Autopoiesis and Cognition in the Game of Life

Randall D. Beer
Dept. of Electrical Engineering and Computer Science
Dept. of Biology
Case Western Reserve University
Cleveland, OH 44106

To appear in *Artificial Life*

Abstract

Maturana and Varela's notion of autopoiesis has the potential to transform the conceptual foundation of biology, as well as the cognitive, behavioral and brain sciences. In order to fully realize this potential, however, the concept of autopoiesis and its many consequences require significant further theoretical and empirical development. A crucial step in this direction is the formulation and analysis of models of autopoietic systems. This paper sketches the beginnings of such a project by examining a glider from the Game of Life in autopoietic terms. Such analyses can clarify some of the key ideas underlying autopoiesis and draw attention to some of the central open issues. This paper also examines the relationship between an autopoietic perspective on cognition and recent work on dynamical approaches to the behavior and cognition of situated, embodied agents.

Please address all correspondence to:

Randall D. Beer
Dept. of Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, OH 44106-7071

Phone: (216) 368-2816
Fax: (216) 368-2801
Email: beer@eecs.cwru.edu
URL: http://vorlon.cwru.edu/~beer

1. Introduction

When is a physical system alive? When is a physical system cognitive? These questions are so fundamental that practicing biologists, neuroscientists, and cognitive scientists rarely ask them, let alone try to answer them. However, there is every indication that such questions will become central to 21st century science. An increasingly pressing concern in postgenomic biology is how to reassemble the living organisms that molecular biology has spent the last 50 years taking apart. This necessarily involves identifying and characterizing living systems as coherent spatiotemporal structures that are generated and maintained through the interactions among their physical constituents. Likewise, neuroscience and cognitive science are beginning to show a newfound appreciation for how behavior and cognition arise in dynamic interaction between a brain, a body and an environment. Again, this necessarily involves identifying and characterizing the coherent behavioral structures that such interactions produce.

Maturana and Varela's work on autopoiesis and the biology of cognition (Maturana & Varela, 1980; Varela, 1979; Maturana & Varela, 1987; Varela et al., 1991) taught me how to think about these questions in a new way. I first encountered Maturana and Varela's ideas in 1985, while reading a preprint of Winograd and Flores' (1986) book *Understanding Computers and Cognition*, a scathing critique of classical artificial intelligence by one of its own. As a graduate student in AI, I had very little interest in the first question at the time, but, like many others in the mid 80s, was deeply concerned with classical AI's answer to the second. Maturana and Varela's work showed me how closely related the two questions really are, and gave me a vocabulary and a conceptual framework with which to express my dissatisfaction and to formulate a path forward. Indeed, there is a very real sense in which much of my subsequent work on the dynamics of adaptive behavior and cognition can be seen as an attempt to concretely express, illustrate and apply some of the insights that Maturana and Varela's ideas bring to understanding biological behavior and cognition (Beer, 1990; Beer, 1995; Beer, 2000).

I can offer no better tribute to Francisco Varela than to trace the threads of this intellectual debt. Along the way, I hope to concretize some of the key ideas of Maturana and Varela's framework using a simple model, so that these ideas might be made accessible to a wider audience. I also hope to suggest how the careful analysis of such models might advance the state of autopoietic theory. Finally, I will argue that recent work on the dynamics of adaptive behavior and cognition follows quite naturally from Maturana and Varela's biological perspective on cognition.

2. Preliminaries

The concept of *autopoiesis* provides the foundation for the rest of Maturana and Varela's framework (Maturana & Varela, 1973). Roughly speaking, an autopoietic (lit. self-producing) system is a network of component-producing processes with the property that the interactions between the components generate the very same network of processes that produced them, as well as constituting it as a distinct entity in the space in which it exists. The paradigmatic example of autopoiesis is a cell, in which the components are molecules, the interactions are chemical reactions, and the cell membrane serves as a physical boundary that spatially localizes these reactions into an entity (or "unity") distinguishable from its environment.

This is a simple yet surprisingly powerful idea. The physicist John Wheeler once said that time is defined in such a way as to make motion look simple. In contrast, it often seems as if life is defined in such a way as to make organisms look complex. To many biologists, life is either a long list of phenomenological properties (e.g., the ability to reproduce and evolve) or a long list of physical components (e.g., DNA). But a mule does not cease to live simply because sterility removes it from the stream of evolution, and a puree of biomolecular constituents is no more alive than a bowl of soup. Instead, Maturana and Varela offer a view of life as a specific organization of physical processes that has as its principal product the maintenance of its own organization: an organizational homeostasis. In their view, the phenomenology of life become

secondary consequences of its autopoietic organization, and the components of life become one particular physical instantiation of this organization.

Unfortunately, while the basic idea of autopoiesis seems clear enough, there is considerable subtlety and controversy in the details (Varela, et al., 1974; Varela, 1979; Maturana and Varela, 1987; Mingers, 1995; McMullin, 1999), and the debates are often carried out in somewhat obscure and idiosyncratic language. What exactly does it mean for the components to generate the network of processes that produced them? Must all of the necessary components be generated by the network itself? How essential is a physical boundary to autopoiesis? What constitutes an acceptable physical boundary (e.g., how permeable is this boundary allowed to be?). Can any systems other than a living cell be autopoietic?

3. The Lives of Gliders

Studying simple concrete models can be an excellent way to sharpen our thinking about difficult concepts. For autopoiesis, several cellular automata models have been developed (Varela et al., 1974, Zeleny, 1977, McMullin and Varela, 1997). Unfortunately, such models have not moved beyond the stage of demonstrating that computational models of autopoiesis are possible. To my knowledge, none of these models have been analyzed in any depth, nor have they been used to explore any of the key ideas and consequences of Maturana and Varela's framework. A major goal of this paper is to sketch the beginnings of such a project.

The Game of Life is probably familiar to almost everyone. It was introduced by John Conway in 1970 and popularized by Martin Gardner in the pages of *Scientific American* (Berlekamp et al., 1982; Gardner, 1983; Poundstone, 1984). Life is a two-dimensional binary cellular automata in which the next state of each lattice cell depends only on its own state and the sum of the states of its eight immediate neighbors (the Moore neighborhood of radius 1). The rules are simple: (1) A dead cell with exactly three live neighbors becomes a live cell (*birth*); (2) A live cell with two or three neighbors remains alive (*survival*); (3) Otherwise, a cell dies or remains dead (*overcrowding* or *loneliness*). With these three simple rules, Life is capable of

generating patterns of bewildering complexity. Indeed, it is known to be Turing universal (Berlekamp et al., 1982).

Consider a glider, the simplest oscillatory structure that moves in the Life universe (Figure 1). A glider is a configuration of five ON cells that undergoes a sequence of transformations which ultimately leave the original glider displaced by one cell both horizontally and vertically. These transformations repeat every four cycles, so that, over time, a glider moves diagonally across the Life universe. As usual, we assume that the Life universe is closed, with periodic boundary conditions. Gliders appear quite commonly from random initial configurations, and they play an important role in the Turing universality of the Game of Life.

[Insert Figure 1]

Is a glider a useful model of an autopoietic system? A glider certainly consists of spatially localized configurations of components (the pattern of ON cells) that participate in networks of processes (mediated by the Game of Life update rules acting through the overlapping Moore neighborhoods of these components) that regenerate the configurations of components necessary to maintain that network. In short, a glider is a coherent localized pattern of spatiotemporal activity in the Life universe that continuously reconstitutes itself. On the other hand, I suspect that many would hesitate to call the glider an autopoietic system. Are self-maintaining spatiotemporal patterns really analogous to physical self-construction? Do the states of individual cells in the lattice really deserve to be called components? Does turning a cell on or off really count as production of components? Does a glider really possess a boundary that generates and constrains it? These are exactly the kinds of questions that a careful analysis of idealized models would eventually hope to sharpen. For now, let us tentatively agree that gliders model at least some features of autopoietic systems, and are therefore worthy of further study in those terms.

In order to begin, we must state clearly what a glider is. This already engages deep issues concerning maintenance of identity in distributed dynamic processes. Normally, the label "glider" refers to the distinctive pattern of five ON cells shown in Figure 1. However, this

characterization seems incomplete from an autopoietic perspective. In order for these configurations of ON cells to undergo the sequence of transformations and renewal that we associate with a glider, they must be surrounded by a layer of OFF cells. This layer of OFF cells forms the environment necessary for the ON cells to undergo the required transitions, and the ON cells themselves participate in the production and maintenance of the OFF layer. In other words, the layer of OFF cells serves as a boundary necessary for the continued existence and propagation of a glider. Thus, it seems more correct to associate the term "glider" with not only the five ON cells, but also the union of all OFF cells in their Moore neighborhoods. These regions are highlighted in gray in Figure 1. The extent to which this boundary is analogous to the cell membrane of a living cell is an interesting open question. On the one hand, this boundary does delineate the spatial extent of the glider and serves as its interface with the rest of the Life universe. On the other hand, this boundary does not really compartmentalize the glider processes in the way that a cell membrane does, since ON cells in Life do not diffuse away like molecular components would.

In principle, a glider in one of four possible states could be centered about any cell in the Life universe and moving in any of four different diagonal directions. Unless the Life universe is otherwise empty, a glider's particular phase, position and orientation at a particular time matter to its future evolution. However, the very fact that we refer to all of these different configurations as "gliders" suggests that they share some essential identity.

Maturana and Varela use the terms "structure" and "organization" to distinguish these two different levels of abstraction. The abstract relations that define some unity as a member of a particular class constitute a system's organization. When we as scientific observers distinguish a unity for study from the universe at large, it is by virtue of its organization that we do so. The specific material properties of the components and processes that realize that organization constitute its structure. When we as scientific observers wish to measure or manipulate a unity, it is through its structure that we act. A key feature of this distinction is that the same organization can be instantiated in different structures. Thus, a glider might be materially

instantiated using pennies on a chess board rather than electrical signals in a computer. Or, more to the immediate point, two gliders in different phases and positions and moving in different directions can still exhibit the same organization. This separation of structure and organization will become crucial in the next section.

If a glider's structure consists of the locations and states of its constituent lattice cells, how can we describe a glider's organization? A first thought is that the glider organization consists of any collection of cells that undergoes the sequence of four transformations shown in Figure 1, irrespective of position or orientation. Then the glider organization is characterized by a particular spatiotemporal limit cycle (Figure 2A). This would certainly place all possible isolated gliders in the Life universe into a single class. However, if we take orientation-independence seriously, then we have a redundancy in our characterization. States 0 and 2 (Figure 1) are equivalent under a 1/4 rotation and a reflection, as are states 1 and 3. By identifying each of these pairs of states, the organization reduces to a limit cycle consisting of two abstract states (Figure 2B). In this figure, the state labels have been rotated to emphasize their similarities. In the glider's intrinsic "coordinate system", these two abstract states differ only in the contents of the center cell and the cell below it.

[Insert Figure 2]

Finally, what about glider precursor configurations, such as the one shown at the right in Figure 2C? This precursor evolves into a glider after one update. Patterns such as these, in all possible positions and orientations, as well as their precursors in turn, form the basin of attraction of the glider limit cycle (Figure 2C). Should these also be included as part of the glider organization? My inclination is to say no, because they do not yet exhibit the characteristic glider limit cycle. They are more akin to the biochemical precursors to life than they are to life itself. Note, however, that this decision has a nontrivial implication. Suppose that we perturb a glider into the precursor shown in Figure 2C and it returns to the glider limit cycle one step later. If we consider all precursors to be part of the glider organization, then this perturbation is just that: a perturbation to a single persistent entity. If we do not, then the "perturbation" destroys

the first glider and its debris spontaneously organizes into a second one. With the proper patterns of environmental activity, the appearance of the second glider can be delayed for an arbitrarily long period of time. Such destruction/recreation occurs quite frequently when gliders interact with one another or with other structures in the Life universe.

The fact that the continuity of glider identity depends crucially on how we choose to define glider organization demonstrates a potential problem. The original formulation of autopoiesis is quite absolute about whether or not a given system exhibits an autopoietic organization; there is no middle ground. But how can we judge this from an observation of the current state of the cell? Are molecules we observe passing through the cell membrane a part of its normal operation, or is the cell slowly leaking its contents into the extracellular matrix? Will a rupture we observe in the cell membrane ultimately prove to be fatal, or will it subsequently be repaired? In order to answer such questions, we need to know how both the dynamics of the cell and the dynamics of its environment will unfold over time. But simply shifting to a longer timescale of observation raises another problem: How long an observation is necessary? Strictly speaking, no system is autopoietic if it is observed over a long enough interval of time. The formal characterization of organization is an open problem (Fontana and Buss, 1994), some of whose difficulty is already apparent in a simple glider.

3. Glider/Environment Interaction

When we as scientific observers identify a glider, we also leave behind a glider-shaped "hole" in the Life universe (Figure 3A). By defining a unity, we also define its environment. The interaction between a unity and its environment takes the form of mutual structural perturbations. Since a glider can interact with its environment only through its boundary, these perturbations consist of cell state changes at the interface between the glider's boundary and the part of the environment that immediately surrounds that boundary (cross-hatched regions in Figure 3). The states of all other cells in the environment are not directly observable by the glider, although they can of course indirectly influence the observable cells at some future point in the interaction.

Thus, whatever the complexity and structure of the environment may be, it can only manifest itself to the glider as a pattern of activity of the cells in the cross-hatched region. The glider's internal state can likewise only influence its environment through the states of its OFF boundary cells. Indeed, because the cells on either side of this glider/environment interface have Moore neighborhoods that overlap both, their states are co-specified by both the glider and the environment.

[Insert Figure 3]

Maturana and Varela distinguish two broad classes of outcomes that can result from an interaction between a unity and its environment. In a *destructive* interaction, the unity is unable to compensate for the structural changes induced by an environmental perturbation, and it disintegrates. Figure 3B shows a glider encountering a 1x3 line of cells known as a blinker. Although the glider is still recognizable after one update, the OFF boundary of the glider has already "ruptured". As the interaction continues for another step, the glider disintegrates completely. After several more updates, the debris leaves behind a single 2x2 group of cells known as a block. In contrast, in a *nondestructive* interaction, the unity's organization is preserved, but its structure may be affected. In Figure 3C, a glider encounters a configuration of four ON cells in its environment. Although the glider organization is maintained, its phase is advanced by one time step and it is prevented from moving downward by one cell. A special case of a nondestructive interaction is one in which a unity's structure is unaffected. In this case, from the glider's perspective, it is as if no perturbation took place at all.

Despite the perturbations a unity receives, it is no mere puppet of its environment. Rather, a unity actively determines its own domain of interaction. At any point in time, a unity's structure specifies both the subset of environmental states that can influence it and the interactions with those states in which it can engage without disintegration. An environment can only select from among the interactions available; it cannot in general place a unity into an arbitrary desired state. Consider, for example, a glider in two different states encountering the same environmental perturbation (middle column in Figure 4A). In the first case, the glider exhibits a change of phase

and position similar to the one shown in Figure 3C. But in the second case, the same perturbation leads to a disintegration of the glider. Thus, identical environmental perturbations can affect a glider differently depending on the state the glider is in when they occur. The converse is also true. Different perturbations can leave a unity in the same state, in which case that unity is incapable of making any distinction between them.

[Insert Figure 4]

4. The Minds of Gliders

For Maturana and Varela, the ability of a unity to draw distinctions through its selective response to perturbations is the hallmark of the cognitive. Indeed, they refer to the domain of interactions in which a unity can engage without disintegration as its "cognitive domain". It is because this ability is grounded in a unity's autopoiesis that Maturana and Varela see cognition as an essentially biological phenomenon, since biological unities can engage only in interactions that affect their structure without destroying their biological organization.

By virtue of the structural changes that a perturbation induces in a unity, the effect of a second perturbation can be different than it would otherwise have been. For example, Figure 4A shows two initially identical gliders receiving a sequence of perturbations. Both gliders begin with the same structure, but the upper glider receives the perturbation shown in Figure 3C, while the lower glider does not. This leaves them in different states when they encounter the same second perturbation. Due to its modified state, the upper glider survives this second encounter with its state modified yet again (in fact, restored to its initial state). However, the second glider, not having been "prepared" by the first perturbation, is destroyed by the second one. Thus, each perturbation that a unity experiences, as well as the structural changes that it undergoes even in the absence of perturbations, influences its sensitivity and response to subsequent perturbations.

As long as no interaction leads to a loss of identity, this state-dependent differential sensitivity to perturbations can continue indefinitely, with each perturbation orienting the unity to different possibilities for future interaction. We can capture this structure in what I will call an

interaction graph (Figure 4B). Here, each black node represents a possible state of a unity. Arcs represent the environmental perturbations that are possible in each state given the unity's structure in that state. Black arcs indicate perturbations for which the unity can compensate, while gray arcs indicate perturbations that lead to disintegration. In such a diagram, a unity's cognitive domain is precisely the set of all black nodes and arcs. The way in which different sequences of perturbations encountered in different states shape the potential for subsequent interaction is captured by the structure of the interaction graph. Note that one arc from each node represents the null perturbation (i.e., the state change that the unity would undergo in the absence of any environmental perturbation).

By undergoing a succession of perturbations and corresponding structural changes, any unity that persists must necessarily exhibit a certain congruence or fit with its environment. From among all of the possible sequences of structural change that a unity might follow, its interaction with a particular environment selects a particular pathway through its interaction graph (and, to at least some extent given the vastly larger state space of the environment, vice versa). Maturana and Varela have used the term "structural coupling" to denote this process by which the structural changes of a unity become coordinated with those of its environment. The notion of structural coupling serves to remind us that a unity-centric perspective of environmental perturbation is not the only one we can take. Structurally, we can also view the unity and its environment as a single dynamical system whose state is merely unfolding according to the underlying physical laws of the universe.

An especially interesting and important special case of structural coupling occurs when multiple unities share the same environment. Not only do such unities interact with their environment, they also serve as mutual sources of perturbation to one another. Indeed, to any particular unity, other unities literally are a part of their environment. Ongoing interactions between multiple unities can lead to structural coupling between them, so that the pathways they take through their respective interaction graphs become coordinated. Such a community of interacting unities can form what Maturana and Varela call a "consensual domain", in which

interactions serve to orient the other agents to similar possibilities for future action. It is this mutual orientation within shared domains that forms the basis of linguistic interactions or "languaging" (Maturana, 1978).

Unfortunately, we have far surpassed the ability of a simple glider to concretely illustrate such ideas. There are 24 cells in a glider's immediate environment, and thus 2^{24} possible perturbations that the environment can generate (some – perhaps many – of these will be "Garden of Eden" configurations, which can only occur as initial conditions). However, in my informal explorations, I have found that most perturbations to gliders are destructive. Thus, their cognitive domains are extremely limited. Although this may reflect either the discrete nature of cellular automata or the reactivity of Conway's update rules, I suspect that it is primarily due to the simplicity of gliders. A rich domain of interactions depends on a certain amount of structural degeneracy, so that many different structural configurations can instantiate a given organization. Since gliders are so small, there is very little "distance" between their organization and their structure, and thus very little room for nondestructive structural changes. Thus, we will move to a higher level of abstraction in the next section.

5. The Dynamics of Adaptive Behavior and Cognition

While behavior and cognition are normally considered to be the province solely of brains, it is clear from the discussion above that Maturana and Varela do not view nervous systems as essential to cognition. By virtue of their state-dependent differential sensitivity to perturbation, *any* biological system (including a single cell or a plant) is capable of selectively interacting with its environment and therefore possesses at least a rudimentary cognitive domain. However, there is no question that nervous systems significantly enrich the cognitive domains of the animals that possess them. By increasing the internal state that can be maintained and thus the structural changes that can be tolerated, nervous systems expand enormously the range of interactions that an organism can engage in without loss of organization.

Nervous systems occur only in multicellular animals. Strictly speaking, multicellular animals are second-order unities, because they are instantiated by networks of first-order autopoietic systems (cells). Such second-order systems are organizationally homeostatic and maintain a physical boundary. They have a structure and an organization. They can be perturbed by their environments either destructively or nondestructively, and thus possess a cognitive domain. They also exhibit state-dependent differential sensitivity to perturbations, and can therefore engage in structural coupling with their environments and with one another. Thus, second-order systems exhibit all of the key features of an autopoietic system. However, there is considerable controversy as to whether such second-order systems are truly autopoietic (Varela, et al., 1974; Varela, 1979; Maturana and Varela, 1987; Mingers, 1995; McMullin, 1999). In an attempt to minimize confusion, I will henceforth use the more general term "agent" to refer to either a first-order autopoietic system or a second-order autopoietic-like system.

Within fields as diverse as robotics, neuroscience, developmental psychology and cognitive science, an embodied, situated and dynamical view of adaptive behavior and cognition has been emerging (Brooks, 1991; Varela et al., 1991; Thelen and Smith, 1994; Beer, 1995; Kelso, 1995; Port and van Gelder, 1995; Smithers, 1995; Chiel and Beer, 1997; Clark, 1997). Embodiment and situatedness emphasize the roles played by an agent's physical and contextual circumstances in the generation of its behavior. For a situated, embodied agent, taking appropriate action becomes the primary concern, and an agent's biomechanics, the structure of its environment, and its social context become sources of both significant constraints on and resources for action. Dynamical approaches emphasize the way in which an agent's behavior arises from the ongoing interaction between its brain, its body and its environment. On this view, the focus shifts from accurately representing an environment to continuously engaging that environment with a body so as to stabilize coordinated patterns of behavior that are adaptive for the agent. This perspective is summarized in Figure 5. It owes a great deal to the ideas of Ashby (1952) and Gibson's Ecological Psychology (Gibson, 1979), among others. But that, as they say, is another story.

[Insert Figure 5]

How does this dynamical perspective on adaptive behavior and cognition relate to Maturana and Varela's framework? Maturana and Varela clearly intend their framework to have significant consequences for thinking about what nervous systems do and how they work (Maturana, 1970; Varela et al., 1991), and it was these insights that originally attracted me to their ideas. However, there is a problem. When cognition is so closely tied to biological existence, as it is in Maturana and Varela's framework, a theory of behavior and cognition must be grounded in a theory of autopoiesis. Unfortunately, autopoietic theory is poorly developed at best. The core concepts of autopoiesis require considerably more concrete explication, and the mathematical tools available for working with such constructive or metadynamical systems are very limited (Bagley et al., 1989; Fontanna and Buss, 1994). Must work on behavioral and cognitive implications therefore be postponed until these shortcomings of autopoiesis are fully addressed? Of course not. While many of the key behavioral and cognitive implications of Maturana and Varela's framework follow from autopoiesis, they do not require a complete account of autopoiesis for their exploration and application. Instead, they can be naturally expressed in the language of dynamical systems theory, which has a considerably more mature mathematical structure.

A dynamical perspective on behavior and cognition follows directly from an autopoietic perspective on life when two key abstractions are made.

First, we focus on an agent's *behavioral dynamics.* An agent's behavior takes place within its cognitive domain, which is a highly-structured subset of its total domain of interaction. If we are concerned only with the normal behavior of an agent, then many of the structural details of its physical instantiation may not be directly relevant. A glider, for example, is instantiated in the states and arrangements of individual lattice cells in the Life universe. However, the behavior of a *glider*, as opposed to its constituent lattice cells, is better described in terms of changes in position, orientation, phase, etc. of the entire spatiotemporal pattern. Each nondestructive perturbation to a glider induce changes in these variables, as can the glider's own dynamics.

Thus, our first abstraction involves formulating a higher-level set of neural, somatic and environmental state variables more directly related to an agent's behavioral degrees of freedom, and then describing its behavioral dynamics in these terms (Figure 6A).

[Insert Figure 6]

Second, we abstract the set of destructive perturbations that an agent can undergo as a *viability constraint* on its behavioral dynamics. Since it is meaningful to study an agent's behavior only so long as that agent actually exists, we largely take an agent's autopoiesis for granted in behavioral and cognitive analyses. However, we must somehow represent in our behavioral description the limitations that autopoiesis impose on an agent's interactions or we will have no way of deciding if a given behavior is adaptive or not. Thus, our second abstraction involves collapsing all perturbations that lead to a loss of identity into a single terminal state that serves as a viability constraint on the agent's behavior (Figure 6A). If any actions are ever taken that carry the agent into this terminal state, no further behavior is possible.

To this point, all of our examples have been discrete in both state and time. However, it is more typical for behavioral-level descriptions to be continuous in nature due to the continuity of sensory signals, neural activity and body motions. In this case, the number of states and moments of time becomes infinite, and the cognitive domain becomes a manifold (Figure 6B). Behavior corresponds to a trajectory through this cognitive domain, and the viability constraint becomes a collection of forbidden regions that the behavioral trajectory cannot enter. Note that this constraint may also involve environmental variables and may vary in time. This is the picture that Figure 5 is intended to capture: A nervous system embodied in an animal, which is in turn situated in an environment with which it continuously interacts (Beer, 1995).

Despite the continuous setting, all of the insights summarized in an interaction graph (Figure 4B) still apply. Although the number of states is infinite in a continuous interaction graph (or interaction manifold) and perturbations become continuous signals, behavioral trajectories still exhibit state-dependent differential sensitivity to perturbations (Figure 7A). Each state still presents the environment with a restricted range of possible perturbations from which it can

select. Any particular perturbation sends the agent's behavioral trajectory down a particular path, which influences the agent's sensitivity and response to subsequent perturbations. For example, Figure 7B shows a continuous analogue of the discrete example in Figure 4A. Here, two copies of an agent start in the same behavioral state, with one receiving an initial perturbation beginning at 1 (black) and the other not (gray). With different perturbations, the same initial state evolves along different trajectories in the two agents, which leaves them in different states when a second perturbation begins at 2. By virtue of this difference in state, the same perturbation has different effects on the two trajectories, with the gray trajectory eventually violating the agent's viability constraint while the black trajectory avoids this fate. While an interaction manifold may be more difficult to visualize and understand than a discrete interaction graph, significant structure still exists.

[Insert Figure 7]

There are many consequences of taking an embodied, situated and dynamical perspective on adaptive behavior and cognition. Perception becomes a kind of perturbation to an agent's dynamics. An agent's internal state sets a context in which the effects of a given perturbation play out without necessarily playing any representational role. It also allows an agent to initiate actions independently of environmental perturbations, as well as organize its behavior in anticipation of future events. Behavior becomes a property of an entire coupled agent-environment system rather than being generated by any one component. The joint dynamics of multiple interacting agents leads to such structurally coupled phenomena as languaging, which involves the mutual orientation of agents in their respective cognitive domains to shared possibilities for future interaction. Learning can be understood as dynamics on multiple timescales. And so on. While this is not the place to elaborate on each of these claims, many examples of these phenomena and others are being explored through the detailed analysis of evolved model agents (Beer, 1997; Chiel et al., 1999; Di Paolo, 2000; Tuci et al., 2002; Beer, in press).

The main point I wish to emphasize here is simply that currently popular dynamical perspectives on adaptive behavior and cognition follow directly from Maturana and Varela's framework. Indeed, in my own case, it was exactly the chain of reasoning outlined above that led me to my present theoretical position (Beer, 1995). Although I believe that dynamical systems theory currently provides the best mathematical tools for understanding behavioral and cognitive dynamics, the concept of autopoiesis remains essential. Not only does it ground the existence and viability of the agents that dynamical approaches generally take for granted, but it continues to inspire the ongoing development of such approaches (Di Paolo, 2003).

6. Conclusion

Maturana and Varela's notion of autopoiesis, and the many important implications that follow from it, are revolutionary ideas in biology. They have the potential to make a real impact on the practice of biology, changing the biological phenomena we choose to study, the experimental questions that we ask about these phenomena, the ways in which we interpret the answers we receive, and indeed the very way we conceive of living systems. However, in order to fully realize this potential, the concept of autopoiesis requires significant further development and concretization. A crucial step in this direction is the formulation and analysis of theoretical models of autopoietic systems.

This paper has sketched the beginnings of such a project by examining a glider from the Game of Life in autopoietic terms. Aside from its pedagogical value, such an analysis clarifies some of the key ideas of autopoiesis (e.g., cognitive domains) and draws attention to some of the central open issues (e.g., formalizing organization). Considerable further analysis is possible. For example, it should be straightforward to fully characterize the cognitive domain of a glider, as well as the sorts of structural coupling, if any, that two gliders can undergo. However, if gliders turn out to be too simple, or they are missing some essential feature of autopoiesis, then there are larger and more complex propagating oscillatory structures in the Life universe that can be explored (Callahan, 2001), as well as interesting generalizations (Evans, 2003). Furthermore,

there are distributed automata models in which components have an independent identity as they move around (Thompson and Goel, 1988; Shibata and Kaneko, 2003), in which case the ability of a boundary to compartmentalize processes becomes important. The addition of stochastic perturbations and thermodynamic constraints could also be examined. The challenge before us is to develop a more rigorous formulation of autopoiesis in the simplest possible models that support it, whatever they may turn out to be.

Beyond its biological implications, the potential consequences of autopoiesis for the cognitive, behavioral and brain sciences are equally revolutionary. In this case, however, substantial progress is already being made on realizing this potential. I have argued that dynamical approaches to the behavior and cognition of situated and embodied agents follow directly from Maturana and Varela's framework when defensible abstractions of the behavioral dynamics of autopoietic systems are made. Furthermore, using the more mature mathematical tools of dynamical systems theory, these abstractions allow us to directly study the behavioral and cognitive implications of autopoiesis without first developing a complete theory of autopoiesis. If this analysis is correct, then it makes clear the central role played by state-dependent differential sensitivity to perturbation in the operation of dynamical agents. The behavioral and cognitive challenge before us is also clear. We must learn to characterize the structure of interaction manifolds, as well as the underlying neuronal dynamics that give rise to this structure and the kinds of agent-environment interactions that this structure makes possible. It is precisely this structure that dynamical analyses of brain-body-environment systems seek to understand (Beer, in press).

By the very definition of autopoiesis, to live is to constantly face the possibility of death. In the game of life, as in the Game of Life, it is inevitable that we will eventually encounter a perturbation for which we cannot compensate, and our own particular cognitive domain will disintegrate. But through the structural coupling in which we have engaged, our being persists in the perturbations we have induced in others, and continues to influence their behavior long after

we are gone. By that criterion alone, Francisco Varela lived a very rich life indeed. It is the highest accomplishment to which a life in science can aspire.

Acknowledgments. This paper is dedicated, with admiration and sadness, to the life and work of Francisco Varela. The idea of exploring some of the implications of autopoiesis through a study of gliders in the Game of Life first emerged from a very interesting discussion I had with Barry McMullin in late 1995 at the Santa Fe Institute. I thank Hillel Chiel for stimulating discussions during the preparation of this paper. I would also like to thank Ezequiel Di Paolo and the anonymous reviewers for their feedback. My research has been supported in part by grant EIA-0130773 from the NSF.

References

Ashby, W.R. (1952). *Design for a Brain*. Wiley.

Bagley, R.J., Farmer, J.D., Kauffman, S.A., Packard, N.H., Perelson, A.S., and Stadnyk, I.M. (1989). Modeling adaptive biological systems. *Biosystems* **23**:113-138.

Beer, R.D. (1990). *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology*. Academic Press.

Beer, R.D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence* **72**:173-215.

Beer, R.D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems* **20**:257-289.

Beer, R.D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences* **4**(3):91-99.

Beer, R.D. (in press). The dynamics of active categorical perception in an evolved model agent. To appear (with commentary and response) in *Adaptive Behavior*.

Berlekamp, E.R., Conway, J.H., and Guy, R.K. (1982). *Winning Ways for Your Mathematical Plays*, Vol. 2. Academic Press.

Brooks, R.A. (1991). New approaches to robotics. *Science* **253**:1227-1232.

Callahan, P. (2001). Life pattern catalog. Available online at http://radicaleye.com/lifepage/patterns/contents.html

Chiel, H.J. and Beer, R.D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences* **20**:553-557.

Chiel, H.J., Beer, R.D., and Gallagher, J.G. (1999). Evolution and analysis of model CPGs for walking I. Dynamical modules. *J. Computational Neuroscience* **7**(2):99-118.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again.* MIT Press.

Di Paolo, E.A. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In K. Murase and T. Asakura (Eds.) *Dynamical Systems Approach to Embodiment and Sociality* (pp. 19-42). Adelaide: Advanced Knowledge International.

Di Paolo, E.A. (2000). Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior* **8**(1):27-48.

Evans, K.M. (2003). Larger than Life: Threshold-range scaling of Life's coherent structures. *Physica D* **183**:45-67.

Fontanna, W. and Buss, L.W. (1994). The arrival of the fittest: Toward a theory of biological organization. *Bulletin of Mathematical Biology* **56**(1):1-64.

Gardner, M. (1983). *Wheels, Life and other Mathematical Amusements*. W.H. Freeman.

Gibson, J.J. (1979). The Ecological Approach to Visual Perception. Lawrence Erlbaum.

Kelso, J.A.S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.

Maturana, H.R. (1970). Biology of cognition. In (Maturana & Varela, 1980), pp. 1-58.

Maturana, H.R. (1978). Biology of language: The epistemology of reality. In G.A. Miller and E. Lenneberg (Eds.) *Psychology and Biology of Language and Thought: Essays in Honor of Eric Lenneberg* (pp. 27-63). Academic Press.

Maturana, H.R. and Varela, F.J. (1973). Autopoiesis: The organization of the living. In (Maturana & Varela, 1980), pp. 59-138.

Maturana, H.R. and Varela, F.J. (1980). *Autopoiesis and Cognition*. Boston, MA: Reidel.

Maturana, H.R. and Varela, F.J. (1987). *The Tree of Knowledge.* Boston, MA: Shambhala.

McMullin, B. (1999). Some remarks on autocatalysis and autopoiesis. Presented at the workshop Closure: Emergent Organizations and their Dynamics, May 3-5, 1999, University of Ghent, Belgium. Available online at http://www.eeng.dcu.ie/~alife/bmcm9901/

McMullin, B. and Varela, F.J. (1997). Rediscovering computational autopoiesis. In P. Husbands and I. Harvey (Eds.) *Fourth European Conference on Artificial Life* (pp. 38-47). MIT Press.

Mingers, J. (1995). *Self-Producing Systems: Implications and Applications of Autopoiesis*. Plenum.

Port, R.F. and van Gelder, T., Eds. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press.

Poundstone, W. (1984). *The Recursive Universe*. William Morrow.

Shibata, T. and Kaneko, K. (2003). Coupled map gas: Structure formation and dynamics of interacting motile elements with internal dynamics. *Physica D* **181**:197-214.

Smithers, T. (1995). Are autonomous agents information processing systems? In L. Steels and R.A. Brooks (Eds.) *The Artificial Life Route to Artificial Intelligence* (pp. 123-162). Lawrence Erlbaum.

Thelen, E. and Smith, L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.

Thompson, R.L. and Goel, N.S. (1988). Movable Finite Automata (MFA) models for biological systems I: Bacteriophage assembly and operation. *J. Theoretical Biology* **131**:351-385.

Tuci, E., Quinn, M. and Harvey, I. (2002). An evolutionary ecological approach to the study of learning behavior using a robot-based model. *Adaptive Behavior* **10**:201-222.

Varela, F.J. (1979). *Principles of Biological Autonomy.* New York: North Holland.

Varela, F.J., Maturana, H.R. and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization, and a model. *BioSystems* **5**:187-196.

Varela, F.J., Thompson, E. and Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.

Winograd, T. and Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. New Jersey: Ablex Publishing.

Zeleny, M. (1977). Self-organization of living systems: A formal model of autopoiesis. *International Journal of General Systems* **4**:13-28.

Figure Captions

Figure 1: A glider in the Game of Life. Active cells are represented by a black disk, while inactive cells are empty. As indicated by the arrow, this glider moves diagonally downward and to the right by one cell every four updates. The set of cells that the main text argues should be identified with a glider is indicated in gray. In order to illustrate how the rules of Life produce the $0 \rightarrow 1$ transformation, the number of active cells in the Moore neighborhood of each cell is given for the initial state 0.

Figure 2: Alternative definitions of glider organization. (A) Glider organization as an abstract 4-cycle. Regardless of its position or orientation, any glider goes through the endless cycle of four transformations shown in Figure 1. Here, each abstract state represents the equivalence class under translation and rotation. (B) Glider organization as an abstract 2-cycle. If we take orientation independence to include reflection, then states 0 and 2 in Figure 1 are in the same equivalence class, as are states 1 and 3. This leads to a glider organization with only two abstract states, labeled with canonical representations of the two distinct configurations. (C) Glider organization as the basin of attraction of the abstract 2-cycle. Only a schematic illustration of the basin is shown, with only one abstract precursor state given explicitly at right. The complete basin consists of the union of the two abstract glider states, their precursor states, the precursor states of those states, and so on.

Figure 3: Glider/environment interactions. (A) Identifying a glider also identifies its environment. All interactions between the two take place through the interface shown with cross-hatching. (B) A destructive perturbation that ruptures the glider's boundary and then disintegrates it. (C) A nondestructive perturbation that changes the phase and position of a glider without destroying it. As a result of this perturbation, state 1 (Figure 1) has been skipped and the downward portion of the glider's normal movement has been prevented.

Figure 4: An illustration of state-dependent differential sensitivity to perturbation. (A) The structural change induced by one perturbation can affect a unity's response to a subsequent perturbation. Two gliders begin in the same state, but only the upper one receives the first perturbation shown. Due to the resulting differences in state, the first glider can properly compensate for a subsequent perturbation while the second glider cannot and disintegrates. (B) An illustration of a unity's interaction graph. Each black node represents a possible state of the

unity and each arc represents a possible environmental perturbation and the state change that results. Black arcs represent perturbations that preserver the unity's organization, while gray arcs represent perturbations that lead to disintegration (gray nodes marked with an X). Only a few representative arcs are explicitly shown in this figure. In general, an interaction graph can be disconnected, recurrent, and contain both converging and diverging pathways.

Figure 5: For the purpose of behavioral and cognitive analyses, an agent can be conceptualized as three interacting dynamical systems that represent its nervous system, its body and its environment.

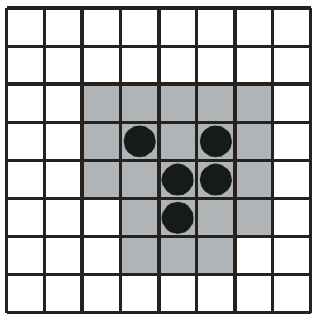
Figure 6: Abstracting the behavioral dynamics of an agent. (A) Abstracting an agent's behavioral dynamics into a discrete dynamical system. An agent's interaction graph has been rearranged to segregate its cognitive domain (black) from its viability constraint (gray). The cognitive domain is abstracted to a collection of behavioral states (open circles) and perturbation-triggered transitions between them, while the viability constraint is abstracted into a single terminal state representing disintegration. (B) Abstracting an agent's behavioral dynamics into a continuous dynamical system. Behavior corresponds to a continuous trajectory. The cognitive domain becomes a manifold (white region), while the viability constraint becomes a collection of regions that the behavioral trajectory cannot enter (gray regions).

Figure 7: State-dependent differential sensitivity to perturbation in continuous behavioral dynamics. (A) An agent's state and dynamical laws specify a limited range of behavioral trajectories that an environmental perturbation can invoke (gray triangle). When a particular perturbation begins, one of these possible paths is selected, causing the agent's behavioral trajectory to evolve along a path (solid curve) different from the one it might otherwise have taken (dashed curve). (B) Contingent chains of perturbations. Two trajectories begin in the same behavioral state, with the black trajectory receiving an initial perturbation at 1 while the gray trajectory does not, causing them to evolve along different paths. The resulting differences in state in turn cause the two trajectories to respond differently when they encounter an identical perturbation at 2, with the gray trajectory eventually violating the agent's viability constraint, while the black trajectory does not.

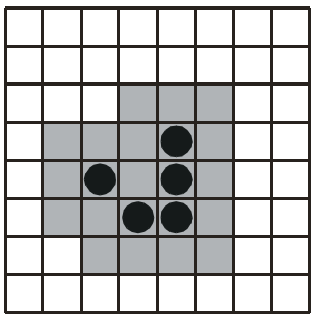
0

0	0	0	0	0	0	0	0
0	0	1	2	1	0	0	0
0	0	1	❶	2	1	0	0
0	1	3	5	❸	2	0	0
0	1	❶	❸	❷	2	0	0
0	1	2	3	2	1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

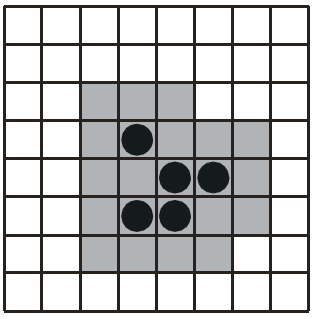
1



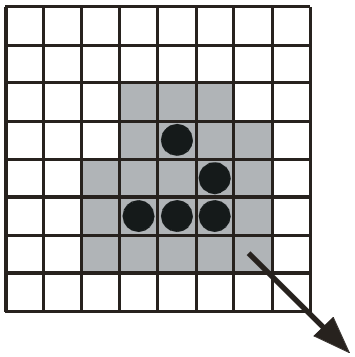
2

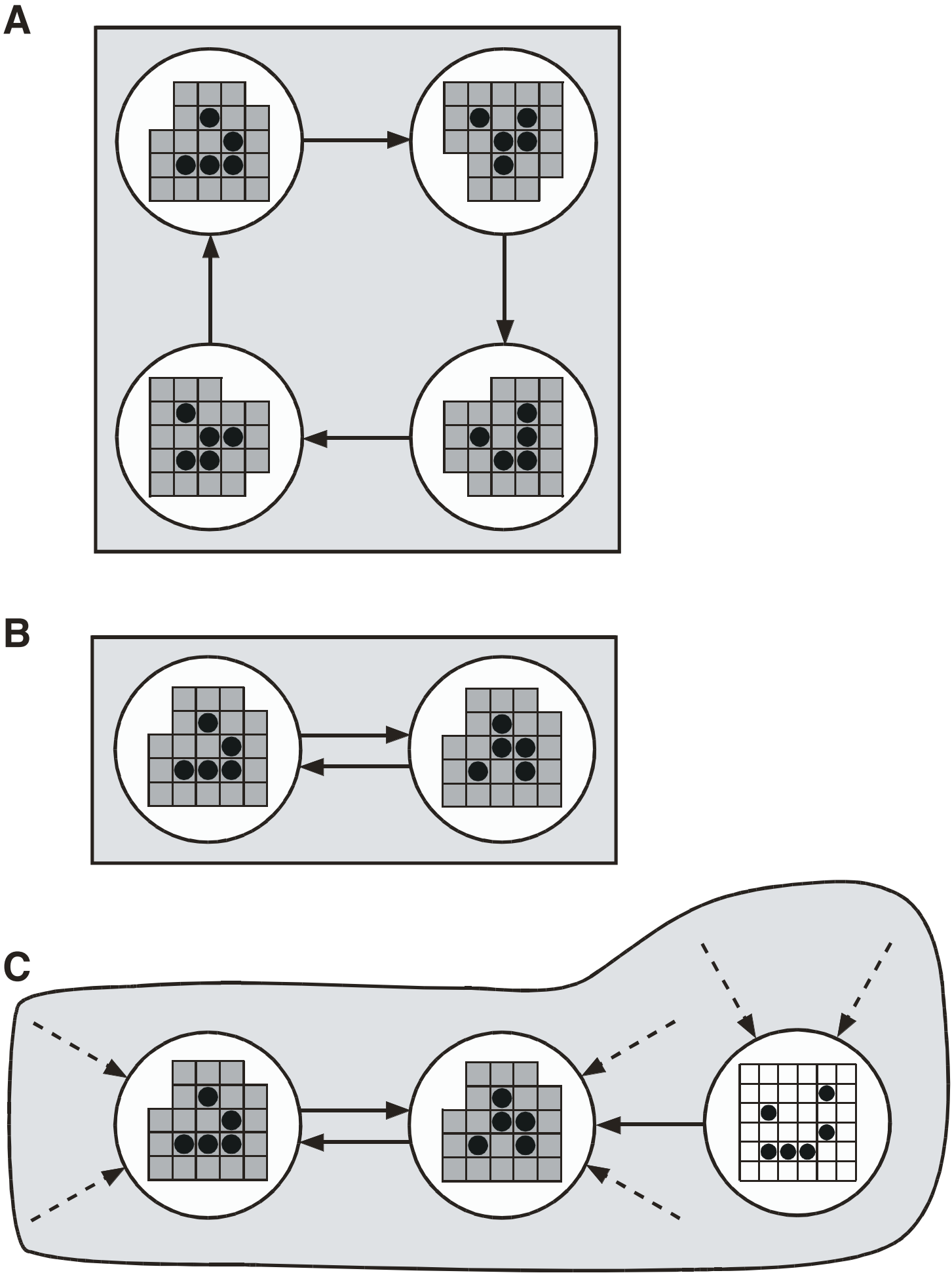


3

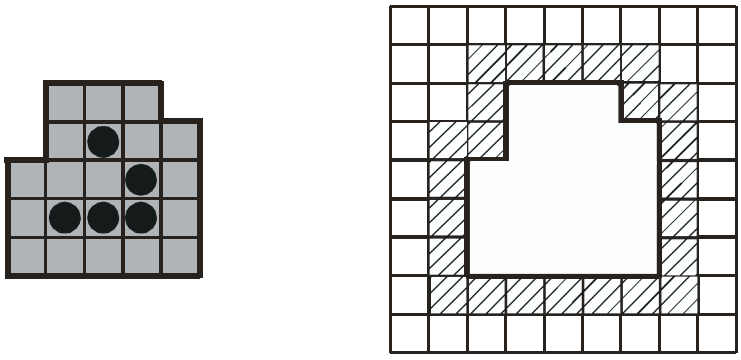


0

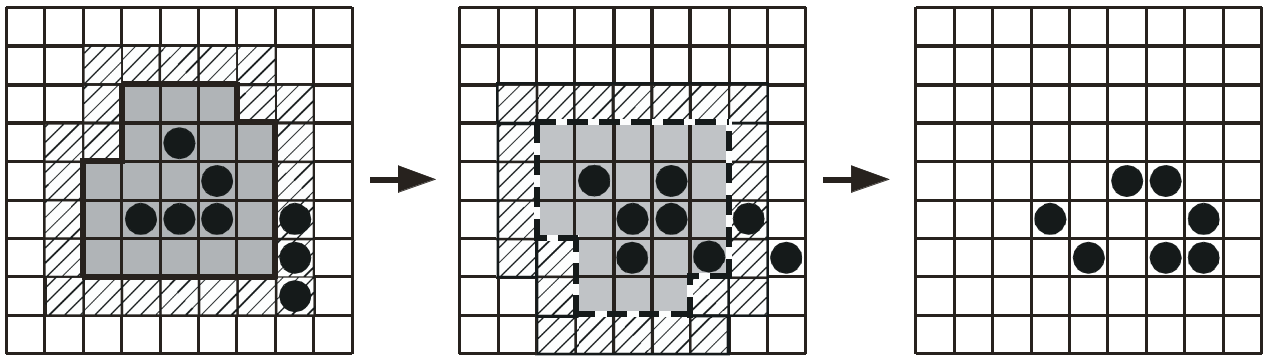



A
B
C

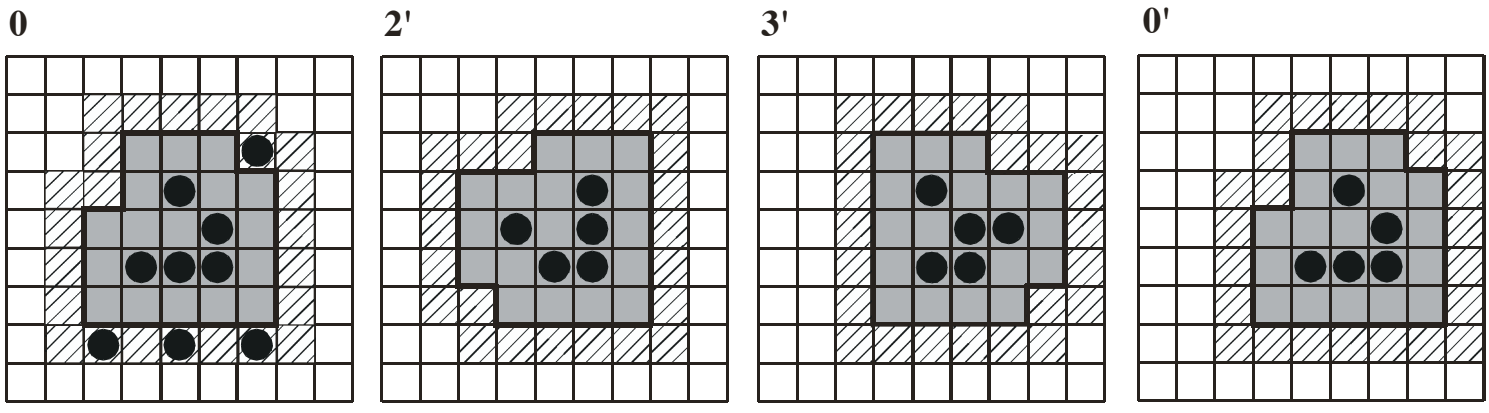
A

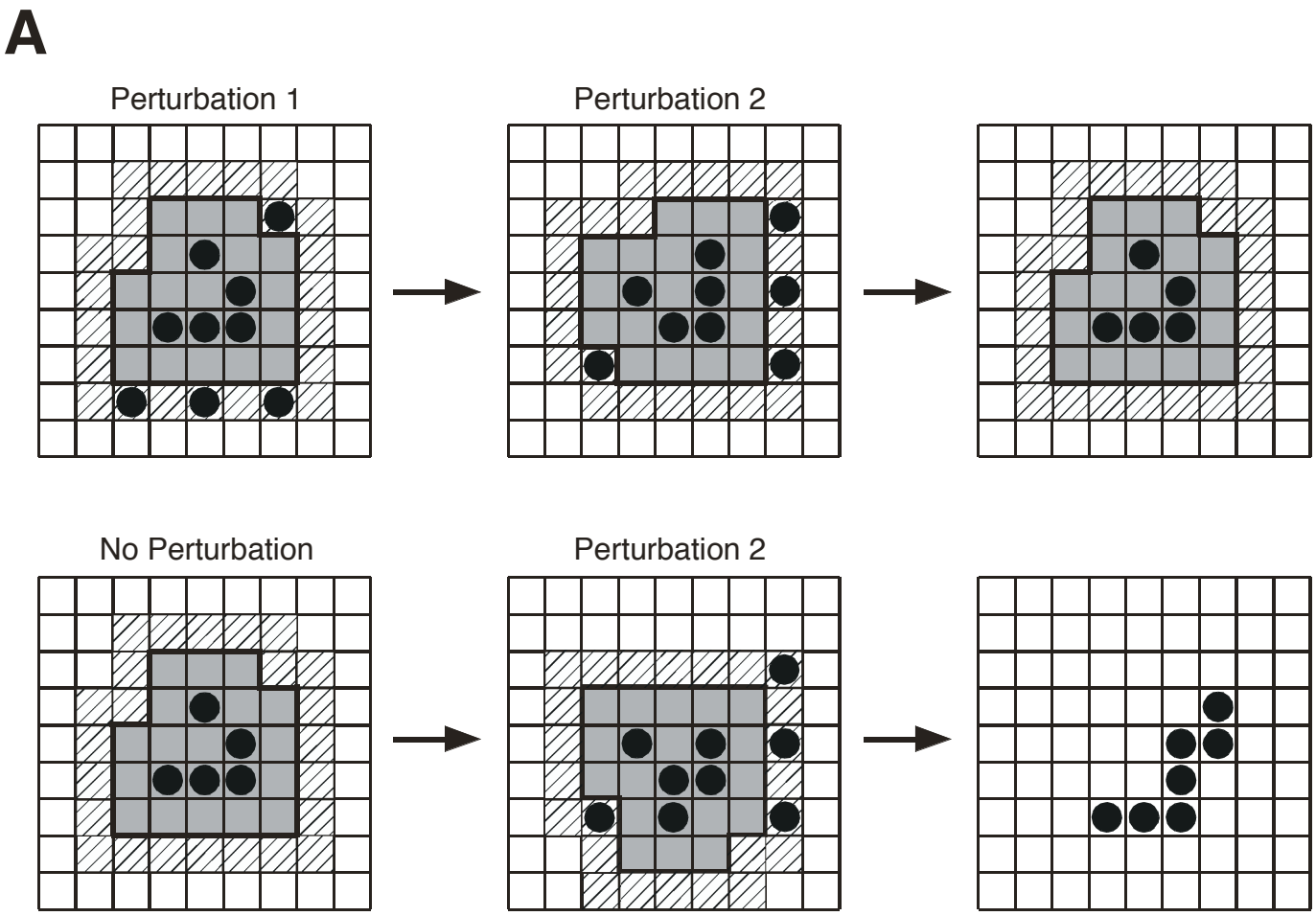


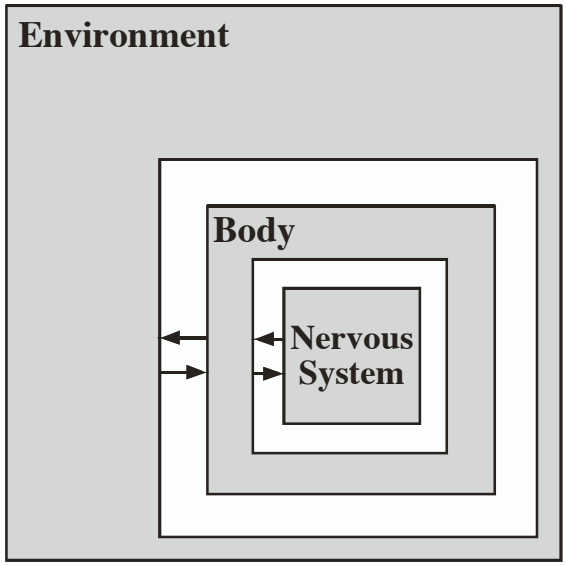
B

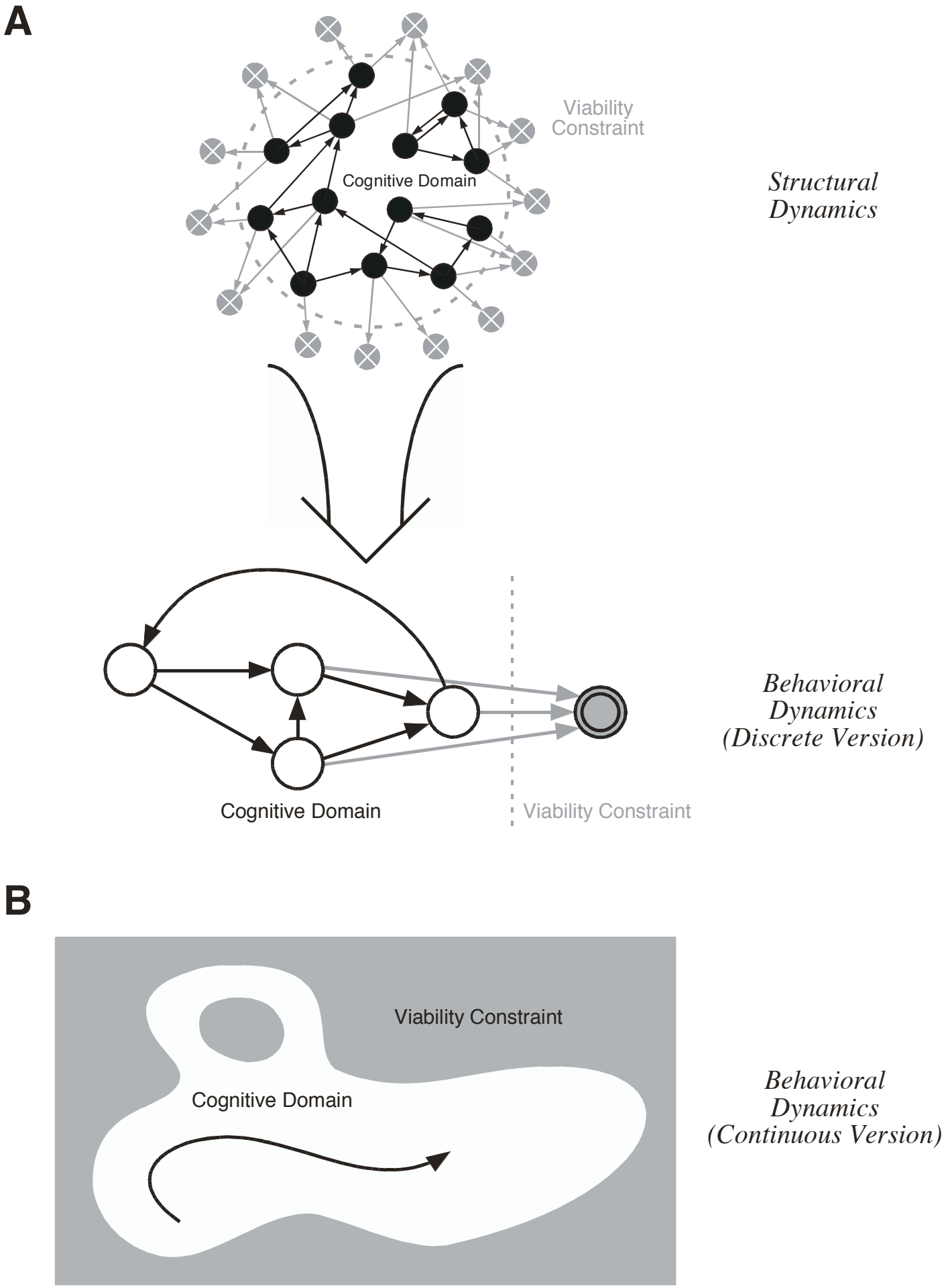


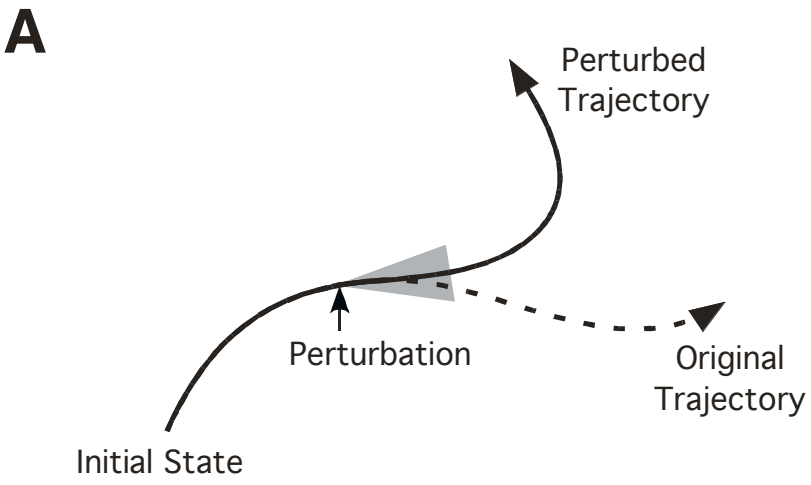
C

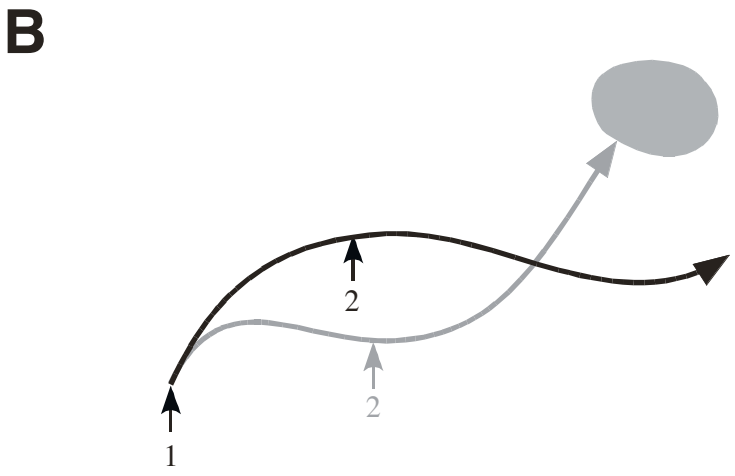



A
Perturbation 1
Perturbation 2
No Perturbation
Perturbation 2
B


Environment
Body
Nervous
System


A
Viability
Constraint
Cognitive Domain
Structural
Dynamics
Behavioral
Dynamics
(Discrete Version)
Cognitive Domain
Viability Constraint
B
Viability Constraint
Cognitive Domain
Behavioral
Dynamics
(Continuous Version)


A
Perturbed
Trajectory
Perturbation
Original
Trajectory
Initial State


B
2
2
1