UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Mental Mechanisms, Autonomous Systems, and Moral Agency

Permalink
https://escholarship.org/uc/item/1nq782s4

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors

Bechtel, William
Abrahamsen, Adele

Publication Date
2007

Peer reviewed

Mental Mechanisms, Autonomous Systems, and Moral Agency

William Bechtel (bechtel@mechanism.ucsd.edu)
Department of Philosophy, 0119, UCSD
La Jolla, CA 92093-0119 USA

Adele Abrahamsen (aabrahamsen@ucsd.edu)
Center for Research in Language, 0526, UCSD
La Jolla, CA 92093-0526 USA

Abstract

Mechanistic explanations of cognitive activities are ubiquitous in cognitive science. Humanist critics often object that mechanistic accounts of the mind are incapable of accounting for the moral agency exhibited by humans. We counter this objection by offering a sketch of how the mechanistic perspective can accommodate moral agency. We ground our argument in the requirement that biological systems be active in order to maintain themselves in non-equilibrium conditions. We discuss such consequences as a role for mental mechanisms in controlling active systems and agents' development of a self concept in which the self is represented as a moral agent.

Keywords: mechanistic explanation, mental mechanisms, autonomous systems, adaptive systems, self concept, agency.

Introduction

Explanation in cognitive science, as in other life sciences, primarily entails identifying and explicating the mechanisms responsible for phenomena of interest. Humanist critics deem such explanations incapable of explaining fundamental features of human existence, especially the capacity for moral agency in which the agent makes choices about what actions to pursue. We argue that even if this objection is warranted against mechanisms as commonly conceived, mechanisms appropriate to biological phenomena have the resources required for explaining such agency. This class of mechanisms can best be understood in terms of autonomous systems and their components.

Mechanistic Explanation

The common conception of mechanism is rooted in mechanical devices of the 13th-16th centuries (e.g., clocks, Gutenberg's printing press, and even mechanical animals). It found uptake in a strategy, relied upon extensively in the Scientific Revolution and still today, of explaining a phenomenon in nature by demonstrating how it could be produced from the activity of a mechanism. In the 17th century, for example, Robert Boyle explained air pressure by construing air as composed of spring-like particles, and Descartes posited that the heart acted like a furnace, heating blood so as to circulate it. Moreover, Descartes championed a *mechanical philosophy* in which this explanatory strategy was to be applied to biological phenomena in general, including any psychological processes that were not unique to humans. The particular biological explanations offered in the 17th and 18th centuries often relied at least as much on speculation as on evidence, given limitations in the available techniques. By the 19th century, though, biologists were obtaining data that allowed them to describe a variety of cellular phenomena (e.g., metabolism) and in some domains were making good progress towards mechanistic accounts (e.g., cell division). By the 20th century the mechanistic project had been extended to the new field of psychology, with models offered for a variety of phenomena in perception (e.g., detecting colors), cognition (e.g., memory encoding and retrieval), and affect (e.g., motivation).

.Despite robust pursuit of mechanistic models in biology and psychology, 20th century philosophers of science focused instead on the laws of physics. Accordingly, they built a deductive-nomological framework in which explanation was construed as subsumption under laws. (Hempel, 1965). Recently, though, certain philosophers of science have called for renewed attention to mechanistic explanation. Focusing on such domains as cell biology, molecular biology, physiology, neuroscience, and cognitive psychology, they note that explanation most often takes the form of identifying component parts and operations within a system and showing how they are organized to realize the phenomenon of interest (Bechtel & Richardson, 1993; Bechtel & Abrahamsen, 2005; Bechtel, 2006; Machamer, Darden, & Craver, 2000; Craver, in press).

To build on this rediscovered insight, philosophers of science must first articulate the essentials of mechanistic explanation. Most crucially, a scientist seeking to offer a mechanistic explanation must identify the system responsible for the phenomenon of interest and decompose it into component parts and operations. Typically quite different techniques are needed to identify the parts within a mechanism than to identify their operations, and these inquiries may be carried out in relative isolation in different disciplines. Eventually specific operations must be localized within specific parts of the mechanism. This requires additional investigations, and perhaps even additional research techniques or disciplines.

Mechanistic explanation offers a distinctive perspective on a number of issues in philosophy of science, notably, that of reduction. Mechanistic explanation is reductionistic insofar as it emphasizes the decomposition of systems into parts and operations. But it equally emphasizes that the parts and operations must be appropriately organized and the mechanism as a whole situated in an appropriate

environment. It therefore rejects the claim, often associated with reductionists working in the deductive-nomological framework, that resources at the lower level are adequate to achieve a complete account of the phenomenon of interest.

Although Descartes was perhaps the leading champion of mechanistic explanation, he also is notorious for attributing human mental activities not to physical mechanisms, but rather to an immaterial mind. Although he advanced several arguments that the mind is non-material, one of his chief motivations was the belief that humans exhibit phenomena that lie beyond the scope of any possible mechanism (e.g., constructing novel sentences or generating solutions appropriate to new circumstances).

The genesis of information processing mechanisms in the mid-20th century provided the resources needed to address Descartes' specific objections. Unlike physiological mechanisms that produce physical products (e.g., protein synthesis or the capture of energy in ATP), cognitive mechanisms are directed to mental and behavioral tasks and consequences (e.g., language comprehension, reasoning, regulation and coordination of movements). The inputs and outputs, as well as the internal states, of a cognitive mechanism serve as representations of entities and events external to the mechanism itself, and the operations upon these representations are designed to respect their content (Haugeland, 1981). Once programmable computers became available, artificial intelligence researchers sought to develop programs that would generate intelligent responses of the sort Descartes denied could be realized in a machine.

While the accomplishments in artificial intelligence and cognitive science give strong reason to believe that the limitations on mechanisms claimed by Descartes can be overcome, many humanists continue to oppose the project of explaining the mind mechanistically. For these critics, humans have a distinctive ability to act for reasons that they choose. Mechanisms, insofar as they involve purely causal processes, are fully determined in their responses and so lack the requisites for moral agency. In what follows we will sketch how mechanistic explanation has resources to answer the critics. In order to even begin to offer an adequate account of moral agency we will have to move beyond the common conception of mechanisms as purely reactive systems responding only when confronted with a stimulus. This is a move that has already been undertaken, though, by a few theorists who construe biological systems as *autonomous systems*.

Autonomous Biological Systems

Mechanists in biology in the 19th century were regularly challenged by vitalists who questioned the ability of mechanisms to exhibit certain features of living systems. Xavier Bichat (1805), for example, focused on the apparent indeterminism in the behavior of biological systems, particularly their capacity to "resist death" by actively countering the physical processes that threatened to destroy them.

Many mechanists simply ignored the vitalists' objections and, quite reasonably, went about their efforts to develop mechanistic explanations of the particular phenomena of interest to them. Others, however, tried to show how mechanists could address the objections. Claude Bernard (1865), for example, responded to Bichat's indeterminism claim by introducing a distinction between internal and external environments. He argued that the components of living systems were in fact responding in a deterministic manner to changes in the environment inside the organism, but that these internal activities might appear non-deterministic if one focused only on factors outside the organism. For example, glucose levels in the blood can remain relatively constant despite changes in the external environment of the organism, but this is due to operations in the liver that transform glycogen to glucose whenever blood glucose levels drop. Bernard further proposed that each organ contributed to *maintaining the constancy of the internal environment*. This provided a part of the answer as to how living organisms resist death—they are organized so as to maintain themselves in a constant state. Walter Cannon (1929) named this capacity *homeostasis*, and it came to be understood as involving negative feedback—a powerful way of organizing operations that was championed by the cyberneticists as providing a general control architecture for biological as well as social and engineered systems (Wiener, 1948).

As Rosenblueth, Wiener, and Bigelow (1943) demonstrated, negative feedback enables a mechanism to regulate its internal processes so as to pursue a goal. But it does not explain how a mechanistic system can establish its own goals. A potentially more productive approach is to inquire into what it would take for a mechanism to achieve Bichat's dictum of resisting death. For living organisms this is a greater challenge than Bichat recognized. Biological organisms must be highly organized to carry out multiple activities. Degeneration or decay is characteristic of any organized system since, as an *organized* system, it is far from thermodynamic equilibrium with its environment. For human artifacts, at least before we became a throwaway society, an independent repair person was commonly summoned to restore a machine when it broke down. But biological mechanisms typically cannot rely on such external agents—they must repair themselves. In the same spirit as the vitalists, Rosen (1991) argued that accounting for self-repair in the manner exhibited by living organisms requires a special kind of system, one closed to efficient causation.

Rosen took this as meaning that organisms are outside the scope of Newtonian science and as pointing to the need for a radically new, non-mechanistic theoretical framework. But in fact what distinguishes a system closed to efficient causation is that it is organized such that for each required operation, there is a part that is appropriate for performing it. What are the necessary operations for a system to maintain itself? Fundamentally, such a system must be able to recruit matter and energy from its environment and

deploy these appropriately in building and repairing itself. In his chemoton model, Tibor Gánti (1975; 2003) sketched how chemical systems might be organized to exhibit such features of life as self-construction and self-repair.

At the core of Gánti's chemoton is a chemical motor—a metabolic system that takes in energy-rich metabolites and transforms them chemically to extract the materials needed to continually remake itself. Using the Krebs cycle as a model, Gánti took these chemical reactions to be organized cyclically. That is, the final product of a sequence of reactions would combine with a new metabolite molecule and reenter the sequence, thereby continually replenishing itself. He in fact proposed reactions that would produce two molecules of the final product for each reentering molecule, thereby creating a continually growing body of material. Such reactions are *autocatalytic*. Some of the intermediate products of the metabolic cycle would be used to build and maintain a membrane surrounding the metabolic system, thereby providing a semi-permeable barrier between the chemoton and its external environment. The membrane allowed the chemoton to control what materials entered or left it and thus assure the appropriate conditions for continuing its metabolic processes. In this way the membrane provides an identity to the chemoton as an enduring entity partially independent of its environment. Finally, but less relevant for purposes of this paper, Gánti included an information system in the form of the construction of polypeptides whose length and sequence, like DNA, could store information.

A shortcoming of Gánti's approach is that, in focusing on balanced formulae to characterize operations in the chemoton, he did not address its energetic requirements. The energetic analysis is critical: highly organized systems like the chemoton are far from thermodynamic equilibrium, and no system can maintain itself in such a state without free energy. It is often noted that the consequence of the second law of thermodynamics—approach to equilibrium—can be avoided only in an open system. But it is not sufficient that the system be open to energy; it must also direct the flow of energy (as the chemoton directs the flow of matter) in ways that maintain its organization. It is with these considerations in mind that Ruiz-Mirazo and Moreno introduced the notion of *basic autonomy*, which they characterized as:

> the capacity of a system to *manage* the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control: (i) internal self-constructive processes and (ii) processes of exchange with the environment. Thus, the system must be able to generate and regenerate all the constraints—including part of its boundary conditions—that define it as such, together with its own particular way of interacting with the environment (Ruiz-Mirazo & Moreno, 2004, p. 240; see also Ruiz-Mirazo, Peretó, & Moreno, 2004, p. 330).

A highly important feature of autonomous systems is that they are inherently active. Without performing the activities needed to maintain themselves far from equilibrium, such systems simply decay and cease to exist. It is noteworthy that such constant activity is characteristic of living things, from single-celled organisms to highly complex primates. Watch a bird, a marmot, or even an amoeba. There is always activity. A marmot might run a bit, stop, look around, sniff the ground, dart in another direction. Organisms are typically not waiting to act. Even in the absence of stimulation from without, they are always doing things—if not acting overtly in their environment, they at least are performing basic physiological functions.

Mental Mechanisms in Autonomous Systems

The realization that biological organisms are inherently active, like all autonomous systems, is a crucial starting point for explaining how mechanical systems could exhibit moral agency. In an autonomous systems framework, the challenge is not to show how a moral agent could initiate activity, but rather to explain how an agent might regulate activity going on within itself. But before we can unpack this account, we need to consider how mental mechanisms might arise in autonomous systems.

Gánti's chemoton is active insofar as it recruits matter and energy from its environment and deploys these in its own growth and repair. If a chemoton were ever actually constructed, however, it would be completely dependent upon its proximal environment to bring resources to it and remove its waste products (which would otherwise be toxic to it). Some organisms (most plants as well of bacteria living in sulfur vents in the ocean) thrive while being dependent upon the reliable provisioning by their environment, but many others have adopted a different strategy, moving through their environments in pursuit of the requisite matter and energy. Accordingly, most bacteria, in addition to mechanisms for metabolism and for constructing their own bodies, possess flagella for swimming and sensory systems designed to detect energy sources (e.g., sucrose gradients). As a result, they are able to move in ways that facilitate self maintenance (e.g., they move forward when they detect an energy gradient or tumble randomly when no gradient is detected). Such systems are agents in that they carry out operations on their environment. In order to be effective as agents, however, autonomous systems need the resources to secure information about their environment and utilize it in directing behavior. That is, they require sensory systems to pick up information and downstream systems to process that information. Provided with such resources, autonomous systems are adaptive—they can regulate their actions appropriately to environmental conditions. Accordingly, Barandiaran and Moreno (2006) characterize systems with these additional resources as *autonomous adaptive agents*.

The fact that biological organisms must continue to capture and transform matter and energy from their environment so as to maintain themselves in existence provides a fundamental teleology to such systems (Bickhard, 2000; Christiansen & Bickhard, 2002). Unlike the teleology provided by cybernetic accounts, this stems

directly from the requirement that autonomous systems must either maintain themselves or die. Specialized mechanisms that evolve in such systems generally facilitate their ability to maintain themselves. Not all such mechanisms need be adaptations in the strict sense of having promoted the ability of their ancestors to reproduce (Brandon, 1990), but minimally, such mechanisms must not interfere seriously with the capacity of the organism to maintain itself. Otherwise they will cease to exist along with the whole organism. Moreover, they must be built and maintained by the organism itself and so are subject to the imperative on autonomous systems to maintain themselves or cease to exist.

This is not the place to pursue a detailed account of how sensory and information processing operations arise within an autonomous system, but one point is worth developing. The simplest autonomous systems are single-celled organisms. Already in such systems there is a division of labor between, for instance, a metabolic system, a membrane system, and a system for generating bodily constituents. When such a system reproduces itself, though, the daughter cells can either separate and live independently, or stay together. In the latter situation, the possibility arises of a division of labor, with individual cells specializing in different tasks needed for the maintenance of the whole rather than continuing as independent autonomous systems. In particular, the processes of securing environmental information could be segregated from the processes enabling locomotion. This, however, brings a new demand of coordinating the specialized components, a demand made especially pressing due to the fact that the tasks are being performed by autonomous systems whose default condition is being active.

In biological systems coordination is facilitated by yet other autonomous systems specializing in the process of communicating between components. In the evolution of biological organisms, this involved some cells capitalizing on the characteristics of semi-permeable membranes, which allowed for the establishment of differences in electrical potential across the membrane and the conduction of disturbances in these potential differences along the membrane (action potentials). Adaptive changes in form enabled these cells (now neurons with axons and dendrites) to facilitate communication between other cells. In organisms such as the jellyfish, a network of such specialized cells facilitates coordinated contraction of muscle cells in the lower rim of the body, resulting in forward propulsion. The neural regulation mechanism in these organisms is tightly coupled to their muscle capacities and is fundamental to their ability to utilize those capacities in maintaining themselves.

The simplest collectives, however, may be transitory, composed of individual cells working together sometimes but going their own way at other times. As division of labor proceeds, the individual cells are no longer able to meet all their needs to remain in existence on their own, and the collectives of autonomous systems become obligatory. The result is that the collective itself becomes an autonomous system, that is, a system that maintains itself in existence by recruiting and utilizing matter and energy from its environment.

As specialization continues, components can develop whose behaviors conflict with each other. Moreover, since the system is inherently active, each component will produce its behavior unless suppressed. Accordingly, the system requires a means of shutting off or down-regulating the mechanisms responsible for some behaviors while others are being performed. Insofar as such regulation enables the organism to pursue one behavior at the expense of others, the regulative processes constitute a decision-making system.

In order to ensure that individual components in the collective do not carry out activities when they would be counterproductive for the whole (on which each component now depends for its existence), there is evolutionary pressure to develop specialized regulatory systems. In the nervous system and in the body, new regulatory systems typically result from replicating existing components, with the daughter components assuming specialized regulatory operations, including the required information processing (Allman, 1999). Within the system for processing visual inputs, for example, new components took on the processing of specific types of information (e.g., motion, shape). The result is an array of specialized information processing mechanisms that together subserve the task of regulating motor activities in the collective autonomous system.

These specialized information processing systems constitute the mental mechanisms that have been the focus of the various cognitive sciences (Bechtel, in press). The key difference between the perspective we are advancing and that which has been common in cognitive science is that we treat these mental mechanisms, and the organism in which they are situated, as inherently active. . They do not passively await an appropriate input before responding. Specific input may modulate their activity, but even without input they are active. Recent modeling strategies in some fields of cognitive science manifest such a perspective. Beer's (1995) mixed controllers for model insect legs, for example, generate output even in the absence of sensory input but can respond to such input when available. Likewise, the perceptual models advanced by van Leeuwen and his collaborators (van Leeuwen, Steyvers, & Nooter, 1997; Raffone & van Leeuwen, 2001) utilize coupled oscillators that can be perturbed by inputs, but are active regardless of input. This enables them to simulate the shifting interpretations exhibited by humans when they look at ambiguous figures.

Mechanisms and Self Representation

A key feature of information processing mechanisms is that they operate on representations. In most autonomous systems, biological or artificial, the referents of representations are objects and events external to the autonomous system itself. But the capacity for

representation brings with it the possibility of self representation, which in turn gives rise to the possibility of an autonomous adaptive agent regulating its behavior in light of its self representation. It is with regulatory capacities of this kind that such agents begin to exhibit the characteristics of moral agents.

Ulric Neisser (1988) differentiated five aspects of self knowledge (he refers to them as different selves) that provide a fruitful framework for further developing the conception of moral agency within the context of autonomous adaptive agents. The first two selves are shared with animals that we do not usually construe as moral agents, but they provide an important foundation for understanding how self representations may acquire roles in controlling active agency. We briefly note these before turning to the other three types of self representation, which are more directly relevant to showing how mechanistic systems can come to exhibit moral agency.

First, the *ecological* self draws upon a Gibsonian ecological account of perception. Neisser emphasizes that visual information specifies not merely what is in the environment, but also the perceiver's position in that environment and the affordances for its action. Such information enables an autonomous system to utilize its motor capacities in the service of self maintenance. Second, the *interpersonal* self likewise specifies the autonomous system relationally, this time in relation to other autonomous systems. For an animal, representing relations to conspecifics, predators, and prey is particularly important, since such information is essential to executing those behaviors that will maintain its existence.

Neisser's first two aspects of self knowledge are necessary for an autonomous system to act in ways that enable it to maintain itself, but they do not support the sort of self regulation we would think of as specifically moral. One type of representation important for moral agency is of oneself as an enduring agent with a past and future. Such knowledge is provided by the Neisser's third aspect of self, the *extended* self, which provides for what Tulving (1983) characterized as *episodic memory*. Episodic memory allows an agent to re-experience its own previous experiences. Since the capacity to represent brings with it the capacity to misrepresent, these mechanisms can be employed to represent alternative pasts in addition to the one that actually occurred. Importantly for Neisser's account and ours, these mechanisms also can be directed to the future to envisage possibilities that might be realized. One way to explore empirically the importance of the extended self for realizing moral agency in autonomous adaptive agents is to examine the impairment of such agency in humans with impaired episodic memory. Extreme impairment of episodic memory in amnesic patients such as HM and KC renders them unable to regulate their actions in light of an envisaged future. (The effects of more limited memory deficits on moral agency requires further investigation.)

Neisser's fourth aspect of self, the ability to represent internal states of the self (Neisser calls this the *private* self),

has been a major focus of philosophers interested in the qualitative character of mental states. Although certain qualitative aspects of experience, such as experience of colors, may not be relevant to realizing moral agency in an autonomous system, the experience of emotion has increasingly been recognized as playing an important role in moral decision making (Damasio, 1995). The emergence of affective cognitive neuroscience offers a useful means of appraising the role of emotions in guiding agency.

Neisser's last aspect of self, which he terms the *conceptual* self, may be the most significant for providing moral agency in autonomous system. It relies upon our ability to conceptualize the world, especially in language, but refers in particular to our ability to represent ourselves conceptually. What is the content of a self concept and how does it affect agency? Wilfred Sellars' (1956) "myth of Jones" provides an instructive way of thinking about the issue. In his myth a group of individuals develops the resources to use language to theorize about the world and even engage in covert speech. One of them, Jones, advances a theory about the inner lives of himself and others that construes covert speech as thoughts and ascribes to thoughts a role in generating behavior. What Sellars is proposing is that characterizing people as having beliefs and desires is in fact a theoretical construct that turns out to be useful in anticipating others' behavior. On this construal, Neisser's conceptual self need not directly report on something mysterious found within oneself. It can instead be a theoretical account based on self-observation of what one is doing and has done in the world.

This self concept, on our proposal, develops as one of the regulatory mechanisms of an autonomous adaptive agent that is already acting in the world and needs to decide between available courses of action. Early in a person's development the options are highly constrained by the available physical capacities and social environment. As the person develops, though, these constraints are loosened and a variety of ways in which the self concept can regulate actions arise. Part of developing the self concept then involves representing the sort of person one is. If one represents oneself as non-violent, for example, that representation may serve to regulate one's behavior (and figure in developing habits of non-violent response). As an individual theorizes about her situation, she may develop moral conceptions that in turn figure in regulating her actions. In part such theorizing about oneself will be influenced by the ideas in one's culture and need not be an individual's own creation. But importantly, insofar as the self concept arises as part of the regulatory systems in an adaptive autonomous agent, it has potency in shaping the activities of the agent.

We have sketched a way in which humans might utilize a concept of self to achieve behavioral regulation. The phenomenon of weakness of will, though, suggests that there are limits to this capacity. Regarding humans as autonomous systems points to one reason why the self-concept may often be ineffectual: it is regulating an already

active system. This perspective suggests avenues of empirical investigation that could lead to better mechanistic accounts in cognitive science as well as a more nuanced mechanistic philosophy of science.

Conclusion

Our objective in this paper has been to sketch an account which reconciles the project of explaining human behavior mechanistically with a conception of humans as moral agents. At the core of this sketch is the idea of humans as autonomous systems and, more specifically, adaptive autonomous agents. Such agents are inherently active and, in order to maintain themselves, such agents require regulatory systems. On this view, mental mechanisms are regulatory systems. As they evolved, they provided agents with the ability to represent themselves. Thus equipped, agents could form representations of their personal characteristics and of their goals. Since these representations were available for use by regulatory systems in the inherently active agent, they provided the foundation for moral agency in a mechanistic system.

References

Allman, J. M. (1999). *Evolving brains*. New York: W. H. Freeman.

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior, 14*, 171-185.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (in press). *Mental mechanisms: Philosophical perspectives on the sciences of cognition and the brain*. Mahwah, NJ: Erlbaum.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421-441.

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

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

Bernard, C. (1865). *An introduction to the study of experimental medicine*. New York: Dover.

Bichat, X. (1805). *Recherches Physiologiques sur la Vie et la Mort* (3rd ed.). Paris: Machant.

Bickhard, M. H. (2000). Autonomy, function and representation. *Communication and Cognition--Artificial Intelligence, 17*, 111-131.

Brandon, R. (1990). *Adaptation and environment*. Princeton: Princeton University Press.

Cannon, W. B. (1929). Organization of physiological homeostasis. *Physiological Reviews, 9*, 399-431.

Christiansen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *Monist, 85*, 3-28.

Craver, C. (in press). *Explaining the brain: What a science of the mind-brain could be*. New York: Oxford University Press.

Damasio, A. R. (1995). *Descartes' Error*. New York: G. P. Putnam.

Gánti, T. (1975). Organization of chemical reactions into dividing and metabolizing units: The chemotons. *BioSystems, 7*, 15-21.

Gánti, T. (2003). *The principles of life*. New York: Oxford.

Haugeland, J. (1981). Semantic engines: An introduction to mind design. In J. Haugeland (Ed.), *Mind design* (pp. 1-34). Cambridge, MA: MIT Press/Bradford Books.

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331-496). New York: Macmillan.

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

Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology, 1*, 35-39.

Raffone, A., & van Leeuwen, C. (2001). Activation and coherence in memory processes: Revisiting the parallel distributed processing approach to retrieval. *Connection Science, 13*, 349-382.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose, and teleology. *Philosophy of Science, 10*, 18-24.

Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*, 235-259.

Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (2004). A universal definition of life: Autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere, 34*, 323-346.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science. I. The foundations of science and the concepts of psychology and psychoanalysis* (pp. 253-329). Minneapolis: University of Minnesota Press.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

van Leeuwen, C., Steyvers, M., & Nooter, M. (1997). Stability and intermittency in large-scale coupled oscillator models for perceptual segmentation. *Journal of Mathematical Psychology, 41*, 319-344.

Wiener, N. (1948). *Cybernetics: Or, control and communication in the animal and the machine*. New York: Wiley.