Heuristics, Descriptions, and the Scope of Mechanistic Explanation

Carlos Zednik

czednik@uos.de
Institute of Cognitive Science, University of Osnabrück
49069 Osnabrück, Germany

To appear in: Pierre-Alain Braillard and Christophe Malaterre eds. *Explanation in Biology. An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Dordrecht: Springer.

Abstract:

The philosophical conception of mechanistic explanation is grounded on a limited number of canonical examples. These examples provide an overly narrow view of contemporary scientific practice, because they do not reflect the extent to which the heuristic strategies and descriptive practices that contribute to mechanistic explanation have evolved beyond the well-known methods of decomposition, localization, and pictorial representation. Recent examples from evolutionary robotics and network approaches to biology and neuroscience demonstrate the increasingly important role played by computer simulations and mathematical representations in the epistemic practices of mechanism discovery and mechanism description. These examples also indicate that the scope of mechanistic explanation must be re-examined: With new and increasingly powerful methods of discovery and description comes the possibility of describing mechanisms far more complex than traditionally assumed.

Keywords:

Mechanistic explanation, scientific discovery, evolutionary robotics, mathematical representation, dynamical systems theory, systems neuroscience, decomposability, heuristics of explanation.

1. Introduction

Many scientific explanations in biology and neuroscience are mechanistic explanations: they describe the mechanisms responsible for the phenomena being explained. Philosophers of science have sought to explicate mechanistic explanation by studying a handful of canonical examples. These include the mechanistic explanations of long-term potentiation (Machamer, Darden, & Craver, 2000), the action potential (Craver, 2006, 2007a), the citric acid cycle (Bechtel, 2006), and edge-detection in vision (Bechtel, 2008; Kaplan, 2011). But although these examples have been enormously useful for developing a philosophical conception of what mechanisms are and how they can be discovered and described, it is questionable whether they actually reflect the epistemic practices that contribute to the discovery and description of mechanisms in contemporary scientific research. Thus, the canonical examples have been taken to suggest that mechanisms are discovered via the dual heuristics of decomposition and localization (Bechtel & Richardson, 1993; Silberstein & Chemero, 2012), and that they are generally described in iconic diagrams and simple animations (Bechtel & Abrahamsen, 2005; Machamer et al., 2000). In contemporary biological practice, however, the practices of mechanism discovery and mechanism description are often grounded on computational and mathematical techniques that go beyond the well-understood principles of decomposition, localization, and diagrammatic representation.[1]

In order to provide a better reflection of contemporary research, this chapter introduces new examples of mechanistic explanation from evolutionary robotics and network approaches in biology and neuroscience. These new examples show how the canonical examples of mechanistic explanation fall short, and how mathematical and computational techniques effectively contribute to the discovery and description of mechanisms in hitherto unappreciated ways. Section 2 briefly reviews the core principles of mechanistic explanation as well as some of the canonical examples that illustrate these principles. Section 3 then shows how graph-theoretic measures and the evolution of simulated model organisms have been used as heuristic strategies to discover a possible mechanism of klinotaxis in *Caenorhabditis elegans* (Izquierdo & Beer, 2013; Izquierdo & Lockery, 2010). Subsequently, Section 4 shows how equations and analytic techniques from dynamical systems theory can be used to describe the organization, composition and activity of mechanisms (Beer, 2003). These new examples show how contemporary epistemic practices of mechanism discovery and mechanism description go beyond decomposition, localization, and pictorial representation.

[1] For related discussions of the role of mathematical modeling in biology and its relation to mechanistic explanations see the contributions to this volume by Baetue, Bechtel, Braillard, Mekios, Brigandt, and Issad & Malaterre.

Finally, this chapter concludes with a brief exploration of the consequences of going beyond the traditionally conception of mechanism discovery and mechanism description. Specifically, Section 5 considers the possibility that novel computational and mathematical techniques might increase the number and types of natural phenomena that can be explained in mechanistic terms. Indeed, the mechanisms discussed in Sections 3 and 4 are of a kind that is often thought to be too large and complex to be effectively decomposed and described in pictures. Thus, the phenomena these mechanisms exhibit are often thought to lie beyond the scope of mechanistic explanation. As novel computational and mathematical techniques make the epistemic practices of mechanism discovery and mechanism description ever more powerful and sophisticated, however, these difficulties might eventually be overcome, thereby extending the scope of mechanistic explanation.

2. Mechanistic explanation and its canonical examples

Mechanistic explanations describe the mechanisms responsible for phenomena of explanatory interest. Therefore, a philosophical conception of mechanistic explanation is answerable to metaphysical as well as epistemological concerns: What are mechanisms and how do they relate to the phenomena being explained? How are mechanisms discovered and subsequently described in scientific practice?

Although the philosophical literature boasts several statements of what mechanisms are, a particularly influential one is due to Carl Craver: “A mechanism is a set of entities and activities organized such that they exhibit the phenomenon to be explained” (Craver, 2007, p. 5. See also: Bechtel & Abrahamsen, 2005; Bechtel & Richardson, 1993; Glennan, 2002; Machamer et al., 2000). This definition captures three widespread ideas, each of which is exemplified by the molecular mechanism of the action potential in nerve cells (Craver, 2006, 2007a).

First, mechanisms consist of *entities* (or parts) on one hand and *activities* (or operations) on the other. Entities are structures or objects in the world with properties that change over time. Activities are what entities do: how entities change over time, and how such changes influence other entities. In the molecular mechanism for the action potential, component entities include sodium and potassium ions, among others, as well as dedicated ion channels on the cell membrane. The mechanism’s component activities include the opening and closing of channels, and the passing of ions through corresponding channels. Second, a mechanism’s component entities and activities are *organized*—they are related to one another in a particular way. Whereas a mechanism’s entities are often organized spatially, related to one another by physical distances and bearings, its activities are typically organized temporally, by occurring at a particular moment in, or for a particular length of, time. In the mechanism for the action potential, ion channels are situated on the cell membrane, and the opening of ion channels allows ions initially situated on one side of the membrane to pass to the other side.

Third and finally, Craver's statement of what mechanisms are specifies that the relationship between mechanisms and phenomena is one of *exhibiting*. Alternative locutions frequently used in this context include "producing" and "being responsible for". If a phenomenon can be understood as a particular pattern of changes to a particular set of properties over time (Bechtel & Abrahamsen, 2010), a mechanism can be said to exhibit this phenomenon if its properties—i.e. the properties of its component entities and activities, as well as the properties of their organization—change in accordance with this pattern. Thus for example, the action potential—a period of rapid depolarization of the cell body followed by gradual repolarization—begins when the opening of sodium channels allows Na^+ ions to permeate the cell membrane, driving the voltage of the cell body toward the sodium equilibrium potential near +30mV. In turn, repolarization occurs when potassium channels open to allow K^+ ions to leave the cell body, thus eventually allowing the voltage to return to its resting potential of approximately -70mV.

Although metaphysical questions concerning the nature of mechanisms themselves often dominate philosophical discourse, an adequate philosophical conception of mechanistic explanation must also address epistemological questions that concern the representation of mechanisms in the scientific literature. To this end, it is useful to distinguish two distinct epistemic practices. In the first, *mechanism discovery*, a mechanism's component entities, activities, and organization are identified by studying both the system in which the mechanism is realized and the phenomenon for which the mechanism is deemed responsible. In the second, *mechanism description*, the mechanism's composition and organization are represented in a way that shows that (and ideally, shows how) the mechanism exhibits the target phenomenon.

Bechtel & Richardson's (1993) *Discovering Complexity* remains the single most comprehensive discussion of mechanism discovery. Bechtel & Richardson adopt a framework pioneered by Herbert Simon (1996), in which scientific discovery in general is likened to a search process through a space of possible solutions to a problem. Thus, the specific case of mechanism discovery is conceived as a search process through a space of possible descriptions of the mechanism for the target phenomenon—also known as "how possibly" models of the mechanism (Craver, 2007a). This search process ends when a true description has been identified—a "how actually" model of the mechanism. Although fallible, heuristic strategies greatly facilitate the search process by allowing scientists to specify and constrain the space of possible mechanisms, as well as to identify regions of the space that merit further exploration, to the exclusion of others.

Bechtel & Richardson place particular emphasis on two such heuristic strategies: *decomposition* and *localization*. Decomposition itself comes in two varieties. Structural decomposition involves breaking down a complex entity into a collection of simpler entities. Typically, the complex entity subject to structural decomposition is the system from which the target phenomenon arises, and which is therefore presumed to realize the mechanism responsible for that phenomenon. In contrast, functional decomposition involves breaking down a complex activity into a collection of simpler activities. The complex activity to which this strategy is usually applied is the phenomenon itself, and the aim is to show how that phenomenon results from the simpler activities whenever these are executed simultaneously or in a particular order (see also: Cummins, 1983). Functional and structural decomposition are linked by the heuristic of localization. The aim of localization is to establish a mapping between entities and activities, which in ideal cases shows that the activities identified via functional decomposition are in fact performed by the entities identified via structural decomposition. Successful localization is frequently used as evidence that the component entities and activities of a particular mechanism have been successfully discovered.

The combined use of decomposition and localization in the service of mechanism discovery is clearly exemplified by the explanation of edge-detection in mammalian vision (see also: Bechtel, 2008; Kaplan, 2011). After the striate cortex had been identified as particularly relevant for visual processing in the early 20th century, David Hubel and Torsten Wiesel (1959, 1968) applied the heuristic of structural decomposition to characterize the different types of cells that composed it. Through single-cell recordings in cats and macaque monkeys, they distinguished three types of cells on the basis of their responses to visual stimuli: *simple* cells that respond to light at specific retinal locations; *complex* cells that respond to bars of light at a particular angle of orientation; and *hypercomplex* cells that respond to bars of light that span the full width of the receptive field. But although Hubel & Wiesel were thus able to identify some of the component entities of the mechanism for visual edge-detection, they described neither their corresponding activities, nor their spatiotemporal organization. Both of these shortcomings were remedied by David Marr and colleagues in the 1970s and 1980s. Specifically, Marr & Hildreth (1980) performed a functional decomposition of vision that showed that edge-detection could be achieved by a sequential process that blurs a visual image with a Gaussian filter, and then applies a Laplacian operator to detect those locations of the blurred image that have the highest changes in intensity—the so-called *zero-crossings*. Marr & Hildreth then showed that a pair of neighboring simple cells with complementary activation profiles could combine to detect a single zero-crossing (figure 1A), and that a larger arrangement of these cells could function as a single edge-detector (figure 1B). In this way, the heuristic of localization was used to link Hubel & Wiesel's structural decomposition of the striate cortex with Marr & Hildreth's functional decomposition of visual edge-detection to provide a description of (some of) the component entities and component activities of the mechanism for the target phenomenon.

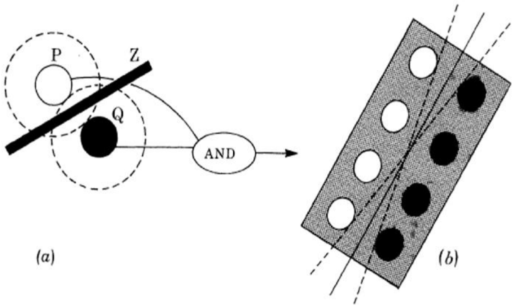

Figure 1. The organization of simple cells that underlies edge-detection in vision, reprinted from Marr & Hildreth (1980). **A:** A pair of cells with opposite response patterns, connected by a simple AND gate, can be used to detect zero-crossings in a blurred image. **B:** An array of several cell-pairs can be used to detect edges.

The epistemic practice of mechanism discovery can be distinguished from the practice of mechanism description. Whereas the former typically involves heuristic strategies to identify the entities, activities, and organization of the mechanism to be described, the latter involves one or more descriptive media to represent these entities, activities and their organization. Notably, the distinction between mechanism discovery and mechanism description is conceptual rather than practical: they need not correspond to distinct periods of time, or be conducted by distinct individuals. Indeed, description and discovery may be mutually constraining, as when representing a mechanism in a certain way reveals errors or ambiguities to be remedied by identifying additional entities, activities or modes of organization.

As is implicit in the aforementioned examples, the descriptive media used to represent mechanisms often include verbal characterizations and iconic or schematic diagrams. In addition, mechanisms are often described as physical and simulated two or three-dimensional models (Bechtel & Abrahamsen, 2005; Wright & Bechtel, 2007). Depending on the nature of the mechanism being described, certain descriptive media tend to be more effective than others. For example, mechanisms in which the spatial relationships between components are crucial—such as the mechanism for visual edge-detection, in which the spatial organization of simple cells is paramount—are more easily represented in diagrams than in words. In contrast, mechanisms in which temporal relationships are critical may be more effectively described by animations that show the time course of relevant events. Finally, mechanisms in which physical details take a back seat to functional relationships between component activities are frequently described with schematic representations such as box-and-arrow diagrams.

No matter the medium, mechanism descriptions provide mechanistic explanations when they adequately represent the mechanism responsible for the phenomenon being explained. When exactly a description is adequate in this way remains controversial. Nevertheless, a widespread idea is that it should refer to those and only those component entities and activities that are actually relevant to the phenomenon being explained. Craver (2007b) has elaborated on this idea by appealing to the notion of *mutual manipulability* (but for criticism see: Leuridan, 2011). On Craver's account, a component is to be included in the description of a mechanism "when one can wiggle the behavior of the whole by wiggling the behavior of the component and one can wiggle the behavior of the component by wiggling the behavior as a whole" (Craver, 2007b, p. 153). Thus for example, the description of the mechanism for edge-detection in vision should refer to simple cells because Hubel & Wiesel used single-cell recordings to show that these cells are activated during episodes of visual edge-detection, and because they showed that interfering with these cells (e.g. through lesions) affects the organism's ability to detect edges (Hubel & Wiesel, 1959, 1968).

The aim of this section has been to briefly review the metaphysical and epistemological principles of mechanistic explanation, as well as to introduce some of the canonical examples on the basis of which these principles are traditionally explored: the mechanisms for the action potential and for visual edge-detection. Although there are other such examples—the mechanisms of long-term potentiation (Machamer et al., 2000) and the citric acid cycle (Bechtel, 2006) are notable omissions—their nature and means of discovery and description differ insubstantially from the examples reviewed here. As the following sections will demonstrate, however, these canonical examples provide an overly narrow glimpse on mechanistic explanation in contemporary scientific practice. Although the core principles of mechanistic explanation remain unchanged—mechanisms can still be viewed as organized collections of entities and activities, and mechanistic explanations are descriptions derived, in part, by the systematic application of heuristic strategies—the mechanisms that figure in contemporary scientific research are far more complex, and the heuristic strategies and descriptive media invoked by practicing scientists are far more numerous and sophisticated, than these canonical examples would suggest.

3. New Heuristics for Mechanism Discovery

The role of heuristic strategies in mechanism discovery is one of the most underexplored aspects of mechanistic explanation. Rather than question or elaborate on Bechtel & Richardson's (1993) discussion, most philosophical treatments assume that the heuristics of decomposition and localization are part-and-parcel of mechanistic explanation. Inspired by Bechtel & Richardson's extended discussion of limits of mechanistic explanation when the heuristics of decomposition and localization fail, it is often assumed that abandoning these heuristics is tantamount to abandoning the search for mechanisms (e.g. Chemero & Silberstein, 2008). But this assumption is false: mechanism discovery is facilitated by any heuristic strategy that aids researchers to efficiently explore the space of "how possibly" models of a mechanism. This section introduces two such strategies: the *evolution of simulated model organisms* and *selective pruning*.

3.1 Evolving Simulated Model Organisms

In two separate but complementary studies, Eduardo Izquierdo and colleagues seek to discover the mechanism for klinotaxis in *Caenorhabditis elegans*. Klinotaxis is a form of goal-directed locomotion in which a chemical source is approached by repeatedly sweeping, and over the long run following, a chemical gradient whose concentration increases with proximity to the source (figure 2).

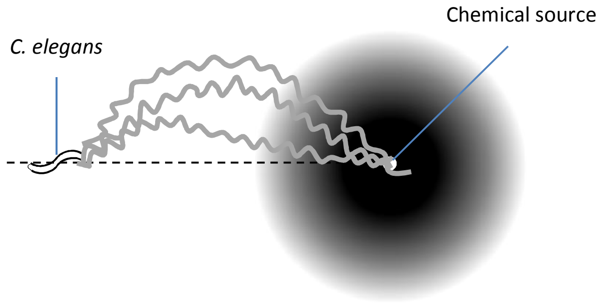

Figure 2. *C. elegans* klinotaxis, adapted from Izquierdo & Lockery (2010). The organism and its movement through a chemical gradient that increases in concentration with proximity to the source. The dotted line denotes the line of steepest ascent, from the organism's current position to the source; wiggly lines denote characteristic trajectories during individual klinotaxis episodes.

As is well-known, efforts to map the *C. elegans* nervous system have resulted in a detailed description of the organism's *connectome*: the 302 neurons and approximately 7000 synaptic connections and gap-junctions that make up its nervous system (Varshney, Chen, Paniagua, Hall, & Chklovskii, 2011; White, Southgate, Thomson, & Brenner, 1986). Although incredibly detailed, this descriptive knowledge falls short of explanation: it is as of yet unknown which individual parts or properties of the connectome contribute to particular behavioral capacities, and exactly how they do so. By invoking the descriptive knowledge of the *C. elegans* connectome to identify a possible mechanism for klinotaxis, Izquierdo and colleagues take a first step from description to explanation.

In the first study, Izquierdo & Lockery (2010) identify the possible component activities of the mechanism for klinotaxis. Previous ablation studies of the *C. elegans* connectome suggest the involvement of at least two kinds of chemosensory neuron (ASER and ASEL) that respond to increases and decreases of a chemical gradient in the environment, respectively, as well as two kinds of motor neuron (SMBD and SMBV) that control neck-muscle contractions on either side of the organism's body. What remains unknown is exactly how these chemosensory and motor neurons interact through mediating interneurons, and how motor action feeds back on chemosensation through the environment. Rather than address these questions through further ablation studies meant to isolate the contributions of specific interneurons, Izquierdo & Lockery adopt a simulation-based approach. Specifically, they invoke a simulated body model of *C. elegans* that is controlled by an artificial neural circuit (figure 3), and determine the range of circuit parameter values that enable the production of klinotaxis in a simulated environment.

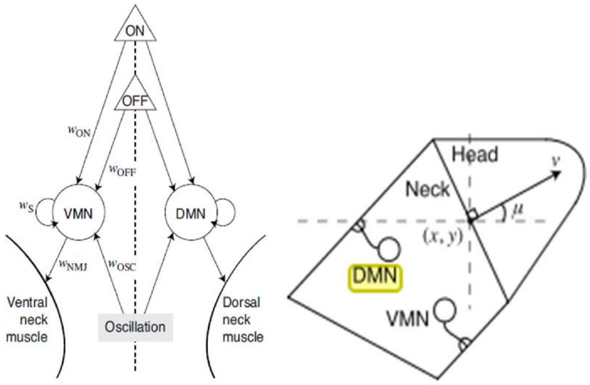

Figure 3. Left: The *C. elegans* neural circuit, reprinted from Izquierdo & Lockery (2010). Triangles correspond to chemosensory neurons; circles to neck-motor neurons. Arrows indicate neural connections whose strength and direction is determined by the evolutionary algorithm. **Right:** The body model. Neck-motor neurons controlled by the neural circuit govern head angle μ, influencing the body's direction of motion in the simulated environment.

This simulation relies on several simplifying assumptions. For example, the neural circuit does not include any interneurons, and approximates the nervous system's background activity as an oscillating signal that drives a snake-like motion characteristic of real-world *C. elegans*. In other respects, the neural circuit is true to biological detail. Specifically, it includes a pair of motor neurons analogous to SMBD (dorsal) and SMBV (ventral) that determine muscle contractions on either side of the body model, and a pair of chemosensory neurons (“ON” and “OFF”) that detect increases and decreases in the concentration of a chemical trace in accordance with the known response patterns of ASEL and ASER (Suzuki et al., 2008).

The aim of the study is to determine the functional relationships between chemosensation and motor action that contribute to klinotaxis in the simulated environment. To this end, Izquierdo & Lockery invoke an evolutionary algorithm (Mitchell, 1996) that, from a random “population” of neural circuits with distinct connectivity profiles, selects those circuits that lead to particularly efficient and reliable klinotaxis across varying environmental conditions. After evolving the population over several generations, the authors identify 77 successful neural circuits. Perhaps remarkably, although distinct in their neural connectivity profiles, all 77 circuits are observed to exhibit the same basic motor neuron response pattern to chemosensory stimulation: whereas stimulation of the ON cell (when the chemical gradient increases) reduces the differential activity of ventral and dorsal motor neurons and causes the body model to align itself with the direction of the source, stimulation of the OFF cell (when the chemical gradient decreases) has the opposite effect, causing the body model to turn away from the source. This motor neuron response pattern is one component activity of the mechanism for klinotaxis in the simulated environment.

A second component activity is environmental feedback from motor action back to chemosensory stimulation. Notably, Kaplan (2012) and Zednik (2011) have already argued that a mechanism may be physically distributed in this way, crossing the physical boundaries between brain, body and environment. In this particular simulation, the body’s snake-like motion makes the head repeatedly oscillate about the line of steepest ascent to the source (see figure 2). This oscillation results in alternating stimulation of the ON and OFF cells, which in turn cause alternating to-and-fro movements with respect to the source. Thus, neural feed-forward processing and environmental feedback together produce an effective displacement in the direction of the source: klinotaxis.

There are reasons to believe that the two component activities of the mechanism for klinotaxis in simulation are also operative in the mechanism for klinotaxis in real-world *C. elegans*. For one, targeted lesions in the simulated neural circuit have behavioral effects very similar to those of corresponding lesions in the biological organism (Izquierdo & Lockery, 2010: 12915). For another, the precise details of the sweeping motion observed in the simulated organism closely resemble those of real-world *C. elegans* in a range of environmental conditions (Izquierdo & Lockery, 2010: 12912). Importantly, this behavioral correspondence emerges unexpectedly: the evolutionary algorithm only selects for the efficiency and reliability of the resultant behavior, not its similarity to real-world klinotaxis. At the end of their study, therefore, Izquierdo & Lockery hypothesize that the component activities of the simulated mechanism may in fact closely resemble those of the real-world mechanism. The "how possibly" model of the mechanism identified by evolving a simulated organism may turn out to be a "how actually" model of the mechanism for klinotaxis in biological *C. elegans*.

It may seem surprising that an empirical hypothesis about the mechanism for klinotaxis in a biological organism can be derived on the basis of a mere simulation. Barbara Webb (2009) has recently questioned the legitimacy of inferences from simulated to biological mechanisms. Indeed, evolutionary algorithms such as the one adopted by Izquierdo & Lockery are notoriously exploitative of specific details of the simulated environments in which they operate. In the present study, the fact that all 77 successful neural circuits realize the same pair of activities might just be an artifact of the specific details of the simulated neural circuit, body model, and environment; in the real world, klinotaxis might be performed by entirely different means, or be produced by a multitude of distinct but redundant mechanisms. But this line of reasoning does not imply that simulation-based strategies are useless, only that they are fallible. Fallibility is a signature feature of heuristic strategies, and is usually offset by the simplicity and speed with which such strategies can be deployed (Bechtel & Richardson, 1993; Gigerenzer, 1991). Indeed, exploring the space of possible mechanisms for klinotaxis via the evolution of simulated model organisms is likely to be a far more effective use of time and resources than exploring it via behavioral and lesion studies of the biological organism—especially insofar as evolutionary algorithms and simulations are particularly adept at identifying unintuitive and complex solutions that might otherwise be overlooked (see also: Wheeler, 2005). Of course, conclusively determining whether or not any individual "how possibly" model of a mechanism is in fact a "how actually" model requires further testing, refinement, and eventual confirmation or falsification on the basis of empirical investigation. Still, viewed as a heuristic device for developing testable "how possibly" models of a mechanism, the heuristic role of evolving simulated model organisms is clearly significant.

3.2 Selective Pruning

Izquierdo & Lockery's simulation-based strategy identifies the (possible) component activities of the mechanism for klinotaxis in *C. elegans*, revealing these to be, on the one hand, particular motor responses to distinct types of chemosensory stimulation, and on the other hand, a specific kind of environmental feedback. What about the mechanism's component entities? As was mentioned above, the direct links between chemosensory and motor neurons in the simulated neural circuit are considerable simplifications of biological reality. In reality, several interneurons mediate between ASE chemosensory and SMB motor neurons. In a recent study, therefore, Izquierdo & Beer (2013) seek to determine which particular interneurons of the full *C. elegans* connectome contribute to the neural feed-forward processing described in Izquierdo & Lockery's earlier work.

To this end, Izquierdo & Beer invoke the heuristic strategy of *selective pruning*, in which the authors combine past experimental results with graph-theoretic measures to distinguish those elements of the connectome that are likely to be the component entities of the mechanism for klinotaxis from those that are not. The starting point of Izquierdo & Beer's study is a graph-theoretic representation that includes: all 12 chemosensory neurons known to detect concentrations of chemical gradients in the environment; all 28 head- and neck-motor neurons that determine the worm's movement; and all 234 interneurons, 6246 chemical contacts and 890 gap junctions that make up the structural links between them (figure 4A).

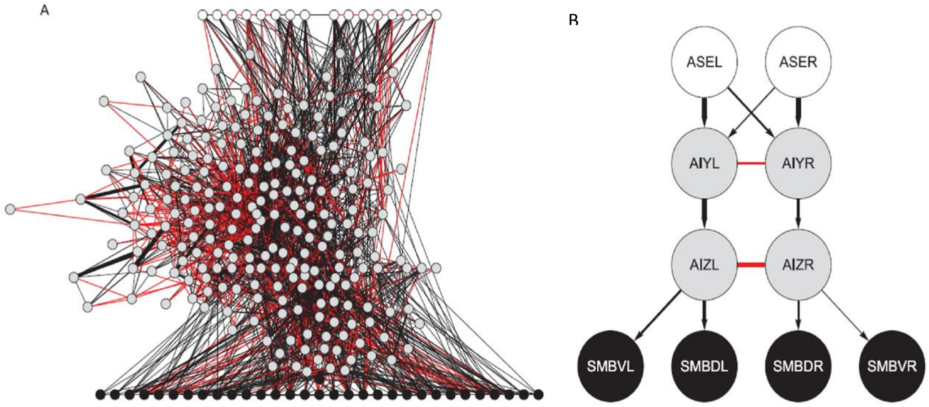

Figure 4. A: A subgraph of the *C. elegans* connectome depicting the network of neurons potentially relevant to klinotaxis, reprinted from Izquierdo & Beer (2013). Gray units represent chemosensory neurons, black units represent motor neurons, white units represent interneurons. **B:** The minimal network derived by selectively pruning the graph in figure A.

Although the 274 elements of the graph in figure 4A fall short of the 302 neurons in the full *C. elegans* connectome, they include all neurons that are *potentially* relevant to the production of klinotaxis. The excluded neurons have no inbound connections to the relevant motor neurons, and can therefore be removed from consideration. As the structural intermediaries between chemosensory and motor neurons, all interneurons represented in figure 4A are potentially relevant to the production of klinotaxis. However, not all of these interneurons need be functionally relevant: their contribution to klinotaxis may be negligible or redundant. In order to separate the relevant elements from the irrelevant ones, Izquierdo & Beer first appeal to previous experimental results to prune those chemosensory and motor neurons that have not previously been associated with klinotaxis (Bargmann & Horvitz, 1991): only ASE chemosensory and SMB motor neurons remain under consideration. Subsequently, the number of interneurons that link ASE and SMB is reduced by applying several graph-theoretic measures: removing all weakly connected elements, such as those that have less than two outgoing connections, and removing all long-range pathways by excluding those interneurons that are not immediately adjacent to either a chemosensory or a motor neuron (Izquierdo & Beer, 2013, p. 3). Through this kind of selective pruning, which is motivated by empirical as well as graph-theoretic considerations, the graph in figure 4A is reduced to the graph in figure 4B: the *minimal network* for klinotaxis by *C. elegans*.

The elements of this minimal network are probable candidates for the component entities of the mechanism for klinotaxis in biological *C. elegans*. First, the interneurons identified by the selective pruning strategy (AIZ and AIY neurons on the ventral and dorsal sides) are consistent with those identified in previous ablation studies of klinotaxis (Iino & Yoshida, 2009; Kocabas, Shen, Guo, & Ramanathan, 2012). Importantly, this consistency was achieved despite the fact that the selective pruning strategy was not designed to reproduce this empirical result. Second, when used as an artificial neural circuit controller for a *C. elegans* body model similar to the one discussed above, the minimal network (fleshed out with appropriate connection weights) produces effective and realistic klinotaxis behavior closely analogous to the behavior observed by Izquierdo & Lockery (2010). Third, and perhaps most important, Izquierdo & Beer show that this klinotaxis behavior is produced by the same interdependence of neural feed-forward and environmental feedback described in the earlier study. Indeed, the interneuron response to chemosensory stimulation in this network is qualitatively identical to the pattern of neural activity described by Izquierdo & Lockery: the minimal network implements one of the two component activities of the mechanism described earlier.

Whether or not the minimal network in figure 4B actually describes (some of the) component entities of the mechanism for klinotaxis in biological *C. elegans*, what is important for current purposes is the strategy of selective pruning that was used to discover this network. This strategy can be clearly distinguished from the heuristic approaches to mechanism discovery discussed in Section 2. Recall that the heuristic strategy of decomposition involves breaking apart a complex system or entity into a collection of simpler entities. In contrast, selective pruning already presupposes that the relevant system has been decomposed and its component entities have been identified. In this particular example, from a previously available description of individual parts (the *C. elegans* connectome), Izquierdo & Beer identify a particular subset of these parts as the possible component entities of the mechanism responsible for the phenomenon being explained. Unlike many previous studies of the *C. elegans* connectome, Izquierdo & Beer thus take a first step from detailed description to genuine explanation. Although these studies generally agree that not all parts of the connectome are the component entities of particular behavioral mechanisms—hence the appeal of targeted ablation studies—it is hard to know how to isolate the relevant connectome elements from the irrelevant ones. Using graph theoretic measures top selectively prune the connectome in the way exemplified here is likely to be of great help.

But the role of selective pruning in the epistemic practice of mechanism discovery is not limited to connectome research. Although technical advances in imaging, mapping, and computer modeling make it increasingly feasible to identify the individual components of different kinds of biological systems, it remains hard to know exactly which components actually contribute to the phenomena exhibited by such systems. For example, although researchers have successfully sequenced the genome of a variety of species, it remains unclear how best to systematically characterize the interactions between gene products, and thereby specify interdependencies in gene expression. Graph-theoretic methods similar to the ones invoked by Izquierdo & Beer have been used to separate strong protein interactions from weak ones (Schlitt & Brazma, 2007), as well as to identify patterns of interaction between multiple proteins that are repeated throughout a genetic regulatory network: network motifs (Banks, Nabieva, Chazelle, & Singh, 2008). Insofar as protein interaction networks can be viewed as mechanisms for gene expression, here again the heuristic strategy of selective pruning facilitates the discovery of biological mechanisms.

In summary, decomposition and localization are far from being the only useful heuristic strategies for mechanism discovery. The evolution of simulated model organisms and selective pruning can both be viewed as heuristic strategies that facilitate the identification of biological mechanisms and their components. But there are likely to be many others. Insofar as the true diversity of heuristic strategies remains unknown, the philosophical literature on mechanistic explanation is well-advised to consider more—and more recent—examples of biological research than the canonical ones outlined above.

4. Beyond Pictures: Mathematical Mechanism-Descriptions

The extant philosophical conception of mechanistic explanation, bolstered by the canonical examples reviewed in Section 2, emphasizes the distinctly visual character of the epistemic practice of mechanism description (see e.g. Bechtel & Abrahamsen, 2005; Bechtel & Richardson, 1993;

Wright & Bechtel, 2007). In contrast to deductive-nomological explanations, which take the form of logical arguments that link linguistic propositions (Hempel, 1965), the canonical examples of mechanistic explanation center on iconic or schematic diagrams. The prevalence of diagrams is due to the fact that most mechanistic explanations describe mechanisms that exist in space and time: their component entities have spatial properties, their component activities can be characterized in terms of changes to those spatial properties over time, and their overall organization is determined by the spatiotemporal arrangement of their components. Diagrams are ideally suited to represent this kind of spatiotemporal information because they are iconic in a way that other descriptive media are not: the spatial properties of diagrams can be used to "mirror" the spatial properties of the entities being represented, and diagram-sequences or animations can be used to visualize changes to the properties of a mechanism's components over time, such as the movement of an ion through a channel.

But mechanistic explanations are only contingently diagrammatic; mechanisms can also be described mathematically. The most straightforward way in which a mechanism might be described mathematically is by way of equations. Variables and parameters can be used to represent the properties of individual entities, such as their size, location, velocity, activation, or charge. Changes to these properties can be represented as changes in the values of the relevant variables over time, and relationships between individual entities or activities (e.g. their spatiotemporal or functional organization) can be captured in mathematical relationships between variables or coupled equations. Common examples once again include network models in cognitive neuroscience and in the study of protein interaction networks. Although such network models are often represented diagrammatically, as in figure 4, these diagrams are nearly always grounded on mathematical equations that precisely specify the interactions between elements of the network as well as the processing or transformation of information that occurs within individual elements.

Despite the existence of such examples, there is substantial disagreement concerning the suitability of mathematical equations for mechanism description. Consider:

> "Equations do not offer the right kind of format, however, for constructing a mechanistic explanation—they specify neither the component parts and operations of a mechanism nor how these are organized so as to produce its the behavior" (Abrahamsen & Bechtel, 2006, p. 171).

In later work, Bechtel & Abrahamsen (2010) explore the role of equations in mechanistic explanation, and conclude that these are generally used to complement, rather than provide, descriptions of mechanisms. For example, of the detailed mathematical models of the time course of circadian rhythms in *drosophila*, they say:

> “the models are not proposals regarding the basic architecture of circadian mechanisms; rather, they are used to better understand the functioning of a mechanism whose parts, operations, and organization already have been independently determined” (Bechtel & Abrahamsen, 2010, p. 322).

But although this analysis is surely true for certain examples, the claim that mathematical equations are generally unsuited for mechanism description is unfounded. As Craver (2006, 2007a, 2008) and others have already argued, what matters is not how a mechanism is represented, but simply that it is represented (see also: Glennan, 2002; Kaplan & Craver, 2011; Machamer et al., 2000; Zednik, 2011). Insofar as mathematical equations can be used to describe the very same properties as diagrams, there is no *a priori* reason to discount the former as vehicles of mechanism description.

Given the possibility of mathematical mechanism-descriptions, when would such descriptions be beneficial? Whereas diagrams are particularly useful for representing mechanisms whose relevant properties are spatiotemporal, mathematical descriptions seem particularly useful for representing mechanisms whose relevant properties are distinctly mathematical. Indeed, some mechanisms are best understood by describing the abstract mathematical properties of their component activities, such as the limit of their activation, their probability of occurring, or the general tendency of their motion. As a concrete example, consider the role of different kinds of mathematical representation in Randall Beer’s (2003) mechanistic explanation of perceptual categorization in a simulated brain-body-environment system (figure 5).

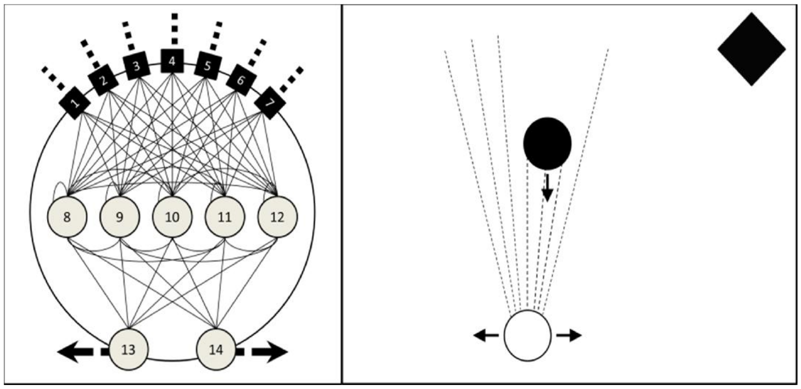

Figure 5. Adapted from Beer (2003). **Left:** The simulated agent and its continuous-time neural network brain. **Right:** The task environment.

This simulated system consists of a single “minimally cognitive” model organism that is embedded in an environment that features a single circular or diamond-shaped object. The model organism is equipped with a continuous-time recurrent neural network brain that mediates between visual inputs and motor outputs. The system’s behavior is determined by a set of 16 coupled differential equations:

(1)…(7)
$$\tau_i \dot{s_\imath} = -s_i + I_i(x, y; \alpha) \quad i = 1, \ldots, 7$$

(8)…(12)
$$\tau_i \dot{s_\imath} = -s_i + \sum_{j=1}^{7} w_{ji}\sigma(g(s_j + \theta)) + \sum_{j=8}^{12} w_{ji}\sigma(s_j + \theta_j) \quad i = 8, \ldots, 12$$

(13), (14)
$$\tau_i \dot{s_\imath} = -s_i + \sum_{j=8}^{12} w_{ji}\sigma(s_j + \theta) \quad i = 13, 14$$

(15)
$$\dot{x} = 5(\sigma(s_{13} + \theta_{13}) - \sigma(s_{14} + \theta_{14}))$$

(16)
$$\dot{y} = -3$$

Equations (1)-(16) define the change over time in the brain’s neural activity (*s_1*… *s_{14}*), the organism’s horizontal position (*x*), and the object’s vertical position (*y*). The brain’s neural activity is continuously affected by the changing sensory input vector *I*, a function of shape parameter *α* and of the relative positions of organism and object. In contrast, neural parameters *w*, *τ*, *σ*, and θ are fixed by an evolutionary algorithm that selects for successful categorization behavior in which a falling object is classified according to its shape. Specifically, the organism is evolved to “catch” circular objects by moving directly beneath them as they fall, and to “avoid” diamond-shaped objects by moving horizontally to either side. As an unexpected result of the artificial evolutionary process, successful organisms perform perceptual categorization via an “active scanning” strategy: they repeatedly move from side to side to “scan” the object before eventually settling on a position either directly beneath it or away to one side.

Insofar as equations (1)-(16) perfectly describe the parts of the brain-body-environment system as well as their interdependencies, there is a sense in which they explain the system’s behavior by describing the mechanism responsible for that behavior. Nevertheless, there is also a sense in which the description provided by these equations is not particularly insightful: it remains quite unclear exactly how the active scanning behavior arises from the interactions between the individual parts of the system. Indeed, as Craver (2012) has already argued, mechanistic explanation always involves choosing an appropriate level at which to describe describing a particular mechanism’s components. Thus, human circulation is typically explained at the level of organs and tissue rather than at the level of molecules, and long term potentiation is more easily illuminated by describing mechanisms at the level of molecules than at the level of atoms. In much the same way, Beer argues that it is far more insightful to decompose the system into two interacting components at a level above the individual neurons: the brain, embodied in the simulated agent (entity *B*), and the environment, defined by the relative positions of agent and object (entity *E*). The activities that correspond to entities *B* and *E* are, on the one hand, the influence of visual input on motor output, and on the other hand, the sensorimotor feedback in which motor output at any given time governs the accumulation of visual input at later times.

This way of verbally characterizing the two component entities and their corresponding component activities provides a “sketch” (Machamer et al., 2000) of the mechanism for perceptual categorization via active scanning. Beer turns this sketch into a detailed mechanistic explanation by invoking analytic techniques from dynamical systems theory. Specifically, *steady-state velocity fields* (shaded regions in figure 6; see original color version in: Beer, 2003) describe the activity of *B*, and superimposed *motion trajectories* (lines in figure 6; see original color version in: Beer, 2003) describe the activity of *E*. Whereas the latter provide a straightforward description of the relative positions of the model organism and the object over time, the former describe the organism’s steady-state (or long-term) velocity for every possible pattern of perceptual input: the horizontal velocity the organism *would* achieve if its motion were stopped and its perceptual inputs were held constant for an extended period of time.

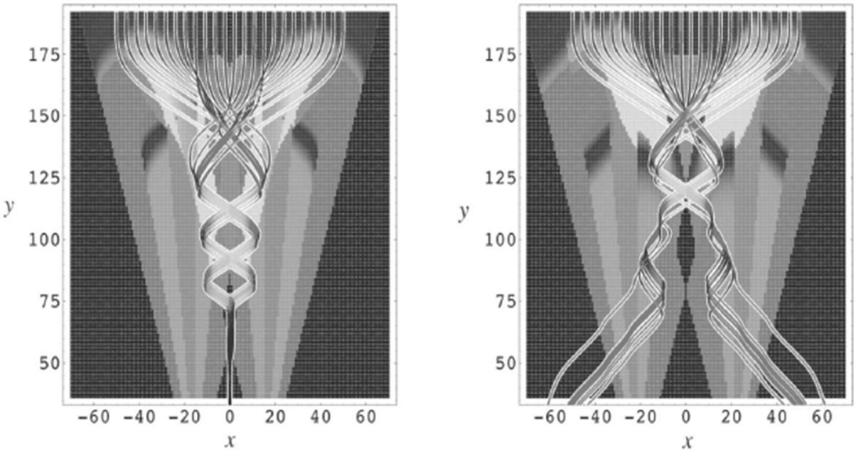

Figure 6. Steady-state velocity fields with superimposed motion trajectories for "catching" circles (**left**) and "avoiding" diamonds (**right**), adapted from original color version in Beer (2003). Axes designate the relative positions of agent (*x*) and object (*y*). Shaded regions describe the activity of the embodied brain (*B*): the way perceptual input constrains the agent's horizontal motion. Different shades (colors in the original) indicates different steady-state velocities, directed either toward or away from the object. Lines describe the activity of the environment (*E*): the way the changing relative positions of agent and environment affect perceptual input. Their shade (color in the original) indicates the agent's instantaneous velocity; their shape indicates the way the relative positions of object and agent change over time.

Gaining explanatory leverage from descriptions of a system's steady state-behavior, especially when that system's actual behavior is hard to describe, is a hallmark of dynamical explanation (Chemero, 2009; Kelso, 1995). Although dynamical explanations come in many different varieties (Zednik, 2011), this is also true when a particular dynamical explanation describes the components of a mechanism. In the current example, the organism's steady-state velocity acts as a constraint that limits its instantaneous velocity, and can therefore be used to approximate the activity of entity *B*, the embodied brain. Consider the way in which the motion trajectories in figure 6 overshoot some shaded regions while reversing their direction over others. What determines whether a particular motion trajectory performs an overshoot or a reversal within any particular region is that region's shade (or rather, its color) as well as the amount of time spent moving through it. Specifically, a motion trajectory of a particular shade or color performs a reversal whenever it is situated over a region of the opposite shade or color, and remains in that region long enough for the instantaneous velocity to approach the steady-state velocity denoted by that region. Because active scanning is just a particular pattern of overshoots and reversals, it can be reconstructed from the specific details—shape and shade/color—of the motion trajectories and steady-state velocity fields in figure 6. Indeed, differences between the left and right side of figure 6 such as the different sizes of the central black region explain the differences in behavior between circle (catch) and diamond (avoid) trials. Notably, modifying the size, shape, shade or color of particular regions (which can be done by e.g. changing certain network parameters) leads to novel and predictably correct or incorrect categorizations (Beer, 2003, pp. 228–230). That is, this description of the activities of *B* and *E* renders the mechanism for active scanning amenable to mutual manipulation—Beer's mathematical description of the two-component mechanism for perceptual categorization is adequate for the purposes of mechanistic explanation (for further discussion see: Zednik, 2011).

This example shows how equations, but also more sophisticated means of mathematical representation, can be used to describe the component entities and activities of mechanisms. Although figure 6 is of course also a diagram, it differs markedly from the diagrams invoked by the canonical examples described in Section 2 above. Whereas those diagrams typically describe a mechanism's spatiotemporal properties and "mirror" those properties in a relatively straightforward manner, the steady-state velocity fields in figure 6 describe a particular component entity's mathematical properties, and can be interpreted only on the background of the mathematical framework of dynamical systems theory. Irrespective of the perhaps unintuitive or hard-to-grasp nature of this description, what matters for current purposes is that it adequately describes (the component activities of) a mechanism.

In closing, there is of course no reason to believe that such mathematical descriptions can only be given for artificial examples like Beer's, and there is equally no reason to believe that only the framework of dynamical systems theory offers the right mathematical methods to describe mechanisms. Indeed, simulated model organisms analogous to Beer's have already been studied in information-theoretic terms (Williams & Beer, 2010), and descriptive techniques from dynamical systems theory are regularly invoked to describe developmental mechanisms and mechanisms for spatial memory in cognitive and developmental psychology (e.g. Spencer & Schöner, 2006). Moving even further beyond the traditional conception that mechanism descriptions are simple and diagrammatic in character will require paying closer attention to these and many other examples in which mathematical representations take center stage.

5. Extending the Scope of Mechanistic Explanation

Although the definition of mechanisms presented in Section 2 is intentionally broad—it includes all sets of "entities and activities organized such that they exhibit the phenomenon to be explained"

(Craver, 2007a, p. 5)—not everything that satisfies this definition can actually figure in a mechanistic explanation of a natural phenomenon. This is because mechanistic explanation is an epistemic activity that centers on the act of describing a mechanism, and because not all mechanisms can be feasibly discovered and subsequently described by practicing scientists. But what exactly distinguishes the mechanisms that can be feasibly discovered and described from those that cannot? Bechtel & Richardson (1993) invoke an influential analysis due to Herbert Simon (1996; see also: Wimsatt, 1986), in which systems are classified according to the degree of interactivity between their components. Systems within which the degree of interactivity is "negligible" (Simon, 1996, p. 207) are deemed *decomposable*: their behavior is an aggregation of the behavior of their components, and can typically be analyzed as such. Another class of systems—those in which the degree of interactivity between components is "weak, but not negligible" (ibid)—are deemed *nearly decomposable*. Although an analysis of a nearly decomposable system's behavior in terms of the behavior of its components will typically be approximate, it suffices to "understand, describe, and even 'see' such systems and their parts" (ibid). Thus, phenomena that arise from the activity of nearly decomposable systems are still amenable to analysis via the heuristics of decomposition and localization, and thus on Bechtel & Richardson's account, are subject to mechanistic explanation.

In Simon's classification, decomposable and nearly decomposable systems can be contrasted with *non-decomposable* systems, in which the degree of interactivity between components is on a par with the degree of activity within components. According to Bechtel & Richardson, non-decomposable systems resist the heuristic strategies of decomposition and localization, and thus, lie beyond the scope of mechanistic explanation. With respect to decomposition, this is because, absent independent criteria such as molecular composition or structure, interactivity is frequently a principle by which the components of a system are individuated. When the degree of interactivity is fairly uniform throughout a system, however, there may be no principled way to tell where one component ends and the next one begins. As concerns localization, although it is always possible to decompose systems arbitrarily, e.g. into equal-sized chunks of matter, it will be exceedingly difficult to identify each chunk's specific contribution to the activity of the system as a whole. In Craver's (2007a) terminology, it will be difficult to show that such arbitrarily individuated parts of a system are in fact the *working* parts of a mechanism—those parts that perform particular component activities.

Despite the widespread appeal of Bechtel & Richardson's account of the scope of mechanistic explanation, its ties to Simon's classification of system interactivity seem ill-motivated. Whereas Simon's classification is metaphysical—it concerns the structure of systems in the world—

Bechtel & Richardson's aims are distinctly epistemological: to distinguish those mechanisms that can be discovered and described from those that cannot. The success and failure of epistemic practices of discovery and description is dependent not only on the structure of the mechanisms being investigated, but also on the epistemic capacities of human investigators (see also: Glauer, 2012). What the examples introduced in the preceding sections of this chapter show is that scientists' capacities to discover and describe mechanisms are continuously evolving, especially with the influx of increasingly powerful computer simulations and sophisticated methods of mathematical representation and analysis. Therefore, it seems fair to wonder whether such simulations and mathematical methods might be used to extend the scope of mechanistic explanation.

Consider again the mechanism for perceptual categorization via active scanning. This mechanism features dense reciprocal interactions not only within the agent's neural network brain (component entity *B*), but also between the brain and the agent's changing environment (component entity *E*). Thus, the system is a non-decomposable system in Simon's sense, whose behavior, on Bechtel & Richardson's account, lies beyond the scope of mechanistic explanation. Nevertheless, Section 4 above shows how the system can be decomposed into two interacting entities, and how analytic techniques from dynamical systems theory can be used to approximately but still adequately (for the purposes of mechanistic explanation) describe the corresponding activities. Therefore, Beer's mechanistic explanation of perceptual categorization via active scanning is a counterexample to the claim that Simon's notion of non-decomposability determines the scope of mechanistic explanation: some non-decomposable systems (in Simon's sense) might after all be decomposed (in the sense relevant to mechanistic explanation).

Might Bechtel & Richardson's account be rescued by divorcing it from Simon's classification of system interactivity, and making it entirely dependent on the success and failure of decomposition and localization? On such a modified account, although the mechanism for perceptual categorization via active scanning is non-decomposable in Simon's sense, it would still lie within the scope of mechanistic explanation just because it can be decomposed and its component activities localized in the way demonstrated by Beer. But this account leads to an overly narrow conception of scientific practice. Section 3 shows that many heuristic strategies other than decomposition and localization contribute to mechanism discovery. It is not difficult to imagine that some of these alternative strategies may succeed even when decomposition and localization fail. Of the novel heuristics introduced above, the evolution of simulated model organisms seems particularly promising. The practice of artificially evolving a mechanism to reproduce a phenomenon in a simulated environment has a rich history of yielding particularly unintuitive or complex examples (Harvey, di Paolo, Tuci, Wood, & Quinn, 2005), many of which resist decomposition and localization but can nevertheless be described using sophisticated mathematical methods (Wheeler, 2005). Although it remains to be seen to what extent such simulated mechanisms can be used to reason about mechanisms in the real world in the way exemplified by Izquierdo & Lockery's study of klinotaxis, this is an empirical question best resolved by scientific research rather than by philosophical reflection.

In summary, the fact that practicing researchers frequently invoke heuristic strategies other than decomposition and localization, together with the fact that they rely on descriptive techniques other than verbal characterization and simple diagramming or animation, suggests that the scope of mechanistic explanation extends beyond the boundaries specified by Bechtel & Richardson. Exactly how far beyond? It is unclear that this question can—or should—be answered *a priori*. Insofar as the scope of mechanistic explanation depends (at least partly) on practicing researchers' epistemic capacities, and insofar as these capacities are continuously evolving, answering this question will involve closely considering future development in the strategies, methods, tools and concepts of scientific research.

6. Conclusion

One reason for considering novel approaches to mechanism discovery and mechanism description is to develop an improved conception of contemporary scientific practice. To this end, Section 3 shows that mechanism discovery goes beyond the heuristics of decomposition and localization, and Section 4 shows that mechanism description goes beyond verbal characterizations and iconic or schematic diagrams. Of course, the number and heterogeneity of heuristic strategies and descriptive techniques that contribute to mechanistic explanation in the life sciences is likely to even go beyond the examples considered here. Insofar as the philosophical conception of mechanistic explanation seeks to capture this number and diversity, there is no way around considering more, and more recent, examples from actual scientific research.

A second reason for considering such novel approaches is to force a reconsideration of the scope of mechanistic explanation—to delineate the class of phenomena that can be explained by describing the mechanisms responsible for them from the class of phenomena that cannot. Might these novel approaches be used to discover and describe mechanisms deemed too complex or too large to be discovered and described by the heuristic strategies and descriptive techniques thus far considered in philosophical discourse? It appears so. At the same time, it is unclear that the scope of mechanistic explanation can be properly determined until after the discovery and description of particularly challenging mechanisms is actually attempted. Unsurprisingly, the question of which phenomena can or cannot be scientifically explained is probably best answered by scientists themselves.

References

Abrahamsen, A., & Bechtel, W. (2006). Phenomena and Mechanisms: Putting the Symbolic, Connectionist, and Dynamical Systems Debate in Broader Perspective. In R. J. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 159–185). Oxford: Blackwell.

Banks, E., Nabieva, E., Chazelle, B., & Singh, M. (2008). Organization of physical interactomes as uncovered by network schemas. *PLoS Computational Biology*, *4*(10), e1000203. doi:10.1371/journal.pcbi.1000203

Bargmann, C. I., & Horvitz, H. R. (1991). Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in C. elegans. *Neuron*, *7*(5), 729–742.

Bechtel, W. (2006). *Discovering Cell Mechanisms*. Cambridge: Cambridge University Press.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441. doi:10.1016/j.shpsc.2005.03.010

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A*, *41*(3), 321–333. doi:10.1016/j.shpsa.2010.07.003

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Beer, R. D. (2003). The Dynamics of Active Categorical Perception in an Evolved Model Agent. *Adaptive Behavior*, *11*(4), 209–243; discussion 244–305. doi:10.1177/1059712303114001

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Chemero, A., & Silberstein, M. (2008). After the Philosophy of Mind: Replacing Scholasticism with Science. *Philosophy of Science*, *75*(1), 1–27. doi:10.1086/587820

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376. doi:10.1007/s11229-006-9097-x

Craver, C. F. (2007a). *Explaining the Brain*. Oxford: Oxford University Press.

Craver, C. F. (2007b). Constitutive Explanatory Relevance. *Journal of Philosophical Research*.

Craver, C. F. (2008). Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential. *Philosophy of Science*, *75*(December), 1022–1033.

Craver, C. F. (2012). Functions and Mechanisms: A Perspectivalist View. In P. Huneman (Ed.), *Functions: Selection and Mechanisms*. Springer.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

Gigerenzer, G. (1991). From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology. *Psychological Review*, *98*(2), 254–267.

Glauer, R. (2012). *Emergent Mechanism: Reductive Explanation for Limited Beings*. Mentis.

Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, *69*(S3), S342–S353. doi:10.1086/341857

Harvey, I., di Paolo, E. A., Tuci, E., Wood, R., & Quinn, M. (2005). Evolutionary Robotics: A new scientific tool for studying cognition. *Artificial Life*, *11*, 79–98.

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: Free Press.

Hubel, D., & Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, (148), 574–591.

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, (195), 215–243.

Iino, Y., & Yoshida, K. (2009). Parallel use of two behavioral mechanisms for chemotaxis in Caenorhabditis elegans. *Journal of Neuroscience*, *29*(17), 5370–5380.

Izquierdo, E. J., & Beer, R. D. (2013). Connecting a Connectome to Behavior: An Ensemble of Neuroanatomical Models of C. elegans Klinotaxis. *PLoS Computational Biology*, *9*(2). doi:10.1371/journal.pcbi.1002890

Izquierdo, E. J., & Lockery, S. R. (2010). Evolution and analysis of minimal neural circuits for klinotaxis in Caenorhabditis elegans. *The Journal of Neuroscience*, *30*(39), 12908–12917. doi:10.1523/JNEUROSCI.2606-10.2010

Kaplan, David M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*, 339–373. doi:10.1007/s11229-011-9970-0

Kaplan, David M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science*, *78*(October), 601–627.

Kaplan, David Michael. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy*. doi:10.1007/s10539-012-9308-4

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

Kocabas, A., Shen, C. H., Guo, Z. V., & Ramanathan, S. (2012). Controlling interneuron activity in Caenorhabditis elegans to evoke chemotactic behaviour. *Nature*, *940*, 273–277.

Leuridan, B. (2011). Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. *The British Journal for the Philosophy of Science*, *63*(2), 399–427. doi:10.1093/bjps/axr036

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *207*(1167), 187–217.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Schlitt, T., & Brazma, A. (2007). Current approaches to gene regulatory network modelling. *BMC Bioinformatics*, *8*(Supplement 6), S9. doi:10.1186/1471-2105-8-S6-S9

Silberstein, M., & Chemero, A. (2012). Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences. *Philosophy of Science*.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, MA: MIT Press.

Spencer, J. P., & Schöner, G. (2006). An Embodied Approach to Cognitive Systems: A Dynamic Neural Field Theory of Spatial Working Memory. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2180–2185).

Suzuki, H., Thiele, T. R., Faumont, S., Ezcurra, M., Lockery, S. R., & Schafer, W. R. (2008). Functional asymmetry in Caenorhabditis elegans taste neurons and its computational role in chemotaxis. *Nature*, *454*(7200), 114–7. doi:10.1038/nature06927

Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., & Chklovskii, D. B. (2011). Structural Properties of the Caenorhabditis elegans Neuronal Network. *PLoS Computational Biology*, *7*(2), e1001066.

Webb, B. (2009). Animals Versus Animats: Or Why Not Model the Real Iguana? *Adaptive Behavior*, *17*(4), 269–286. doi:10.1177/1059712309339867

Wheeler, M. (2005). *Reconstructing the Cognitive World*. Cambridge, MA: MIT Press.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *314*, 1–340. doi:10.1098/rstb.1986.0056

Williams, P. L., & Beer, R. D. (2010). Information Dynamics of Evolved Agents. In *Adaptive Behavior*.

Wimsatt, W. C. (1986). Forms of Aggregativity. In A. Donagan, A. N. Perovich, & M. V. Wedin (Eds.), *Human Nature and Natural Knowledge: Festschrift for Marjorie Grene* (pp. 259–293). Dordrecht: Reidel.

Wright, C., & Bechtel, W. (2007). Mechanisms and Psychological Explanation. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science* (pp. 31–79). New York, NY: Elsevier.

Zednik, C. (2011). The Nature of Dynamical Explanation. *Philosophy of Science*, *78*(2), 238–263.