Representation in Natural and Artificial Agents

Mark Bickhard

I will be arguing that a primitive form of representation is naturally emergent in certain kinds of far-from-equilibrium dynamic systems—in particular, in agents—and that more complex forms of representation are constructable out of these primitive forms. If this is so, then interactive—agent focused—dynamic systems theory is the proper framework for the study of cognitive phenomena.

Kinds of Dynamic Systems

The first task is to outline the relevant kinds of dynamic systems. The kind that I am aiming toward is one that is called recursively self-maintenant systems. These are a special type of dissipative system, so that is where I begin.

Dissipative systems are necessarily far-from-equilibrium. They are necessarily in interaction with an environment in order to maintain their far-from-equilibrium conditions: cessation of that maintenance will yield a system going to thermodynamic equilibrium and the far-from-equilibrium system will cease to exist. Chemicals being pumped into a chemical bath, for example, may yield very interesting interactions within that bath so long as the far-from-equilibrium conditions of the pumped influx is maintained (Nicolis 1995; Nicolis & Prigogine 1977, 1989; Prigogine 1980). For another, the heat under a pan of water can yield the emergence of Benard cells of boiling turbulence in the water, which will cease if the temperature differential between the bottom and top of the liquid is not maintained.

Some dissipative systems help maintain the conditions that are essential to their own existence. These are called self-maintenant (Bickhard 1993). A candle flame is a simple example. The flame maintains above combustion threshold temperature, vaporizes wax into fuel, and, in standard gravity and atmospheric conditions, creates convection, which both pulls in new oxygen and gets rid of combustion waste products. A candle flame, then, contributes to the maintenance of its own existence in several different ways.

Recursively self-maintenant systems are those that can function to help maintain their own property of being self-maintenant (Bickhard 1993). That is, they can adopt differing ways of contributing to their persistent existence depending on what might be useful in differing conditions. To do so, such shifts in self-maintenance processes must be appropriate to environmental shifts. There-

fore, a recursively self-maintenant system must be sensitive to those shifts in environmental conditions. The candle flame, for example, might expand to find new fuel, if it could sense that it was running out of fuel, and if the candle flame were recursively self-maintenant in that respect and not just self-maintenant per se.

Environmental Sensitivity

Such environmental sensitivity is a crucial step toward interactive representation. It is also a source of massive confusion in the standard literature, with such sensitivity often being considered to *constitute* representation (Bickhard 1993; Bickhard & Terveen 1995). The most general form of such sensitivity is via interactive differentiation. The basic idea is that, for any (sub)system interacting with various environments, some environments will yield the same final internal state as certain other environments, while still different environments will yield different final internal states. That is, the final state that such a system arrives at in interaction with an environment will serve to differentiate some possible environments together and to differentiate them from others. These differentiations are a kind of categorization into those environments yielding final state **A** (say) from those that yield final state **B**. But it is not a representation of those environments because there is no information about what kind of environment it is that will yield **A** rather than **B**, only that they differ, and that the system is currently in (say) an **A**-type environment.

The most powerful form of such differentiations are produced by full interactions with an environment, but the basic logic is the same if the weaker version in which there are no outputs is considered. In such passive differentiations, input streams alone suffice to yield differentiating internal final states. In important cases, such passive differentiations suffice.

They are of critical importance for another reason: such passive differentiations, in the form of perceptual input processing, constitute one of the primary seductions into thinking that such differentiations yield full representations. It is assumed that the products of visual sensory processing, for example, yield encodings of the world that reflected the light signals being processed. It is assumed that the *factual* correspondence between system states and environmental states that is created by such differentiation constitutes an *epistemic* correspondence in which the system states are somehow representations of those environmental states (Bickhard & Terveen 1995; Fodor 1987, 1990a, 1998). There are myriads of problems with such an assumption (Bickhard 1996b, 1997b; Brooks

1991; Dartnall 1997; Fodor 1990b; Loewer & Rey 1991; Port & van Gelder 1995; Prem 1995; Shanon 1993). For one example, if such a correspondence exists, then the presumed representation exists, and it is correct, but if the correspondence does not exist, then the representation does not exist, and it cannot be incorrect. Accounting for the possibility of representational error on such correspondence accounts of representation has proven to be a vexing—I argue an impossible—problem (Bickhard 1993, 1996b). In any case, the interactive model does not depend on such an assumption that factual correspondences, set up by differentiations, do anything more than differentiate. In particular, the interactive model does not assume that differentiations constitute representations.

Instead, the model points out that such differentiations, even if not representational per se, can nevertheless be useful in contributing to the system shifting its self-maintenant interactions in ways that are appropriate to those differentiations (Bickhard 1980, 1993, 1996b). The system does not need to have any representation of environments of type **A** in order to be functionally capable (whether innately or via learning) of indicating that, when type **A** environments are encountered, then the system should engage in, or at least *can* engage in, processes of type **Q**.

The simplest form of such differentiation dependent shifting is triggering. The outcome of a differentiation process simply triggers the initiation of the appropriate self-maintenant interactive process. Triggering, however, does not suffice in all cases. In particular, when the environment is not sufficiently reliable in its responses to the interactions, the self-maintaining contributions may not occur. Such unreliability can occur either if the environment itself is too complex or stochastic, or if the system isn't engaging in the right differentiations (hasn't learned the best differentiations) in order to yield the most reliable triggerings. Whatever the source, such unreliability puts a premium on the system being able to monitor its interactions to determine if the appropriate consequences have in fact occurred.

Anticipation

The solution to unreliability, then, is for the system to anticipate the outcomes of its interactions in order to be able to check to see if they occur. These anticipated outcomes must be outcomes internal to the system itself. Otherwise we encounter the problem of detecting them, perhaps of representing them, and representation is what the model is trying to explicate. If the anticipated outcomes are internal, they can be functionally checked, and no conceptual problem obtains. Pointers, for example, to anticipated internal outcome states could support such anticipatory processes (Bickhard & Terveen 1995).

Interaction Selection

Conversely, if such anticipations of interaction outcomes are available, they can be used in the selection of interactions: select interactions with anticipated outcomes that are relevant to current goals.[1] If the anticipated outcomes do not occur, then do the interaction again, or do something different—or engage in learning processes if they are available. Anticipated outcomes, then, permit the selection of interaction by the system, and permit feedback to the system about the successes or failures of those interactions (Bickhard & Terveen 1995).

Selected interactions by agents are called actions, and the problem of interaction selection is 'just' a general version of the problem of action selection. Action selection, then, in complex versions, should occur via indicated anticipated outcomes.

Any agent of more than minimal complexity will need to select actions and interactions—triggering does not suffice as a general strategy. The general action selection problem, therefore, is common to living agents and to artificial agents. It is the bridge between representation in natural and in artificial systems.[2]

Nevertheless, although the model to this point has functional anticipations based on environmental differentiations, and action selections based on func-

---

[1] If the concept of 'goal' needed here were itself necessarily representational, then this would be a point of circularity in the model. Representation would be explicated in terms of goal, which is itself already representational. Representational goals, however, are not necessary here. Something closer to internal conditions for switching suffice. For more extensive discussion of this issue, see Bickhard (1993) and Bickhard & Terveen (1995).

[2] The relevant literature is voluminous. See, for example, Beer (1990, 1995a, 1995b); Beer, Chiel, Sterling (1990); Beer, Gallagher (1992); Bickhard (1996a, 1997a); Brooks (1990, 1991a, 1991b, 1991c); Cherian & Troxell (1995a, 1995b); Clark (1997); Cliff (1991); Maes (1990a, 1990b, 1991, 1992, 1993, 1994); Malcolm, Smithers, Hallam (1989); Malcolm & Smithers (1990); Nehmzow & Smithers (1991, 1992); Pfeifer & Verschure (1992a, 1992b).

tional anticipations, there is as yet no model of representation per se. In particular, there is no account thus far of representational truth value—of being representationally correct or incorrect.

System Detectable Truth Value

The anticipations on which interaction selections are based can be in error. The anticipated outcomes may not occur. And such errors can be detected by the system itself—by checking if the indicated outcomes obtain. This is a system detectable error of anticipation.

That is, indications that interaction **X** will yield outcome **Y** in the current environment can be false, and can be detected to be false by the system. We have the emergence of system detectable *truth value.*

Environmental Predications

An indication of an interaction and its anticipated outcomes predicates of this environment that it—the environment—will yield those outcomes if that interaction is engaged in. This is a predication that has emergent truth-value—it is true or false. A predication with truth-value constitutes at least a simple form of *representation.*

Such a predication attributes to the environment whatever those properties are that would in fact support the indications if the interaction were engaged in. These properties are implicit in the predication, not explicit. Representational content in this model is implicit, not explicit.[3] The predication does not specify what any of those properties are, yet it predicates them implicitly.

What About Objects?

We can find the simplest interactive representations, representations of potential interactions and their anticipated outcomes, in very simple organisms and agents. What about more complex kinds of representations, such as objects, or numbers? How can the interactive model account for them?

The general answer is "Piaget", but making that connection requires a little contextualizing. The interactive model of representation construes representation as emergent out of action (Bickhard 1993, in press-a, in press-b; Bickhard & D. Campbell, forthcoming; Bickhard & R. Campbell 1989). It thereby makes contact with the pragmatist tradition (Rosenthal, 1983, 1990), and, in particular,

[3] Implicitness solves and dissolves many problems in epistemology. For some explorations, see Bickhard (1993, forthcoming), and Bickhard & Terveen (1995).

with Piaget (Piaget 1954). Piaget also modeled representation as emergent out of action systems, though the specifics of his model differ from the interactive model (Bickhard & Campbell 1989; Campbell & Bickhard 1986). Nevertheless, his models of how complex representations can be constructed out of basic action systems are available.

Making this connection requires outlining how interactive representations can attain complexity at all, instead of being merely single indications and anticipations. The first step is to recognize that indications can be multiple. One differentiation can indicate many possible interactions and associated outcomes. Second, indications can iterate. If an indicated outcome is obtained, then other interactive possibilities may be available. Putting together such branchings and iterations, we get the possibility of linking such anticipations into complex webs, with each indication connecting to further indications should the focal indication be executed and its outcomes attained.

Some of these web organizations will have the crucial properties of closure and reachability. That is, all of the actions within such a web will yield conditions that are still in that web—closure—and every state in such a web can be attained or reached from every other state upon completion of appropriate intermediate interactions—reachability. Still further, some of these closed and reachable webs of interactive indications will exhibit the further property of being invariant under classes of other interactions. That is, the webs will not change even when various other kinds of interactions are engaged in.

For example, a child's toy block offers a complex web of possible eye scans, manipulations, pushing, chewing, and so on. All such actions will leave the possibilities of the block unchanged, and therefore in the web—it is closed. All such states are reachable from any other via the appropriate intermediate manipulations, rotations, and so on. And the whole web is invariant under all such actions, and under still others such as leaving in the toy box, walking into the other room (which will introduce having to walk back in order to re-attain the immediate interactive possibilities), and so on, but it will *not* be invariant under such interactions as crushing or burning (Bickhard 1980).

This is a basically Piagetian model of (manipulable) object representation. Similarly Piagetian models of abstractions such as numbers are also available (Piaget 1954). For current purposes, the key point is that such representations as of objects and numbers do not constitute aporia for the interactive model.

Representing is Dynamic

Representation emerges with truth-value, and truth-value emerges in a special kind of anticipatory function.[4] In this view, representing is an activity, a process (Bickhard 1993, in press-a; Bickhard & Terveen 1995; Hooker 1996). Representations as entities or states can be derivative from this basic anticipatory functional process, but are not foundational themselves.

In general, then, representation is emergent in a particular class of dynamic systems. Representation is inherently dynamic, and *cannot* be properly understood in a non-dynamic, non-interactive, framework.[5] Interactive dynamic systems form the proper framework for attempting to understand representation and representational phenomena. That is, the study of autonomous agents is the proper framework for the study of cognition.

---

[4] It is important to note that the normativity of representation in this model is derivative from the normativity of function—the function of contributing to the maintenance of the far-from-equilibrium conditions that maintain the system. This paper has not focused on function per se, and it might appear that simple mechanical systems—not dissipative systems—could realize the interactive model as outlined. That is correct, except that all representational normativity would be lost. No mechanical system has any intrinsic normativity about true representation as distinct from false representation. It is the asymmetry between continued far-from-equilibrium existence and the terminal "relaxation" into equilibrium that yields that basic normative emergence. For further discussion, and alternative models of function, see Bickhard (1993); Christensen (1996); Christensen, Collier, Hooker (in preparation); Cummins, (1975); Godfrey-Smith (1994); Millikan (1984, 1993).

[5] The arguments for this negative point are vast in scope and cannot be surveyed here. The difficulty in accounting for error that was mentioned earlier is just one of a multifarious array of such problems for non-interactive models. See Bickhard (1980, 1993, 1996b, in press-a, in press-b), Bickhard & Terveen (1995).

References

Beer, R. D. (1990), *Intelligence as Adaptive Behavior*. Academic.

Beer, R. D. (1995a), "Computational and Dynamical Languages for Autonomous Agents", in: R. Port, T. J. van Gelder (eds.) *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge, MA: MIT Press, 121–147.

Beer, R. D. (1995b), "A Dynamical Systems Perspective on Agent-Environment Interaction" in: *Artificial Intelligence, 73*(1/2), 173.

Beer, R. D., Chiel, H. J., Sterling, L. S. (1990), "A Biological Perspective on Autonomous Agent Design", in: P. Maes (ed.) *Designing Autonomous Agents*. Cambridge, MA: MIT Press, 169–186.

Beer, R. D., Gallagher, J. C. (1992), "Evolving Dynamical Neural Networks for Adaptive Behavior", in: *Adaptive Behavior, 1*(1), 91–122: "evolved" chemotaxis and six-legged walking with genetic algorithms.

Bickhard, M. H. (1980), *Cognition, Convention, and Communication*. New York: Praeger.

Bickhard, M. H. (1993), "Representational Content in Humans and Machines", in: *Journal of Experimental and Theoretical Artificial Intelligence, 5*, 285–333.

Bickhard, M. H. (1996a), "The Emergence of Representation in Autonomous Embodied Agents", in: Papers from the *1996 AAAI Fall Symposium on Embodied*
Cognition and Action. Chair: Maja Mataric. Nov 9–11, MIT, Cambridge, MA. Technical Report FS-96-02. Menlo Park, CA.: AAAI Press.

Bickhard, M. H. (1996b), "Troubles with Computationalism", in: W. O'Donohue, R. F. Kitchener (eds.) *The Philosophy of Psychology*. (173–183). London: Sage.

Bickhard, M. H. (1997a), "Emergence of Representation in Autonomous Agents", in: Cybernetics and Systems: Special Issue on Epistemological Aspects of Embodied Artificial Intelligence, 28(6), 489–498.

Bickhard, M. H. (1997b), "Is Cognition an Autonomous Subsystem?" In" S. Ó Nualláin, P. McKevitt, E. MacAogáin. Two Sciences of Mind: Readings in Cognitive Science and Consciousness. (115–131). Amsterdam: John Benjamins.

Bickhard, M. H. (forthcoming). "Critical Principles: On the Negative Side of Rationality", in: Herfel, W., Hooker, C. A. (eds.) Beyond Ruling Reason: Non-formal Approaches to Rationality.

Bickhard, M. H. (in press-a). "Dynamic Representing and Representational Dynamics", in E. Dietrich, A. Markman (eds.) Cognitive Dynamics: Conceptual Change in Humans and Machines. Cambridge, Mass.: MIT.

Bickhard, M. H. (in press-b). "Levels of Representationality", in: Journal of Experimental and Theoretical Artificial Intelligence.

Bickhard, M. H. with Campbell, Donald T. (forthcoming). "Emergence", in: P. B. Andersen, N. O. Finnemann, C. Emmeche, & P. V. Christiansen (eds.) Emergence and Downward Causation.

Bickhard, M. H., Campbell, R. L. (1989), "Interactivism and Genetic Epistemology", in: Archives de Psychologie, 57(221), 99–121.

Bickhard, M. H., Richie, D. M. (1983), On the Nature of Representation: A Case Study of James J. 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: Elsevier Scientific.

Brooks, R. A. (1990), "Elephants don't Play Chess", in: P. Maes (ed.) Designing Autonomous Agents. Cambridge, MA: MIT Press, 3–15.

Brooks, R. A. (1991), "Intelligence without Representation", in: Artificial Intelligence, 47(1–3), 139–19.

Brooks, R. A. (1991a), " Challenges for Complete Creature Architectures", in J.-A. Meyer, S. W. Wilson (eds.) From Animals to Animats. Cambridge, MA: MIT Press, 434–443.

Brooks, R. A. (1991b), "How to Build Complete Creatures Rather than Isolated Cognitive Simulators', in: K. VanLehn (ed.) Architectures for Intelligence. Hillsdale, NJ: Erlbaum, 225–239.

Brooks, R. A. (1991c), "New Approaches to Robotics", in: Science, 253(5025), 1227–1232.

Campbell, R. L., Bickhard, M. H. (1986), Knowing Levels and Developmental Stages. Basel: Karger.

Cherian, S., Troxell, W. O. (1995a), "Intelligent behavior in machines emerging from a collection of interactive control structures", in: Computational Intelligence, 11(4). Blackwell Publishers. Cambridge, Mass. and Oxford, UK.,565–592.

Cherian, S., Troxell, W. O. (1995b) "Interactivism: A Functional Model of Representation for Behavior-Based Systems", in: Morán, F., Moreno, A., Merelo, J. J., Chacón, P. Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life, Granada, Spain, Berlin: Springer, 691–703.

Christensen, W. D. (1996), "A complex systems theory of teleology", in: Biology and Philosophy, 11, 301–320.

Christensen, W. D., Collier, J. D., Hooker, C. A. (in preparation). "Autonomy, Adaptiveness, Anticipation: Towards autonomy-theoretic foundations for life and intelligence in complex adaptive self-organising systems".

Clark, A. (1997), Being There. MIT/Bradford.

Cliff, D. (1991), "Computational Neuroethology: A Provisional Manifesto", in: J.-A. Meyer, S. W. Wilson (eds.) From Animals to Animats. Cambridge, MA: MIT Press, 29–39.

Cummins, R. (1975), Functional Analysis. Journal of Philosophy, 72, 741–764.

Dartnall, T. (1997), "What's Psychological and What's Not? The act/content confusion in cognitive science, artificial intelligence and linguistic theory", in: S. Ó Nualláin, P. McKevitt, E. MacAogáin. Two Sciences of Mind: Readings in Cognitive Science and Consciousness. (77–113). Amsterdam: John Benjamins.

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

Fodor, J. A. (1990a), A Theory of Content. Cambridge, MA: MIT Press.

Fodor, J. A. (1990b), "Information and Representation", in: P. P. Hanson (ed.) Information, Language, and Cognition. Vancouver: University of British Columbia Press, 175–190.

Fodor, J. A. (1998), Concepts: Where Cognitive Science went wrong. Oxford.

Godfrey-Smith, P. (1994), A Modern History Theory of Functions. Nous, 28(3), 344–362.

Hooker, C. A. (1996), "Toward a naturalised cognitive science", in: R. Kitchener and W O'Donohue (eds.) Psychology and Philosophy. London: Sage, 184–206.

Loewer, B., Rey, G. (1991), Meaning in Mind: Fodor and his critics. Oxford: Blackwell.

Maes, P. (1990a), Designing Autonomous Agents. Cambridge, MA: MIT Press.

Maes, P. (1990b), " Situated Agents Can Have Goals", in: P. Maes (ed.) Designing Autonomous Agents. Cambridge, MA: MIT Press, 49–70.

Maes, P. (1991), "A Bottom-Up Mechanism for Behavior Selection in an Artificial Creature", in: J.-A. Meyer, S. W. Wilson (eds.) From Animals to Animats. Cambridge, MA: MIT Press, 238–246.

Maes, P. (1992), "Learning Behavior Networks from Experience", in: F. J. Varela, P. Bourgine (eds.) Toward A Practice of Autonomous Systems. Cambridge, MA: MIT Press, 48–57.

Maes, P. (1993), "Behavior-Based Artificial Intelligence", in: J.-A. Meyer, H. L. Roitblat, S. W. Wilson (eds.) From Animals to Animats 2. Cambridge, MA: MIT Press, 2–10.

Maes, P. (1994), "Modeling Adaptive Autonomous Agents", in: Artificial Life, 1, 135–162.

Malcolm, C. A., Smithers, T., Hallam, J. (1989), "An Emerging Paradigm in Robot Architecture", in: T. Kanade, F.C.A. Groen, & L.O. Hertzberger (eds.) Proceedings of the Second Intelligent Autonomous Systems Conference. Amsterdam, 11–14 December 1989. Published by Stichting International Congress of Intelligent Autonomous Systems, 284–293.

Malcolm, C. Smithers, T. (1990), "Symbol Grounding via a Hybrid Architecture in an Autonomous Assembly System", in: P. Maes (ed.) Designing Autonomous Agents. Cambridge, MA: MIT Press, 123–144.

Millikan, R. G. (1984), Language, Thought, and Other Biological Categories. Cambridge, MA: MIT Press.

Millikan, R. G. (1993), White Queen Psychology and Other Essays for Alice. Cambridge, MA: MIT Press.

Nehmzow, U., Smithers, T. (1991), "Mapbuilding Using Self-Organizing Networks in "Really Useful Robots", in: J.-A. Meyer, S. W. Wilson (eds.) From Animals to Animats. Cambridge, MA: MIT Press, 152–159.

Nehmzow, U., Smithers, T. (1992), "Using Motor Actions for Location Recognition", in: F. J. Varela, P. Bourgine (eds.) Toward A Practice of Autonomous Systems. Cambridge, MA: MIT Press, 96–104.

Nicolis, G. (1995), Introduction to Nonlinear Science. Cambridge.

Nicolis, G., Prigogine, I. (1977), Self-Organization in Nonequilibrium Systems. New York: Wiley.

Nicolis, G., Prigogine, I. (1989), Exploring Complexity. New York: Freeman.

Pfeifer, R., Verschure, P. (1992a), "Beyond Rationalism: Symbols, Patterns and Behavior", Connection Science, 4(3/4), 313–325.

Pfeifer, R., Verschure, P. (1992b), "Distributed Adaptive Control: A Paradigm for Designing Autonomous Agents", in: F. J. Varela, P. Bourgine (eds.) Toward A Practice of Autonomous Systems. Cambridge, MA: MIT Press, 21–30.

Piaget, J. (1954), The Construction of Reality in the Child. New York: Basic.

Port, R., van Gelder, T. J. (1995), Mind as Motion: Dynamics, Behavior, and Cognition. Cambridge, MA: MIT Press.

Prem, E. (1995), "Grounding and the Entailment Structure in Robots and Artificial Life', in: Morán, F., Moreno, A., Merelo, J. J., Chacón, P. Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life, Granada, Spain. Berlin: Springer, 39–51.

Prigogine, I. (1980), From Being to Becoming. San Francisco: Freeman.

Rosenthal, S. B. (1983), "Meaning as Habit: Some Systematic Implications of Peirce's Pragmatism", in: E. Freeman (ed.) The Relevance of Charles Peirce. La Salle, IL: Monist, 312–327.

Rosenthal, S. B. (1990), Speculative Pragmatism. La Salle, IL: Open Court.

Shanon, B. (1993), The Representational and the Presentational. Hertfordshire, England: Harvester Wheatsheaf.