

Arches and Stones in Cognitive Architecture

A Response to Commentary

Randall D. Beer

Dept. of Electrical Engineering and Computer Science

Dept. of Biology

Case Western Reserve University

Cleveland, OH 44106

To appear in *Adaptive Behavior*

Marco Polo describes a bridge, stone by stone.

“But which is the stone that supports the bridge?” Kublai Khan asks.

“The bridge is not supported by one stone or another,” Marco answers, “but by the line of the arch that they form.”

Kublai Khan remains silent, reflecting. Then he adds: “Why do you speak to me of the stones? It is only the arch that matters to me.”

Polo answers: “Without stones, there is no arch.”

(Calvino, 1974, p. 82)

1. Introduction

The original goals of the work described in the target article were threefold. First, I wanted to encourage a direct confrontation of situated, embodied and dynamical ideas with cognitive phenomena, in the hopes of furthering debate on the role of these ideas in cognitive science. Second, I wanted to describe a specific research methodology that could concretely ground such a debate, and to illustrate in some technical detail how a situated, embodied, minimally-cognitive agent could be developed and dynamically analyzed. Third, I wanted to use this model agent as a springboard to begin to explore some of the larger implications of these ideas for explanation in cognitive science. Judging from the range of the commentary, it appears that, at the very least, the paper was successful at provoking a broader discussion of these issues. I would like to thank

all of my colleagues for taking the time to provide detailed comments on my paper. I have certainly found the process constructive, and I hope that others do as well.

2. Building Bridges

Too often, cognitive scientists and advocates of situated, embodied and dynamical (SED) approaches simply talk past one another. They seem to be interested in completely different things, and to use entirely different tools to study entirely different problems. The notion of “minimally cognitive behavior” was an attempt to identify some common ground on which these two worldviews could productively engage (Beer, 1996). The word “behavior” was intended to remind cognitive scientists that any intellectual capacities are expressed by embodied agents acting in environments; no disembodied ratiocination allowed. The word “cognitive” was intended to remind SED advocates that the behavior must be sufficiently complex to be cognitively interesting; no wall-followers need apply. The word “minimal” was intended to remind both camps that progress requires tractable model systems. As the commentators themselves aptly demonstrate, there is no universally-accepted definition of cognition. Thus, my approach has been pragmatic: Minimally cognitive behavior is the simplest behavior that raises issues of genuine interest to cognitive scientists. My own intuition has been that the more “offline” a behavior appears to be (the less driven it is by immediate circumstances), the more cognitive it will be considered to be. However, the general point is simply that a fruitful collaboration can only begin when SED approaches produce agents that cognitive scientists actually care about understanding.

Some commentators are clearly not yet convinced. To **Edelman**, for example, discriminating circles from diamonds is a “toy task” that is “hardly worth the effort” of analysis. Even on this toy task, he finds the temporal evolution of the agent’s decision to be indicative of a “deep-seated incompetence”. Instead, Edelman emphasizes a “capacity for hierarchical abstraction” as essential to cognition. There are apparently two senses in which he means this. While I agree with Edelman that, as in any science, “complex cognitive systems cannot be

understood without resort to hierarchical abstraction of details”, I do not agree that cognitive systems themselves are “incapable of dealing with complex reality” without “appropriately structured mediating states” that support hierarchical abstraction. As I will argue later, it seems to me that this is a hypothesis that needs to be demonstrated, not an established empirical fact. In the spirit of minimally cognitive behavior, I propose the following challenge. Can Edelman suggest a specific task that would begin to engage this issue? Is he willing to help identify the simplest possible form of this task that preserves the essential features that concern him? If so, then it may be possible to apply the methodology described in the paper. The results of such experiments could then speak for themselves.

Clark is concerned that what the model agent described in the paper knows is too closely tied to what it does. Instead, he suggests that, in order to be cognitively interesting, there must be room for a mismatch between decision and action. Distinguishing between an agent’s intention and its behavior necessarily involves language. While the model agent’s decision-making was operationalized as commitment to action in the paper, operationalizing judgment requires verbal reports. As Clark acknowledges, this move constitutes something of a retreat from his earlier position that tasks with a significant “offline” component were sufficiently “representation-hungry” to engage cognitive issues (Clark, 1997). It is also somewhat problematic, since many phenomena studied in cognitive science do not directly involve language. Requiring verbal reports of intention thus represents a significant narrowing of the cognitive domain. There is also a great deal of work on the cognitive capacities of nonverbal animals, including chimpanzees, dolphins and octopi. Nevertheless, language certainly does constitute an important frontier for SED approaches, and work in this area is ongoing (Elman, 1995; Kirby, 2002).

Interestingly, **Di Paolo & Harvey** suggest a straightforward extension to the object discrimination task that might begin to address both **Edelman’s** and **Clark’s** concerns about the decision-making limitations of the model agent without requiring full-blown language. Following Di Paolo and Harvey, let us suppose that two light bulbs were added to the agent, one

communicating an “intention” to catch the current object and the other communicating an intention to avoid it. We could force the agent to make a discrete decision, as Edelman demands, by requiring that it irrevocably communicate a correct intention as early as possible. By separating intention and action in this way, we open a space in which to explore the relation between the two. Indeed, we could even probe situations in which there are mismatches between intention and action, as Clark requests. I encourage those who find categorical perception cognitively inadequate to look beyond this specific task to the general approach it was meant to illustrate. To slightly paraphrase Warren McCulloch (1965): “Don’t bite my finger, look where it’s pointing!”

A very different sort of concern about the cognitive adequacy of the model agent is raised by **Keijzer**, who argues that it is a poor example of what he calls “radically embodied cognition” because it “can do only one thing”. On his view, minimal agency, and therefore minimally cognitive behavior, requires an “unspecified, but significantly large number of perception-action capabilities”. Even putting aside the fact that the model agent can actually do *two* things (catch circles and avoid diamonds), this would seem to be a relatively trivial criticism to address. I see no reason in principle why exactly the same approach couldn’t be applied to tasks involving several different kinds of interactions with more objects of a larger variety of types. Indeed, ongoing work on selective attention has already begun to address the issue of handling multiple objects (Slocum et al., 2000).

However, it appears that Keijzer has something more fundamental in mind than simply increasing the available behavioral options. After making the curious claim that the model agent does not exhibit “organism-environment coupling through perception and action”, his real target becomes clear: Keijzer wants an account of what it means to be an organism, and is disappointed that the target article does not provide one. An organism’s behavior is ultimately geared toward its own continued existence and reproductive success, and I certainly agree that understanding the nature of the processes that constitute an organism is an important goal with important implications for cognition (Beer, in press). But I think it is a serious mistake to

postpone the application of SED ideas to cognitive phenomena until a rigorous account of biological agency is available. In the meantime, I see no reason why we cannot take the existence of agents as self-maintaining units of interaction for granted, at least provisionally, and use the available mathematical tools of dynamical systems theory to study the rich, and sometimes cognitive, behavior that agents produce in the service of that existence.

3. One Stone At A Time

Assuming that understanding minimally cognitive behavior is a reasonable goal, there are still many ways in which this goal might be pursued. The target article described a particular research methodology, “frictionless brains”, involving the construction and analysis of idealized models of complete brain-body-environment systems that exhibit minimally cognitive behavior. The target article also described a specific instantiation of this general methodology, in which genetic algorithms are used to evolve dynamical “nervous systems” for minimally cognitive agents, and the resulting brain-body-environment systems are then analyzed using the mathematical tools of dynamical systems theory. A key feature of the methodology is that it is incremental. Tools and insight developed in the construction and analysis of simpler model agents are then applied to more complicated ones.

Several commentators raised important questions about the ability of the proposed methodology to scale to more realistic cognitive behavior. There are at least two different sorts of scaling issues that are relevant here. First, the specific technique of evolving model nervous systems may not extend to more complex agents due to limitations of the evolutionary algorithm, limitations of the neural model, or limitations of the available computational resources. While this is indeed a serious practical concern, I do not think that anything terribly fundamental hangs in the balance here. There is nothing sacred about the particular evolutionary algorithm or neural network model utilized in the target article. If they someday fail to deliver successful agents beyond some threshold of cognitive complexity, there are many other variations that can and are being explored (e.g., see Nepomnyashchikh & Podgornyj in this issue). After all, we know that

biological nervous systems can generate cognitive behavior and that they do, in fact, evolve. The challenge is to capture in our models the essential biological features that make this possible.

A second scaling issue, and the one that most concerned several commentators, is the applicability of dynamical analysis to more complicated cognitive agents. **Koehne & Schank**, for example, express concerns about both the computational expense of simulating and analyzing larger models, and the experimental difficulty of simultaneous recording and/or manipulation of large numbers of neurons in intact, behaving animals. While these are certainly legitimate concerns, computational power and multiunit neural recording techniques are improving much faster than our ability to understand the results. It is this latter problem that I find most pressing.

Nepomnyashchikh & Podgornyj suggest that parametrically-coupled maps might be more tractable to analysis than CTRNNs, particularly for the higher-dimensional behavioral switches that they emphasize. I fully support the exploration of different dynamical models. However, I do not see how this really makes the problem any more tractable. Either parametrically-coupled maps are dynamically universal, in which case their analysis will in general be as difficult as CTRNNs, or they are not, in which case they may miss important dynamical features. Furthermore, from a scientific (as opposed to an engineering) point of view, we are not free to choose models based solely on the convenience of analysis. Rather, we are faced with understanding the particular brain/body/environment systems that Nature has presented to us.

Finally, **Edelman**, complains that the tools of dynamical systems theory are barely up to analyzing even the “toy” task of circle/diamond discrimination. Strictly speaking, a theory’s correctness is logically independent of both its intuitive appeal and its technical difficulty. Over a century after quanta were first introduced into physics, quantum mechanics remains deeply puzzling. Yet it has passed every experimental test that has ever been performed. Similarly, the equations of quantum chromodynamics (the quantum theory of the strong nuclear force) are extraordinarily difficult to solve, yet all the predictions that have so far been extracted from QCD have been experimentally verified. It could very well be that, although a situated, embodied, dynamical perspective is correct, both the explanatory expectations we bring to cognition and the

mathematical, computational and experimental tools necessary to do it justice require considerable further development.

There is no question that better tools are needed, particularly in the areas of dimension reduction and nonautonomous dynamics. However, I am more optimistic about the prospects of dynamical systems theory (DST) as it presently exists than some of the commentators seem to be. Pessimism about DST is often grounded in a misunderstanding of dynamical analysis. Complete visualization of the entire state space of a dynamical system is rarely either necessary or desirable, and none of the local techniques of DST require visualization for their application. Limit sets can be found, and their local stabilities and bifurcations studied, in many thousands of dimensions. Even when more global analysis is required, there are often considerably fewer essential degrees of freedom within a given behavioral interaction than there are total degrees of freedom in a system. The fact that the model agent's discrimination behavior was largely restricted to a bundle of trajectories in the 16D state space enabled low-dimensional projections to be used in the target article. While 16 dimensions may not be considered large, it is already much larger than we can ever hope to directly visualize.

Interestingly, the hard part of the work described in the target article was not the analysis itself, which employed only the most elementary mathematical tools. Rather, the most difficult part was the conceptual work of finding the right questions to ask and of understanding the implications of their answers. When a sufficiently focused, well-defined mathematical question has been formulated, it can usually be answered, either analytically or numerically, by the tools we already have at hand. There may come a day when we understand the conceptual structure of cognition so well that only mathematical difficulties prevent the completion of cognitive science, but I do not think that that day has yet arrived.

Other commentators suggest how the methodology described in the target article could be usefully extended. For example, Li argues that developmental processes also need to be incorporated. There is no question that developmental processes play a central role in shaping adult behavior, and I fully agree that a situated, embodied and dynamical perspective on

ontogenetic change is likely to be quite productive. As Li certainly knows very well, there has been considerable effort along exactly these lines in the developmental literature over the past ten years, and I suspect that the methodology may have contributions to make to the ongoing reconciliation of dynamical and connectionist models of development (Thelen and Bates, 2003). However, as a practical matter, I would caution against trying to do too much at once. While the framework sketched in the target article and SED approaches to development have much to teach one another, perhaps both would benefit from, pardon the pun, further “development” before a complete synthesis is attempted.

Fajen & Turvey suggest that model agents be constructed that can perceive possibilities for action rather than object categories. But the paper’s emphasis on perceiving objects merely reflects the nature of the literature on categorical perception, rather than ignorance of the lessons of ecological psychology. I think that SED approaches have much to offer ecological psychology, and vice versa. Ecological psychology provides a framework for characterizing the behavioral opportunities that a given environment offers to an agent. In return, SED approaches provide a framework for understanding how an agent’s behavior is codetermined by the structure of its environment, the design of its body, and the dynamical potential of its nervous system. I would very much like to see closer cooperation between these two complementary perspectives. In fact, we have previously evolved agents that can perceive whether or not an opening affords passability (Slocum et al., 2000) based on a Gibsonian analysis of human walking through apertures (Warren and Whang, 1987), and it would be interesting to extend this work to the perception of affordances that change over time.

Di Paolo & Harvey propose that the methodology be extended to include more inter-trial variability during evolution. I certainly agree that inter-trial variability during evolution ensures generality of the resulting agents. While the target article randomized over only the initial horizontal offset of the object, one could also randomize over such things as initial vertical offset, object velocity, object size, and object orientation. The only concern here is that too much randomness during evolution can confound a genetic algorithm if the variability swamps the

actual structure of the fitness space that is being searched (although incremental shaping protocols can sometimes help with this problem). However, as Di Paolo and Harvey nicely demonstrate, at least object size variability is easily incorporated, and the other forms of variability should certainly be explored.

Di Paolo and Harvey also suggest that noise be added to the dynamics, in the form of stochastic perturbations to sensors, motors and neurons. Noise is important because it ensures the structural robustness of a model. My only caution here is that the analysis of stochastic dynamical systems is considerably more mathematically challenging than the analysis of deterministic ones (Lasota and Mackey, 1994). An N -dimensional continuous-state dynamical system becomes infinite-dimensional when noise is added because the state of the system is no longer described by a length- N vector, but by an N -dimensional *function* representing its probability distribution. In addition, such familiar notions as attractor, stability and bifurcation must be carefully redefined for stochastic dynamical systems. If I have so far focused only on deterministic dynamics, it is not because I think that noise is unimportant. Rather, my goal has been to better understand the simpler case first, before considering the complications introduced by noise.

Through the experimental results that they present in their commentary, Di Paolo and Harvey demonstrate that they have understood the essential point of the methodology perfectly: Formulating simplified models allows us to replace vague intuition with concrete investigation. Whether or not one agrees with the specific tasks I have chosen, or the particular simplifications and modeling decisions I have made, is almost irrelevant. The important thing is that we build and analyze a wide variety of concrete SED models of minimally cognitive behavior, starting with the simplest possible models and then incrementally complicating them as our understanding improves. Rather than fearing simplification, we should embrace it. The problem with the sort of microworlds studied in classical AI was not that they made idealizations, but that they made the *wrong* idealizations. Science cannot proceed without idealization. A classic textbook in general relativity advises students to

Study one idealization after another. Build a catalog of idealizations, of their properties, of techniques for analyzing them. This is the only way to come to grips with so complicated a subject as general relativity! (Misner et al., 1973, p. 943)

This advice seems equally sound for the vastly more complicated subject of cognitive science.

4. Arches and Stones

Much work obviously remains to be done in developing and extending the proposed methodology. However, it is already becoming clear that SED ideas have important implications for explanation in cognitive science. A final goal of the paper was to begin to explore some of these implications. Unsurprisingly, the most controversial issue concerned the role (or lack thereof) of internal representation in dynamical explanation.

While it may come as a surprise to some of the commentators, I do *not* consider myself to be an anti-representationalist. I fully understand the intuitive appeal of representation-talk, and I have no desire to ban *a priori* any explanatory framework from cognitive science. Rather, I consider myself to be a representational skeptic. The difference may be subtle, but it is important. Despite the enormous explanatory weight that the notion of internal representation is required to bear in cognitive science, there seems to be very little agreement about what internal representations actually are. We need look no further than the commentary for ample proof of this assertion. For example, while both **Edelman** and **Clark** chastise the model agent for having no representations, **Li** praises it for bringing “representational states back into dynamic systems”! Clearly, they must be talking about different things. In the face of such ambiguity, an informal attitude of “I’ll know it when I see it” benefits no one. Rather than conceiving of the representation debate as a battle, as does Edelman, I would think that both “sides” would welcome a critical examination of this fundamental concept.

The strategy proposed in the target article for a critical examination of internal representation was straightforward: Evolve agents on tasks that are rich enough to be representationally-

interesting, then examine whether or not these agents actually use representations in their operation. When the only selection pressure operating on an agent's evolution is the efficacy of its behavior on a given task, this is an interesting open question, one whose answer can only sharpen our understanding of representation regardless of how it comes out. **Bullock** worries that a dynamical analysis may be inherently biased against finding representations, but I think that this worry is misplaced. Even **Edelman** recognizes that “dynamical approaches as such are not inherently anti-representation”. It is relatively easy to imagine at least simple ways in which dynamical systems might represent, and more sophisticated schemes have also been proposed (e.g., Spencer and Schöner, 2003). More generally, dynamical systems can be as representational as computers, since every computational system is a dynamical system. If dynamical systems can be – but need not be – representational, then this approach really is representation-neutral.

Of course, it may very well be that discovering and characterizing representations in a dynamical system will be extremely difficult, as **Bullock** suspects. But, to some extent, that is exactly the point. If we do not assume the existence of representations *a priori*, then we have to work for them. The harder we have to work for them, the less straightforward a foundation for cognitive science they really provided in the first place. An inability to imagine any other way it might be can no longer be accepted as evidence for internal representation. The methodology outlined in the target article actually makes it possible to empirically test such assertions as “inner states or processes that seem to code for the presence or absence of [...] features” are necessary for cognition (**Clark**) or “without appropriately [hierarchically] structured mediating states, a cognitive system would be incapable of dealing with complex reality” (**Edelman**). But this methodology will only work if advocates of particular sorts of representation are willing to specify clear criteria for identifying them, regardless of the difficulty. Otherwise, we quite literally won't know what we're talking about when we talk about representation.

Scientific explanation works by spanning at least two levels of description. The fact that much of chemistry can be described in terms of the making and breaking of chemical bonds is not sufficient. It must also be possible, at least in principle, to understand how chemical bonds

themselves arise from quantum mechanics, even if this can only be carried out in practice for the very simplest cases. While we may certainly posit a representational level to explain the regularities we observe in cognitive behavior, representation itself must be explained in terms of sub-representational mechanisms. It cannot be representations all the way down, or we have explained nothing. While **Clark** suggests that perhaps cognition may be special in some way, I think it is wiser to be forced by the facts to this unhappy conclusion, rather than to assume it from the outset. At the very least, the notion of representation will be significantly clarified in the process.

Of course, inter-level explanations always raise the possibility that a higher-level story may have to be fundamentally reconceived. For every quantum chemistry success in the history of science, there is also a phlogiston failure. Why do I suspect that a fundamental reconceptualization of internal representation may be necessary? Because the dynamical analysis of evolved model agents suggests a broader set of possibilities. Whether it involves reflexively withdrawing a hand from a hot stove or linguistically negotiating a complex social landscape over the course of a lifetime, it is only overt behavior that is directly selected for or against during evolution. From a SED perspective, all behavior is soft-assembled from the agent's unfolding circumstances. Thus, the only requirement on an agent's nervous system is that it be endowed with a latent potential to engage in the patterns of behavior necessary to that agent's survival and reproduction. It is only when embedded within the agent's particular body and situated within the agent's particular environment that this potential is actually realized through the resulting dynamical interaction. The belief that an agent's nervous system must explicitly represent any regularities that the behavior of the entire brain/body/environment system exhibits is an additional theoretical hypothesis, subject to critical analysis and possible rejection. I make no claim to have falsified this hypothesis, only to have suggested how we might begin to test it in a concrete way.

Even if it should turn out that internal re-representation dissolves in the facts of situated, embodied, dynamical existence, I do not think that the utility of higher-level representational

explanation will disappear. Not only will it continue to be the best language we have for making everyday sense of our own and others' behavior, but, as **Pineda & Noble** and **Koehlne & Schank** emphasize, it will likely still have an important scientific role to play. Explaining chemical bonding in terms of quantum mechanics in no way reduced chemistry to physics. But it did clarify both the strengths and limitations of the bond concept as an approximate summary of the actual underlying physical processes. No modern chemist seriously believes that molecules are literally held together by tiny rods, although they sometimes find it convenient to talk as if they do. Likewise, cognitive science must learn to distinguish the symbolic terms we use to describe cognitive behavior from the mechanisms that actually underlie it. As human beings, we may be more interested in the intricate arches that our introspection reveals to be supporting the cathedral of cognition, but as scientists we have to understand how these arches arise from the interaction of stones.

References

- Beer, R.D. (in press). Autopoiesis and cognition in the Game of Life. To appear in *Artificial Life*.
- Beer, R.D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J.A. Meyer, J. Pollack and S. Wilson (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (pp. 421-429). MIT Press.
- Calvino, I. (1974). *Invisible Cities* (W. Weaver, translator). Harcourt Brace Jovanovich.
- Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. MIT Press.
- Elman, J.L. (1995). Language as a dynamical system. In R.F. Port and T. van Gelder (Eds.), *Mind as Motion* (pp. 195-225). MIT Press.
- Kirby, S. (2002). Natural language from artificial life. *Artificial Life* **8**(2):185-215.
- Lasota, A. and Mackey, M.C. (1994). *Chaos, Fractals and Noise: Stochastic Aspects of Dynamics* (2nd Edition). Springer-Verlag.
- McCulloch, W.S. (1965). *Embodiments of Mind*. MIT Press.
- Misner, C.W., Thorne, K.S. and Wheeler, J.A. (1973). *Gravitation*. W.H. Freeman.
- Slocum, A.C., Downey, D.C. and Beer, R.D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In J. Meyer, A. Berthoz, D. Floreano, H. Roitblat and S. Wilson (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior* (pp. 430-439). MIT Press.
- Spencer, J.P. and Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science* **6**(4):392-412.
- Thelen, E. and Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science* **6**(4):378-391.
- Warren, W.H. and Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance* **13**:371-383.