of the innateness controversy. In J. Fodor (Ed.), *Representations* (pp. 257–316). Cambridge: MIT Press.
- Fodor, J. A. (1990). *A theory of content*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1998). *Concepts: where cognitive science went wrong*. Oxford: Oxford University Press.
- Franklin, S. (1995). *Artificial minds*. Cambridge, MA: MIT Press.
- Freeman, W. J. (2000). *How brains make up their minds*. New York: Columbia University Press.
- Gardner, H. (1987). *The mind's new science*. New York: Basic.
- Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, *28*(3), 344–362.
- Levine, A., & Bickhard, M. H. (1999). Concepts: where Fodor went wrong. *Philosophical Psychology*, *12*(1), 5–23.
- Loewer, B., & Rey, G. (Eds.). (1991). *Meaning in mind: Fodor and his critics*. Oxford: Blackwell.
- Maes, P. (1990). *Designing autonomous agents*. Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*. In *Psychological and biological models, Vol. 2*. Cambridge, MA: MIT Press.
- Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Nedergaard, M. (1994). Direct signaling from astrocytes to neurons in cultures of mammalian brain cells. *Science*, *263*, 1768–1771.
- Nedergaard, M., Ransom, B., & Goldman, S. A. (2003). New roles for astrocytes: redefining the functional architecture of the brain. *Trends in Neurosciences*, *26*(10), 523–530.
- Newman, E. A. (2003). New roles for astrocytes: regulation of synaptic transmission. *Trends in Neurosciences*, *26*(10), 536–542.
- Nieuwenhuys, R. (2000). Comparative aspects of volume transmission, with sidelight on other forms of intercellular communication. In L. F. Agnati, K. Fuxe, C. Nicholson, & E. Syková (Eds.), *Volume transmission revisited. Progress in Brain Research*, *125* (pp. 49–126). Amsterdam: Elsevier.
- Piaget, J. (1954). *The construction of reality in the child*. (M. Cook, Trans.). New York: Basic Books. (Original work published 1937).
- Port, R., & van Gelder, T. J. (1995). *Mind as motion: dynamics, behavior, and cognition*. Cambridge, MA: MIT Press.
- Ransom, B., Behar, T., & Nedergaard, M. (2003). New roles for astrocytes (stars at last). *Trends in Neurosciences*, *26*(10), 520–522.
- Ruiz-Mirazo, K., & Moreno, A. (1998). Autonomy and emergence: how systems become agents through the generation of functional constraints. In G.L. Farre & T. Oksala (Eds.), *Emergence, Complexity, Hierarchy, Organization*. Selected and Edited Papers from the ECHO III Conference. *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series*, *91*, 273–282.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing*. In *Foundations, Vol. 1*. Cambridge, MA: MIT Press.
- Stich, S., & Warfield, T. A. (1994). *Mental representation: a reader*. Oxford: Blackwell.
- van Gelder, T. J. (1995). What might cognition be, if not computation? *Journal of Philosophy*, *92*(7), 345–381.