

EXPLAINING HUMAN FREEDOM AND DIGNITY MECHANISTICALLY: FROM RECEPTIVE TO ACTIVE MECHANISMS

WILLIAM BECHTEL AND ADELE ABRAHAMSEN
UNIVERSITY OF CALIFORNIA, SAN DIEGO

ABSTRACT: Mechanistic explanation is the dominant approach to explanation in the life sciences, but it has been challenged as incompatible with a conception of humans as agents whose capacity for self-direction endows them with freedom and dignity. We argue that the mechanical philosophy, properly construed, has sufficient resources to explain how such characteristics can arise in a material world. Biological mechanisms must be regarded as active, not only reactive, and as organized so as to maintain themselves far from thermodynamic equilibrium. Notions from systems biology make key contributions, particularly Gánti's chemoton, Ruiz-Mirazo and Moreno's basic autonomy, and Barandiaran and Moreno's adaptive autonomous agents. The reconstrual is then extended to mental life by conceiving of cognitive mechanisms as control components in inherently active systems, as illustrated in models offered by Randall Beer and Cees van Leeuwen.

Throughout the biological and behavioral sciences the term most often used in advancing explanations is *mechanism*. Thus, biochemists put forward mechanistic models of cellular respiration, molecular biologists describe the mechanism of protein synthesis, cytologists construct accounts of the mechanism of cell division, physiologists advance mechanistic models of blood cleansing by the kidney, cognitive psychologists generate models of the mechanisms involved in memory encoding, storage, and retrieval, psycholinguists propose accounts of mechanisms of sentence comprehension and production, and so forth. The quest to understand phenomena by identifying the mechanisms responsible for them has been enormously successful in the three centuries since Galileo, Descartes, and

others introduced what Boyle later termed the *mechanical philosophy*. Nonetheless, proponents of mechanism have always faced critics. In biology vitalists focused on what they thought were phenomena exhibited by living organisms that could not be produced by mechanisms. For the most part, vitalism has disappeared from biology with the increasing success of clever mechanists in identifying mechanisms that account for important features of living systems. But some of the challenges identified by the vitalists have remained.

Although Descartes applied the mechanical philosophy to all biological phenomena, and even to a wide range of behavioral phenomena exhibited by various organisms, he made an exception for mental activities that he took to be unique to humans: reasoning and language. For Descartes mind was a different kind of substance than body and was not subject to mechanistic explanation. Descartes' dualism was countered by the introduction of mechanistic models of human information processing in the twentieth century, which offered specific, explicit ways to bring reasoning and language into the explanatory mainstream. Nonetheless, the mechanistic construal of mind continues to be challenged. Many of the critics focus on the idea that humans, at least, act for ends that are in important respects freely chosen by agents. Such critics view human life as possessing an important kind of dignity in its ability to direct itself and regard mechanistic explanation of mental activities as an affront to that dignity.

Our goal in this paper is to show how the mechanical philosophy, properly construed, does not dismiss the features the critics have identified as central to human mental life. Instead, it has the potential to explain how such features can arise in a material world. This will require an important modification of the common construal of mechanisms as reactive, responding only when confronted with a stimulus. Biological organisms are active systems, and any adequate account of mechanism for either biology or psychology must have the resources to handle such systems. We will begin by characterizing the mechanistic program as it has developed in biology and cognitive science, with an emphasis on how it has successfully realized much of what Descartes envisioned in biology and overcame the limitations he thought it faced in psychology. Next we describe the concerns people often raise against proposed mechanistic accounts of mental life. We then turn to the question of how to adapt the notion of mechanism so as to capture the active character of biological mechanisms, showing how properly organized mechanisms can constitute active, autonomous systems. Finally we sketch how, taking this framework from biology as a starting point, we can advance an account of psychological mechanisms that addresses the concerns of skeptics. We offer two examples of modeling in cognitive science that capture the active character of mental mechanisms.

I. REALIZING DESCARTES' MECHANISTIC VISION IN BIOLOGY

Descartes proposed that all activities of non-human organisms, and those activities of humans which are also performed by non-human organisms, could be

explained in terms of mechanisms. Descartes' contemporaries were beginning to acquire the scientific techniques required to identify and understand the mechanisms operative in living organisms. A notable example was Harvey's (1628) discovery of the circulation of the blood, particularly his understanding of the heart as a pump and the valves in the heart as providing a means to control the flow of blood through the system. In subsequent centuries investigators began to articulate detailed accounts of the mechanisms responsible for many of the activities of living organisms. One of the central phenomena targeted for investigation was the generation of animal heat (a phenomenon not recharacterized as involving the procurement of energy for physiological activities from foodstuffs until the 1930s). Lavoisier and Laplace's (1780) demonstration that the process involved a kind of slow oxidation (burning) focused researchers' attention on localizing where in the organism such processes occurred. When nearly a century later Pflüger (1875) demonstrated that the process occurred in individual cells, this only served to narrow the question to the mechanism in cells that enabled them to perform the activity (Bechtel and Richardson 1993, chapter 3), a quest that occupied biochemists in the first half of the twentieth century.

Although the term *mechanism* is widely employed by biologists, most philosophical accounts of explanation have not attended to its use and have instead characterized explanation as involving subsumption of phenomena to be explained under laws (Hempel 1965). This presented a challenge for philosophers interested in biology, since there are relatively few biological laws. Beginning in the 1990s, however, some began developing an account of mechanism and mechanistic explanation in biology, especially targeting cases of mechanistic explanation advanced in cell and molecular biology and neuroscience (Bechtel and Richardson 1993; Glennan 1996 and 2002; Machamer, Darden, and Craver 2000). Although the terminology differs between the various characterizations of mechanism, there is considerable convergence. Our preferred characterization of a mechanism is:

A structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel and Abrahamsen 2005)

The key element in this characterization is that a mechanism consists of parts (organized spatially) which perform operations (organized temporally) that realize the phenomenon of interest. A mechanistic explanation of the phenomenon depicts or represents the mechanism and shows how it generates the phenomenon of interest. To achieve this, the mechanism taken to be responsible for the phenomenon of interest must be decomposed into its parts and the operations they perform. It can then be determined how these component parts and operations are organized such that the mechanism as a whole exhibits the phenomenon. The representation may take the form of a verbal description, but it is often more effective to depict the parts in a diagram and use icons or arrows to represent their operations. Understanding the dynamics of the mechanism—its orchestrated functioning in real time—may

further involve the use of mathematical representations, mental simulations, or computer simulations.

By articulating what is involved in mechanistic explanation, philosophers have begun to gain tools for characterizing the pursuits of biologists. One outcome is an explication of the high value biologists place on taking biological systems apart and explaining them in terms of their parts and operations—an explication that works without embracing the reductionism of philosophers in the longer-standing nomological tradition (Nagel 1961). Mechanisms are not simply parts, but rather organized systems of component parts and operations contextualized in an environment. Explaining phenomena mechanistically requires not just knowing the parts but their organization and context, which requires attending to higher levels of organization (Bechtel and Hamilton in press). A further advantage of the mechanistic perspective is that it quite naturally focuses attention on important questions for a naturalistic philosophy of science, such as how mechanisms are discovered (Craver and Darden 2001; Darden and Craver 2002) and how accounts of them are revised (Darden 1991 and 2006).

II. EXTENDING DESCARTES' MECHANISTIC VISION TO PSYCHOLOGY

Although most of the examples of mechanistic explanation developed by philosophers of science have involved biological (including neuroscience) cases, the notion of mechanism is also widely deployed in psychology. This is especially true in the cognitive tradition that developed in the 1950s and 1960s and the interdisciplinary projects of cognitive science and cognitive neuroscience in which cognitive psychology has participated. A central feature of the cognitive tradition is the idea that the mind is an information processing mechanism in which operations are performed on mental representations that generate new representations or behaviors. Memory, for example, is envisaged as involving the encoding of information, storing that information in a buffer, and subsequent retrieval of the information. A major feature of the development of cognitive inquiry was the development of techniques that could provide evidence about the information processing mechanisms employed. Miller (1956), for example, established a limitation on the size of the buffer for storage of information over short periods of times (e.g., the length of time between looking up a number in a phone book and dialing the phone). Such encoding seemed to be limited to seven plus or minus two items (a number reduced to five or less by more nuanced investigations later). An item might itself be complex as long as it formed a cohesive chunk. Thus, one could retain approximately seven letters or numbers if they did not form meaningful units, but also seven *chunks* of numbers or letters that formed meaningful units (e.g., seven words). Longer term memory did not suffer such limitations—the amount of information that can be retained for weeks or months once successfully encoded is essentially unlimited. Thus, an early focus of mid-century cognitive psychology was a distinction between

short and long term memory. Sperling (1960) showed that for very brief periods after a visual presentation people could, if appropriately cued, access any row or column in an iconic representation of items where the total number of items exceeded the limitations of short-term memory. This finding led to a three-store conception of memory and a model, articulated by Atkinson and Shiffrin (1968), in which attention mechanisms helped direct the flow of information between buffers (see Figure 1).

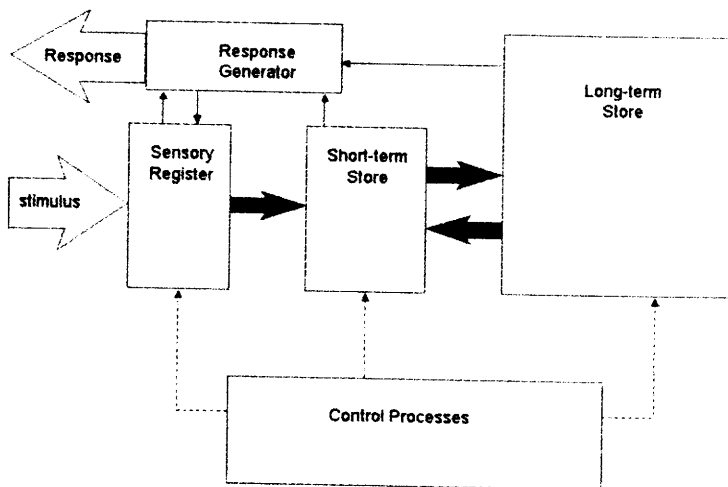


Figure 1. Atkinson and Shiffrin's (1968) memory model involving three different stores (sensory register, short term store and long term store) plus control processes that regulate how information is transferred between memory stores.

Although the notion of information and information theory was articulated independently (Shannon 1948; Miller and Selfridge 1950), the idea of information processing had its roots in the development of the computer. Turing's (1936) provided the conceptual foundation for modern computers and for the idea that information processing could be carried out by machines when he advanced his proposal of an abstract machine that could compute any specific decidable function, and a more complex universal machine that could compute all decidable functions. With the development of actual digital computers in the 1950s, some key researchers, including notably Herbert Simon, Allen Newell, Marvin Minsky, John McCarty, and their students, began to design programs to run on computers and perform the sorts of activities that are judged as intelligent when performed by humans (Newell and Simon 1956 and 1972; Minsky 1968). Underlying many of these efforts was the idea that computers can be interpreted as manipulating symbolic structures and that cognition itself might be explained by identifying rules that manipulate symbols.

A different tradition developed from the idea that individual neurons could be viewed as information processing devices, which got off the ground when McCulloch and Pitts (1943) showed how artificial neurons could compute basic logic functions. Subsequent work dropped the emphasis on logic and focused more on networks as devices for pattern recognition (Pitts and McCulloch 1947; Rosenblatt 1962). One advance that attracted attention to networks of neuron-like units was the development of algorithms for altering the strength of connections between units, which enabled networks to learn to recognize patterns without being specifically programmed to do so. The earliest learning procedures were limited to two layers of units with one layer of connections, but the introduction of a learning algorithm for networks with multiple layers of connections (Rumelhart, Hinton, and Williams 1986) created the possibility of networks learning any computable function provided they had sufficient numbers of units in their intermediate (hidden) layers and used nonlinear activation functions (for an overview, see Bechtel and Abrahamsen 2002).

Both symbolic and neural network modelers have produced impressive simulations of various cognitive activities. Moreover, they have effectively dispelled the doubt Descartes raised as to the ability of mechanisms to exhibit cognitive behavior (although some skeptics remain—see for example Dreyfus 1992). Although their capacities are subject to the creativity of their designers, such systems can generate inferences and construct novel, semantically appropriate sentences. But a notable feature of these models is that they are responders, not initiators—that is, they produce behavior only when input is provided. Thus, a problem solving symbolic model is given a representation of a goal and initial conditions as input and computes a solution to the problem. A neural network is given a pattern to recognize as input and produces an identification of the input on its output units. While there are more complex models within both traditions that partially overcome this limitation, the dominant orientation in thinking of mechanisms in cognitive science and neuroscience conceives of them as systems that respond to inputs. We will argue that this is a serious shortcoming in the understanding of information processing mechanisms, and mechanisms more generally.

III. CHALLENGES TO MECHANISM: EXPLAINING HUMAN FREEDOM AND DIGNITY

Descartes' objections to mechanisms as adequate to explain psychological capacities have largely been answered (at least in principle), but they have contemporary counterparts in many humanists' objections to the mechanistic perspective on human mental life that is integral to cognitive science. A major component of this opposition focuses on the idea that humans are responsible agents possessing freedom and dignity. We make choices in our lives, are responsible for those choices, and are thus unlike mechanisms that merely do what they were caused to do. Notoriously, Skinner (1971) responded to such concerns about behaviorist accounts of humans as products of conditioning by rejecting freedom and dignity as false ideals held over

from the Enlightenment. He maintained that these notions needed to be replaced by causal, deterministic accounts that would show how to arrange human life to make it as pleasant and productive as possible. Our strategy, unlike Skinner's, is to capture a robust account of freedom and dignity within a mechanistic perspective. To do so, we need to probe more deeply at the nature of the humanistic objection to mechanism.

Perhaps the most common way of putting the humanistic concern is to maintain that human agency requires freedom, whereas mechanisms behave deterministically. Only if agents are free to choose their own actions can they be held responsible for them. The notion of free will is itself a vexed one. Sometimes freedom is construed as the absence of causal determination. The appropriate sense of freedom for moral responsibility, however, cannot be that of being uncaused. If a person just did something that was totally lacking in causal origins (e.g., a person jumped up, stretched their arms, and yelled "I am the Devil," with that activity being totally unconnected to any preceding events including beliefs and desires of the person), that would not normally be construed as a free action. Minimally, for behavior to count as free it must originate in the mental activity of the agent—it must stem from the agent's reasoning and decision making. An important part of the pertinent notion of free decision is that such decisions are made in light of a person's reasons and values. Unless one can provide an account of how reasons and values could affect decisions other than causing them, causal considerations seem to play a role in free decision making.

A major reason many people seek nonetheless to advance a notion of freedom outside the causal nexus is the concern that if decisions to act were themselves caused, there would be no positive role for the agent to play in the generation of action. In part this stems from adopting a conception of causation as involving linear chains of events. If decision is only an intermediate in a causal chain between antecedent causes and actions, then the decision seems to be determined by the antecedent causes and alluding to the decision adds nothing to the account. In a related context of considering what is added in positing mental events between stimulus and behavior, Hempel (1958) characterized the problem as a *theoretician's dilemma* according to which positing intermediate theoretical mental events added nothing to the ability to explain and predict behavior. A modest step in showing the importance of the decision itself is to recognize that typically we are dealing not with linear chains of events, but rather with causal webs in which many factors contribute to the decision. Some of these are previous events in the life of the agent that helped to establish in the person the propensities to respond to new situations in particular ways, whereas others are immediate sensory inputs that impact on the person. In such a framework, decision processes cannot simply be passed over: it is precisely in these processes that the causal antecedents are combined to generate the behavior.

While such a perspective emphasizes the agent's decision in the causal determination of the behavior, it is unlikely to appease critics. The decision maker in

such a scenario might be viewed as not an active agent, but merely the container in which some of the causal processes transpire. Even if we emphasize the internal processes, the decision maker seems to be nothing more than a switch box in which various inputs are combined to yield an output—not a true agent in the sense of doing things and making decisions for itself. Viewing the decision maker as an information processing mechanism, one must posit the representation of goals for actions within the mechanism. But the represented goals do not seem to be the agent's own goals but simply ones imposed on the agent by its constitution and history. In an AI system, for example, they are viewed as the goals of the programmer, not of the computational system. The agent is not autonomous in the sense of giving direction to itself.

The difficulty is not limited to AI but applies to psychology and neuroscience as well. For the most part, practitioners in those fields have invoked responsive mechanisms—mechanisms that only produce their behavior when the appropriate stimulus is presented—to explain behavior. Mechanisms of this sort treat the organism as simply a way-station in causal processes, not as an agent. Such a view of mechanism is natural when the psychological phenomena themselves are construed as responses to stimuli. Thinking in terms of such mechanisms is further entrenched as a result of the research tools neuroscientists and psychologists deploy to explain behavior. We will very briefly illustrate with examples from two domains of mental phenomena: vision and memory.

In a vision experiment, the investigator typically presents discrete visual stimuli to subjects and measures some aspect of their responses. Some very robust regularities have emerged. In psychophysics, for example, as early as the nineteenth century Weber and Fechner discovered a simple nonlinear relationship in a variety of modalities between physical variables descriptive of the stimulus (e.g., the intensity of a light) and corresponding psychological variables (e.g., perceived brightness). Psychophysicists have tended not to try to explain the internal mechanism responsible for such regularities, but many other phenomena of vision have attracted considerable explanatory interest on the part of psychologists and neuroscientists taking a mechanistic perspective. Three strategies have been particularly effective. The first is purely behavioral—identifying regularities in the responses to various visual stimuli beyond those of psychophysics. For example, in one productive line of research stimuli that challenge the visual system are presented, generating illusions. Researchers then try to determine what kinds of information processing could produce such illusions. A second strategy has been to study humans or animals in whom parts of the brain have been damaged or removed. Investigators seek to identify the resulting visual deficits and infer the damaged part's normal role in the mechanisms of vision. Finally, using either electrodes or non-invasive techniques such as functional magnetic resonance imaging (fMRI), researchers adopting the third strategy seek to identify the response of different intact parts of the brain to particular kinds of visual stimuli. Such techniques have led to identification of numerous subcortical and cortical areas involved in vision

and to plausible hypotheses regarding the operations each area performs (Felleman and van Essen 1991; van Essen and Gallant 1994). Although vision researchers are generally well aware that there are as many or more recurrent (backwards) projections as forward projections in the brain areas engaged in visual processing, the resulting models focus on how incoming stimuli are processed through the system until a response is issued. That is, their research tends to approach the visual system as a mechanism that responds stepwise to stimuli, with the full implications of bidirectionality in the neural architecture set aside to be pursued later.

Much the same approach has been adopted with memory. A great deal of research has focused simply on identifying different types of memory—short-term or working memory, implicit memory, semantic memory, episodic memory, and so forth—with researchers often arguing that each of these is performed by a different mechanism (Schacter and Tulving 1994; Tulving 1999). Memories of each type are assumed to begin with the encoding event, to involve a storage operation, and to be made accessible to the investigator in a retrieval episode, usually initiated by presenting a memory cue. A further task undertaken by some investigators is to characterize the mechanisms that underlie such processes as encoding and retrieval. In the case of episodic memory, Tulving and his colleagues have argued that encoding involves operations performed in the left prefrontal cortex whereas retrieval relies on processes in the right prefrontal cortex (Tulving, Kapur, Craik, Moscovitch, and Houle 1994).

Such research endeavors in psychology and neuroscience, like many AI models, treat the subject as the locus of mechanisms in which responses are generated. Such a perspective on mechanisms fails to provide any sense of how the operation of these mechanisms renders the subject into an agent so as to answer the humanistic objections to the mechanistic view. What is needed is a means of construing the decision processes in an agent both as parts of a causal nexus and as reflecting self-determination on the part of the agent. This is a tall order, but one we contend the mechanistic view can meet. To do so, however, we must move beyond conceiving of mechanisms merely as responsive systems and understand them as parts of active systems doing things for themselves. Only then will we be in a position to meet the objections of mechanism's critics and account for human freedom and dignity. To see how this is possible, it is useful to turn to biology where a body of theorizing is developing that presents the needed apparatus to develop a conception of active mechanisms.

IV. FROM RESPONSIVE TO ACTIVE MECHANISMS IN BIOLOGY

The eighteenth and nineteenth centuries witnessed a struggle in biology between theorists who sought mechanisms to explain the various phenomena of life and those who maintained that the mechanistic approach was doomed to failure. The latter theorists, generally referred to as *vitalists*, offered a variety of alternatives to the mechanistic approach, but of far greater significance are the arguments they

advanced to demonstrate the inadequacy of mechanism. Two claims advanced by Xavier Bichat (1805) are particularly instructive. First, whereas physical phenomena behave in a deterministic fashion, living systems respond irregularly. Second, living systems tend to resist, in a way ordinary physical objects do not, physical forces exerted against them. He captured the last point with his aphorism that life is the “sum of all those forces which resist death.”

Most mechanists simply ignored the vitalists’ claims as they pursued the project of discovering the mechanisms responsible for particular biological phenomena. Sometimes they were confronted with temporary obstacles. For example, for forty years after Louis Pasteur (1857, 1858, 1860) claimed that fermentation could only be accomplished by whole living cells, more mechanistically minded physiologists and chemists tried but failed to produce fermentation in an extract free of cellular structure so they could study it as an ordinary chemical reaction. But finally Buchner (1897) succeeded, and numerous researchers picked up the trail of investigating the chemical mechanisms underlying fermentation. This effort culminated in the 1930s with a highly detailed account of the Embden-Meyerhof pathway (see Bechtel 2006). But the various individual successes of mechanistically minded investigators did not mean that the objections of the vitalists were successfully answered—the mechanists seldom addressed the phenomena vitalists insisted would prove problematic.

The few mechanistically minded physiologists who did take the vitalists’ objections seriously offered some interesting answers. Bernard (1865), for example, was greatly concerned by Bichat’s contention that physiological processes were indeterministic as this was a direct challenge to his attempt to provide strictly deterministic accounts of physiological phenomena. Bernard argued that indeterminacy was only apparent and was due to looking in the wrong place for the relevant causal factors that figured in deterministic regularities. For physiology, he maintained that one must look not to the external environment in which the organism lived but to the internal environment of other physiological processes within the organism. For example, it might be shown that glucose levels are relatively unresponsive to conditions in the external environment, but highly responsive to conditions inside the organism if that were where the regulatory causal processes were localized.

The introduction of the concept of an internal environment also provided Bernard with the start of a response to the claim that organisms resist death—he proposed that different organs of the body would carry out specific tasks so as to *maintain the constancy of the internal environment*. To follow up on the previous example, Bernard discovered that the liver served to synthesize glucose from glycogen whenever glucose levels in the blood dropped below a target level (Bernard 1848). Walter Cannon (1929) further developed this framework, introducing the term *homeostasis* to describe the tendency of living systems to maintain themselves in a constant internal condition and identifying a number of ways homeostasis was maintained in living organisms.

One means of achieving homeostasis involves negative feedback—the transmission of a signal from a product back to the process itself which can then regulate that process. Negative feedback provides a potent mode of organizing a system to achieve desired behavior, and was so celebrated by Wiener (1948) and other contributors to the cybernetics movement. Rosenblueth, Wiener, and Bigelow (1943) construed negative feedback as a means of introducing the notion of goal or purpose into a mechanistic framework. Whatever state was maintained through the process of negative feedback constituted a goal for the system. Thus, if a thermostat turns a furnace on whenever a room falls below a target temperature and off when it exceeds that temperature, the thermostat can be construed as having maintenance of the target temperature as its goal or purpose. Although providing an account of purpose in terms of negative feedback constituted a major conceptual advance, it did not explain how a system came to have a goal or purpose. At best, one could conceive of a system adopting a goal in the course of trying to satisfy a higher-level goal (e.g., an AI system) or coming to have a goal as a response to design (in the case of artifacts) or natural selection (in the case of organisms). Such accounts are inadequate as accounts of goals or purposes as found in living organisms in two respects. First, they cannot account for the ability of living systems to set goals for themselves. Second, the goals characterized in this manner are ones imposed *on* the system and not those *of* the system. That is, an observer may be able to detect whether the system is pursuing a goal, but the system does not have a goal for itself. This raises the question of how to conceptualize mechanisms that have and set goals for themselves.

We propose that greater headway may be gained by focusing on the second property of living systems that led Bichat to embrace vitalism: their engagement in activities that enable them to resist death. This is often treated in biology as the problem of repair—the problem of how an organism can continually reconstitute itself in the face of deterioration or damage. Degeneration or decay is characteristic of any organized system. With human artifacts, at least before we became a throw-away 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, some biological theorists (Rosen 1991) have argued that accounting for self-repair in the manner exhibited by living organisms requires a special kind of system, one closed to efficient causation. On Rosen's analysis, the efficient cause of all activities must be found within the living organism itself, which he claimed put organisms outside the scope of Newtonian science. He maintained that a radically new, non-mechanistic theoretical framework was required.

A different perspective is to focus on how mechanisms might be organized to exhibit the distinctive properties of living systems, including self repair (Bechtel 2007). What this organization must provide for is the recruitment of matter and energy from the environment and its use in building and rebuilding the organized system itself. In his model of a chemoton Tibor Gánti (1975 and 2003) provided a

sketch of how chemical systems might be organized to exhibit such features of life as self-construction and self-repair. He proposed the chemoton (Figure 2) as the simplest system that controlled its own synthetic activities, both in development and repair. At the core of the chemoton is a chemical motor—a metabolic system that takes in energy rich metabolites and transforms them chemically to extract the materials needed to build itself. Using the Krebs cycle as a model, he viewed these chemical reactions as organized cyclically. That is, the final product of a sequence of reactions would combine with a new metabolite molecule, thereby continually replenishing itself. He in fact proposed reactions that would produce two molecules of the final product on each pass through, thereby creating a continually growing body of material (such reactions are autocatalytic). Some of the intermediate products of the same reactions would be used to build and maintain a membrane surrounding the metabolic system, which would serve as 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 thereby to maintain itself in the appropriate conditions for continuing its metabolic processes. The membrane thus provides an identity to the chemoton. 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) could store information.

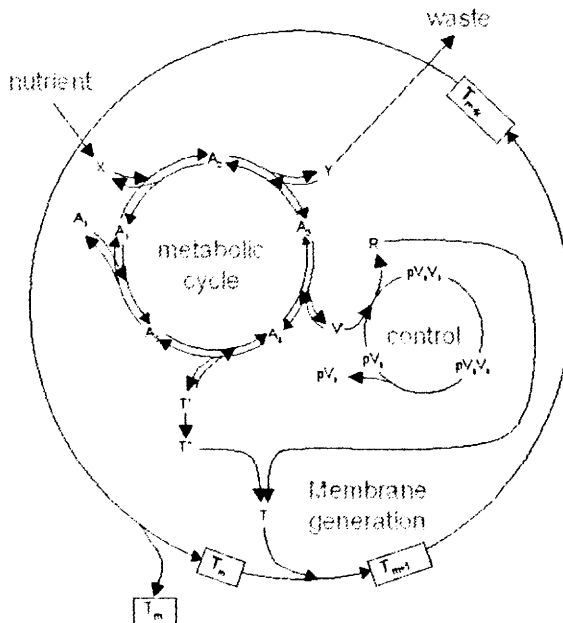


Figure 2. The chemoton proposed by Tibor Gánti as a model of the simplest chemical system exhibiting properties of living organisms.

By coupling a self-renewing metabolic system with a membrane construction system, Gánti found a way of organizing mechanisms that achieved not only self-maintenance but also, basically for free, self-reproduction. The reaction cycles resulted in a continuous increase in the amount of metabolite and membrane. But because the membrane grows faster than the space required for the metabolites, the membrane folds in on itself and buds.

A shortcoming of Gánti's approach is that while he focused on balanced equations in his characterization of the chemoton, he did not address its energetic requirements. The energetic analysis is critical since highly organized systems, such as 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 it is only possible to avoid the consequence of the second law of thermodynamics—approach to equilibrium—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 the 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 and Moreno 2004, 240; see also Ruiz-Mirazo, Peretó, and Moreno 2004, 330.)

(Varela 1979 proposed an account of autonomy that is similar in some respects, but does not emphasize energetic requirements.)

A highly important feature of autonomous systems is that they are inherently active systems. Without performing the activities needed to maintain themselves far from equilibrium, such systems simply decay and cease to exist. Moreover, 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—they are always doing things, whether acting overtly in their environment or performing basic physiological functions. Even in the absence of stimulation from without, they are continuing to perform activities. Thinking that activity is always goal-directed, we are often tempted to try to identify the goals behind all these activities. We might try to construct an account of how a marmot has the goal of exploring its environment, but this seems stretched, at least from the cybernetic perspective in which goals are controlled by feedback. A different perspective, which we find more plausible, is to treat being active as basic to autonomous systems and to view goal-directed behavior as constituting a further development within already acting autonomous systems.

V. MENTAL MECHANISMS AS CONTROL SYSTEMS IN AUTONOMOUS SYSTEMS

Gánti's chemoton provides a valuable tool for analyzing what kinds of organization are required to construct an active, autonomous biological system. If it were ever physically realized, though, it would be completely dependent upon its environment to provide it with a continuous source of matter and energy and to carry away its waste products. Some life forms, such as bacteria living in sulfur vents in the ocean, and many plants, can rely on such resources coming to them. Many life forms, including single-celled organisms, must actively pursue the resources they need. This requires that they move about appropriately in their environments. As Aristotle maintained, the ability to locomote and the ability to sense are tightly coupled since sensation is required to regulate movement. Thus, even most bacteria have, in addition to mechanisms for metabolism and for constructing their own bodies, 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 subserve their own self maintenance (e.g., move forward when moving up an energy gradient or tumble randomly when no gradient is detected). Such systems are agents in that they carry out operations on their environment and are adaptive insofar as they can regulate these actions appropriately to conditions that arise. Accordingly, Barandiaran and Moreno (2006) characterize such systems 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 kind of fundamental teleology to such systems (Bickhard 2000; Christiansen and Bickhard 2002). 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 arisen because they promoted the ability of the organism to reproduce (Brandon 1990), but minimally such mechanisms must not interfere seriously with the capacity of the organism to maintain itself, else they will cease to exist along with the whole organism. Moreover, they must be built and maintained by the organism itself and so are subservient to the imperative on autonomous systems to maintain themselves or cease to exist.

We have sketched a means of introducing a teleological framework that is intrinsic to autonomous adaptive agents. However, to account for goals or purposes of the sort that matter to human beings we need to take another step and consider how mental mechanisms, such as those involved in perception, memory, reasoning, and language, secure a place in autonomous adaptive systems. Minimally, like all other mechanisms within autonomous agents, these must be built by the organism and must not jeopardize the organism's ability to maintain itself. The key to understanding the role of such mechanisms in organisms is to recognize that as new mechanisms for behaving evolved, means for regulating or controlling these capacities were required so that they would operate in ways compatible with

each other and the continued existence of the organism. It is in this context that neurons likely developed. Taking advantage of the electric potential that resulted from cells having semipermeable membranes, certain cells presumably evolved specialized processes (axons and dendrites) through which electrical disturbances (action potentials) could be conducted over long distances. In organisms such as the jellyfish, a network of such specialized cells facilitates coordinated contraction of muscle cells in the lower rims 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.

This is not the place to advance a full account of brain evolution, but simply to suggest a framework in which individual mechanisms that evolved in brains through a process of replication and specialization (Allman 1999) would function to coordinate and regulate simultaneously evolving behavioral capacities. Such regulation is required since some of the behavioral capacities that an autonomous adaptive agent acquires are inconsistent with each other. 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.

Since behaviors have consequences, the decision making by the organism is, implicitly or explicitly, a choice of ends to be pursued at a given time. A tempting view to hold is that all decisions must be based on goals or ends to be achieved, with the most basic goals or ends being fixed by natural selection. This leads back to an account of teleology grounded in natural selection (Wright 1972). But such a move is both problematic and unnecessary. As long as the organism has the capacity to regulate its capacities to act, there is no need to assume that its basic goals or ends are fixed by natural selection. Many of the decisions it makes between courses of action may serve no particular evolutionary end (i.e., they may not render the organism more fit than its competitors). As the autonomous adaptive system lives in the world, situations requiring decisions arise and the regulatory system in place makes some decision and thereby pursues one end and adopts it rather than another. Although one might view the overall capacity to adopt goals and coordinate behavior in their service as a product of evolution, perhaps via natural selection, the resulting agent is typically free to operate on its own, establishing goals through the decisions it makes, pursuing them, and abandoning them. The goals such an organism adopts are then its own in the sense that, but for the system happening to adopt such a goal, it would not have it.

To illustrate this perspective, consider again the marmot described above. For a period it might seem to be going somewhere in particular or looking for something, and these activities may be elicited by what it sees or smells around it. But its interest in one pursuit might be abandoned at any time for another. Much of our own human behavior has a similar character. Consider going for a walk in the

country. Various things capture our interest—a flower, a rock, or a particular view. We move to a particular spot, stop and investigate, maybe move around a bit to get a better look, but soon break off and return to our walk. Temporary goals such as getting a good look at the flower arise and guide our behavior, without these being obviously in the service of any greater goal.

In scenarios such as that above, stimuli from the environment can perturb the behavior of the system, but the system is not just reactive. The marmot or the human is an autonomous adaptive system, acting in the world as it maintains its own existence. As it encounters new stimuli, the operations of its various mechanisms are perturbed. But they were already active. They may alter their response given the perturbation, but the activity is driven from within. A consequence of this is that a goal need not be adopted to somehow energize the organism to act. All that is required is to regulate existing activity.

One obvious difference between marmots and humans is that humans develop and maintain much longer-term goals, such as reading or writing a manuscript, acquiring an education, or saving for retirement. When people are asked to describe themselves they describe goals such as these, not those manifested in on-line regulation of motor activities. Although we can only provide the briefest sketch here (see Bechtel in press, for a more developed account), there is a rather natural way to extend the framework just introduced to handle such cases. What are required are the appropriate mental mechanisms. Many of the mental mechanisms supported by the neocortex in mammals and especially in primates serve to categorize experiences so as to allow knowledge acquired about the category to be applied to new instances. Our human linguistic abilities enable us to name these categories and develop explicit theories about them. If these linguistic representations are to be functional in our lives and not be epiphenomenal, they must figure in the sorts of control over behavior obtained through other mental mechanisms as well.

In the course of developing representations of things relevant to its life, an autonomous system may represent itself as an autonomous adaptive agent operating in the world. People represent many things about themselves, including events they have experienced, but especially aspirations and goals. These contribute to the control mechanisms that regulate behavior, enabling people to refrain from activities incompatible with their goals. Although such control is fallible (witness the phenomenon of weakness of will), it is this capacity to make and pursue goals that enables people to operate as agents with freedom and dignity. Freedom and dignity are real, but they are obtained through mechanisms that enable us to represent goals and regulate our behavior accordingly. Critically, however, such mechanisms develop and operate in a regulative manner in an already active autonomous adaptive system.

VI. MODELING ACTIVE MENTAL MECHANISMS

Our diagnosis of the basis for humanistic opposition to mechanistic accounts of mind is that mechanistic accounts commonly construe mental mechanisms as

responsive, not active. The account of autonomous adaptive agents provides a framework in which mental mechanisms are constituents of active systems, ones that are constantly engaging the environment so as to maintain themselves as autonomous systems. In this final section we will offer two examples from modeling research in cognitive science that point to how we can develop accounts of such mechanisms. The first models a control system for an active agent; the second models mid-level perceptual processing. Both involve neural networks, but these particular networks can be viewed as active systems that operate independently of input and so are not merely responsive systems. Rather, input serves to modulate the ongoing behavior of these systems.

One way to make a neural network model active is to employ interactive rather than purely feedforward architectures. With recurrent (backwards) connections, activity can recycle through the network and affect future activity. The most familiar networks with recurrent connections are designed to respond to stimuli by settling over time into a stable configuration (Hopfield 1982), but that is not the case in the ones we consider here. These networks continue to progress through new states or cycle through old ones as a result of their recurrent activity.

Randall Beer (1995) designed an artificial insect in which each of six legs was controlled by a network of just five units (artificial neurons). They were fully interconnected—connected to themselves and each other. Three of these controlled features of the leg's motion. In the simplest version of his project, Beer employed the genetic algorithm to establish weights on the connections that enabled the network to generate walking behavior. The resulting networks, which Beer referred to as *autonomous controllers*, generated a very stereotyped walking motion that was not responsive to the terrain. To analyze how they did so, Beer deployed tools from dynamical systems theory. He found that the networks had developed limit cycles—that is, the activation values of the units went repeatedly through the same cycle.

The autonomous controllers produced behavior, but not adaptive behavior. In other simulations, Beer also included a sensor unit that fed input to the other five units based on the angle of the leg. Since this angle is determined not just by input from the controller but also by the terrain on which the model insect is walking, this sensor indirectly provides information about the environment. In some simulations (*coupled controllers*) the input was always available as the genetic algorithm was employed to settle on weights, and in others it was only sometimes available (*mixed controllers*). The coupled controllers produced walking that was more responsive to the terrain than the autonomous controllers, but it turned out they were dependent on the sensory input. If it was turned off, they simply stopped. These networks became, in the language we are using in this paper, reactive rather than autonomous mechanisms, albeit ones with some adaptivity.

The most interesting of Beer's networks were the mixed controllers. By virtue of being trained some of the time without input, these networks exhibited a limit cycle which enabled them to generate walking behavior without sensory input. This can be viewed as an intrinsic dynamic of the system. The cycle, however, was

subject to modification when the sensor was active. This enabled walking behavior that was adaptive to the local landscape.

Beer's mixed controllers illustrate how internal dynamics can drive the basic behavior of a motor system but nonetheless allow sensory stimulation to modulate the behavior. The second example, due to Cees van Leeuwen and his collaborators (van Leeuwen, Steyvers, and Nooter 1997; Raffone and van Leeuwen 2001), illustrates the same principle in modeling a more cognitive activity—one in which we can all experience the active nature of our underlying cognitive mechanisms. In most perceptual contexts perception does seem passive—an object is present or an event occurs in front of our eyes and we recognize it. But some stimuli, such as the Necker cube, are ambiguous and, assuming we are able to see the stimulus in two or more ways, we shift spontaneously from one interpretation to another (we can also, sometimes, induce such a shift, but the shifts of interest are those that happen spontaneously as a person continues to gaze at the stimulus). Since there is no change in the external stimulus, the origin of the change must be in us.

To model this phenomenon, van Leeuwen used a rather unusual activation function for the units in his network—a logistic map function which in its simplest version simply updates the activation of a unit (a) based on its previous activation using the following equation:

$$a_{t+1} = A a_t (1 - a_t)$$

As shown in Figure 3, depending on the value of the parameter A , this function results in the unit settling to a fixed activation, oscillating through a limited number of values, or oscillating in a chaotic fashion. In his models, van Leeuwen made use of values of A resulting in chaotic oscillations.

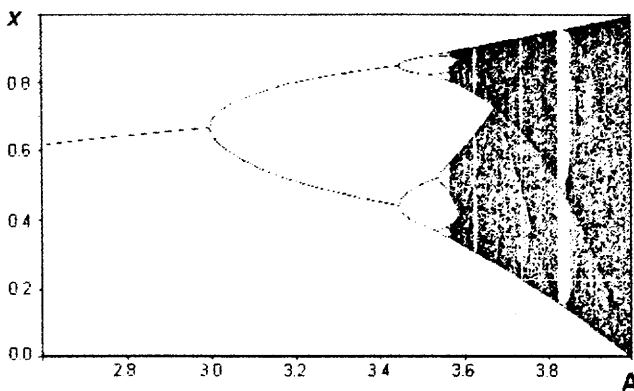


Figure 3. Plot of the logistic map (after transient activity) for values of A between 2.6 and 4.0. For values of A less than 3.0, the function settles into a point attractor. Above 3.0 the attractor bifurcates into two attractors, with successive values of the function falling into different attractors. It then bifurcates into four attractors, then eight, and so forth. Beyond 3.6 it enters a chaotic regime in which each successive value is distinct. The region beyond 3.6 is punctuated by narrow ranges in which there are again periodic attractors.

To model units that can couple their oscillations, van Leeuwen modified this simple activation function by first calculating the net input to each unit based both on the previous activation of the unit (a_i) and of the other units (a_j) to which it is connected according to the equation:

$$\text{netinput}_i = \sum C_{ij} a_j + (1 - C) a_i$$

C is a coupling parameter which determines how much the activation of a unit is affected by its own previous state and how much by the states of other units. The activation of a unit is then computed by the equation:

$$a_{i,t+1} = A \text{netinput}_{i,t} (1 - \text{netinput}_{i,t})$$

As before, the value of A determines whether the units behave periodically or chaotically. Even with values of A that generate chaotic behavior, the units can still synchronize, but when they do they will also have a tendency to decouple.

As in Beer's autonomous and mixed controllers, networks configured in this manner continue to operate without inputs on the basis of their own previous activity. In fact, the network described so far has no external input. But assume each unit is a sensory neuron which has a receptive field and receives an external input whenever something is present in its receptive field. Synchronized firing of multiple units then can be viewed as representing a stimulus that fills the receptive fields of the firing units. To enhance synchronization of activity, van Leeuwen and his collaborators replaced parameter C with variables representing adaptive weights for each connection. When the activation sequences of two units begin to synchronize, the weight between them is increased; this favors greater synchronization in the succeeding time-steps. As a result, synchronization that initially just happened to occur between two chaotic oscillators gets grabbed and used by the system to move it towards more structured activity. In a sense, the weights serve as a short-term memory of recent synchronization that helps to reinstate that synchronization. With this occurring across multiple pairs of units simultaneously, the system can advance towards larger clusters within which all units in the cluster are synchronized. The synchronized firing of a set of units can be viewed as an interpretation of an input stimulus.

Van Leeuwen and his collaborators have employed such networks to interpret visual input patterns such as those shown in the left and right panels of Figure 4. Here the network interprets the black squares on the left as forming vertical lines and those on the right as forming horizontal lines. In these cases van Leeuwen's network exhibited behavior that would be expected of any pattern recognizing feedforward network. The network reveals its distinctive characteristics when presented with stimuli such as that in the middle panel. Such inputs, along with the internal dynamics of the network, prevent it from settling permanently into one pattern of synchronous firing that represents just one interpretation. Rather, the network begins to settle into a pattern of synchronous firing for a period, but then it escapes and begins to settle into a competing pattern, then escapes again and begins to settle into the first pattern, and so forth.

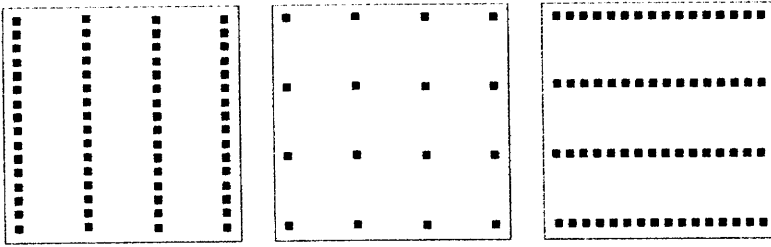


Figure 4. Stimuli used in van Leeuwen, Steyvers, and Nooter (1997). If the gestalt principle of symmetry is used to group items, the black squares in the left display will be grouped vertically and those in the right display horizontally. Those in the center display, however, will be ambiguous and subjects may alternate between grouping them vertically and grouping them horizontally.

Beer's and van Leeuwen's models point to an approach to modeling cognitive mechanisms that is rather different from standard reactive models. Each of them achieved a model that exhibits behavior affected, but not completely determined, by external stimuli. Rather, they are governed by internal processes that maintain them as active even in the absence of stimulation. Employing such continually active components, Beer and van Leeuwen point a way to modeling cognition through active mechanisms.

VII. CONCLUSIONS

In the two cognitive models described in the last section, the systems maintained themselves through internal dynamic processes but were subject to modulation by the environment. As a result, they produce behavior that is more like that of real behaving systems than do the merely reactive mechanisms employed in many models of cognitive mechanisms. Granted, there is still a great conceptual difference between these models and actual agents with freedom and dignity. But by starting with such models of active mechanisms that develop systems of control, we are in a much better position to see how agents that are completely constituted of mechanisms might acquire freedom and dignity.

A critical element of the account is that agency of the sort that involves freedom and dignity arises in systems that are already autonomous adaptive agents. Autonomous systems are active systems; without activity they cannot maintain themselves in existence as organized systems. Various mental mechanisms can be built on this foundation, but they do not supplant the foundation itself. In humans these mental mechanisms involve capacities to represent not just the world around us, but also ourselves as agents. Part of representing ourselves is representing goals we have chosen, and these representations may guide our behavior by modulating the operation of an internally driven dynamical system. Our self concept does not need to initiate activity *de novo*, but only to modulate the activities that we, as

autonomous systems, already are engaging in (see Bechtel in press, chapter 7, for further discussion).

Support for the perspective advanced in this paper comes from noting that human (and animal) development begins with organisms that are active in the environment but relatively unconstrained in their behavior. Babies are actively engaging their environment but without a great deal of discipline or structure. They look around, move their limbs, and vocalize, but without much purpose (the exceptions are when they are hungry, uncomfortable, and want attention). Over time their movements become more coordinated and their vocalizations approximate those of language, first through babbling and eventually through the production of words and sentences. One way to view the process is that internally active systems have gradually become more coordinated and so are able to achieve more complex goals. This opens up the possibility of choosing more things to do and entraining one's behaviors in the service of those selected ends.

Even as mature adults who engage in rational planning and execution of complex goals, the spontaneous, active nature of our mental systems remains apparent, although often overlooked or downplayed. It is difficult to maintain attention and focus—our minds and our perceptual systems wander and we move our bodies without thinking about it. It takes great disciplining to remain steadfast in our chosen pursuits and not to be tempted by other opportunities for action. It is difficult to maintain our focus even when we have a well-articulated set of goals and have trained ourselves to pursue them. The command not to look at anything or think about anything is nearly impossible to obey; meditators must practice extensively before they are even able to approximate such a state by invoking the requisite controls. The underlying active nature is evident in the effort it takes to regulate it.

We have only sketched how the mechanistic perspective can ground an account of agents with freedom and dignity when it is recognized that biological mechanisms operate as autonomous adaptive agents and that complex mental mechanisms are additions to such a foundation. This account may fail to satisfy the humanist critic. The freedom is not freedom from causation, but the freedom of a system that is directing its own engagement with its environment. The dignity is not the dignity of escaping the causal nexus of the world, but of a mechanistic system representing itself and selecting ends for itself among the options before it. But it is, we contend, a substantial sort of freedom and dignity, and one that surpasses what many thought was possible in a mechanistic world.

BIBLIOGRAPHY

- Allman, J. M. 1999. *Evolving Brains*. New York: W. H. Freeman.
- Atkinson, R. C., and R. M. Shiffrin. 1968. Human Memory: A Proposed System and its Control Processes. In *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 2, ed. K. W. Spence and J. T. Spence. New York: Academic. 89–195.

- Barandiaran, X., and A. Moreno. 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.
- . 2007. "Organization and Biological Mechanisms: Organized to Maintain Autonomy." In *Systems biology: Philosophical Foundations*, ed. F. C. Boogerd, F. J. Bruggeman, J.-H. Hofmeyr, and H. V. Westerhoff. New York: Elsevier.
- . In press. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Mahwah, N.J.: Erlbaum.
- Bechtel, W., and A. Abrahamson. 2002. *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*, 2nd edition. Oxford: Blackwell.
- . 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421–441.
- Bechtel, W., and A. Hamilton. In press. "Reduction, Integration, and the Unity of Science: Natural, Behavioral, and Social Sciences and the Humanities." In *Philosophy of Science: Focal Issues*, ed. T. Kuipers. New York: Elsevier.
- Bechtel, W., and R. C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, N.J.: Princeton University Press.
- Beer, R. D. 1995. "A Dynamical Systems Perspective on Agent-Environment Interaction." *Artificial Intelligence* 72: 173–215.
- Bernard, C. 1848. "De l'Origine du Sucre dans l'Économie Animale." *Archives Générales de Médecine* 18: 303–319.
- . 1865. *An Introduction to the Study of experimental Medicine*. New York: Dover.
- Bichat, X. 1805. *Recherches Physiologiques sur la Vie et la Mort*, 3rd edition. 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.
- Buchner, E. 1897. "Alkoholische Gärung ohne Hefezellen (Vorläufige Mittheilung)." *Berichte der deutschen chemischen Gesellschaft* 30: 117–124.
- Cannon, W. B. 1929. "Organization of Physiological Homeostasis." *Physiological Reviews* 9: 399–431.
- Christiansen, W. D., and M. H. Bickhard. 2002. "The Process Dynamics of Normative Function." *Monist* 85: 3–28.
- Craver, C., and L. Darden. 2001. "Discovering Mechanisms in Neurobiology: The Case of Spatial Memory." In *Theory and Method in Neuroscience*, ed. P. K. Machamer, R. Grush, and P. McLaughlin. Pittsburgh, Pa.: University of Pittsburgh Press, 112–137.
- Darden, L. 1991. *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.
- . 2006. *Reasoning in Biological Discoveries*. Cambridge: Cambridge University Press.
- Darden, L., and C. Craver. 2002. "Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis." *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 33: 1–28.

- Dreyfus, H. L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, Mass.: MIT Press.
- Felleman, D. J., and D. C. van Essen. 1991. "Distributed Hierarchical Processing in the Primate Cerebral Cortex." *Cerebral Cortex* 1: 1–47.
- Gánti, T. 1975. "Organization of Chemical Reactions into Dividing and Metabolizing Units: The Chemotons." *BioSystems* 7: 15–21.
- . 2003. *The principles of life*. New York: Oxford.
- Glennan, S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44: 50–71.
- . 2002. "Rethinking Mechanistic Explanation." *Philosophy of Science* 69: S342–S353.
- Harvey, W. 1628. *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus*. Francofurti: Sumptibus Guilielmi Fitzeri.
- Hempel, C. G. 1958. "The Theoretician's Dilemma." In *Minnesota studies in the philosophy of science*, Vol. 2, ed. H. Feigl, M. Scriven, and G. Maxwell. Minneapolis, Minn.: University of Minnesota Press, 37–98.
- . 1965. "Aspects of Scientific Explanation." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, ed. C. G. Hempel. New York: Macmillan, 331–496.
- Hopfield, J. J. 1982. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities." *Proceedings of the National Academy of Sciences* 79: 2554–2558.
- Lavoisier, A. L., and P. S. LaPlace. 1780. "Mémoire sur la Chaleur." *Mémoires de l'Académie Royale des Sciences*. 35–408.
- Machamer, P., L. Darden, and C. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67: 1–25.
- McCulloch, W. S., and W. H. Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics* 7: 115–133.
- Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63, 81–97.
- Miller, G. A., and J. A. Selfridge. 1950. "Verbal context and the Recall of Meaningful Material." *American Journal of Psychology* 63: 176–185.
- Minsky, M., editor. 1968. *Semantic Information Processing*. Cambridge, Mass.: MIT Press.
- Nagel, E. 1961. *The Structure of Science*. New York: Harcourt, Brace.
- Newell, A., and H. A. Simon. 1956. "The Logic Theory Machine." *IRE Transactions on Information Theory* 3: 61–79.
- . 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Pasteur, L. 1857. "Mémoire sur la Fermentation Appelée Lactique." *Comptes Rendus de L'académie des Sciences* 45: 913–916.
- . 1858. "Mémoire sur la fermentation appelée lactique." *Annales de Chimie*, 3e Ser. 52: 404–418.
- . 1860. "Mémoire sur la Fermentation Alcooolique." *Annales de Chimie*, 3e Ser. 58: 323–426.

- Pflüger, E. 1875. "Beiträge zur Lehre von der Respiration: I. Ueber die Physiologische Verbrennung in den Lebendigen Organismen." *Pflüger's Archiv für die Gesamte Physiologie des Menschen und der Thiere* 10: 251–367.
- Pitts, W. H., and W. S. McCulloch. 1947. "How We Know Universals: The Perception of Auditory and Visual Forms." *Bulletin of Mathematical Biophysics* 9: 127–147.
- Raffone, A., and C. van Leeuwen. 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.
- Rosenblatt, F. 1962. *Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms*. Washington: Spartan Books.
- Rosenblueth, A., N. Wiener, and J. Bigelow. 1943. "Behavior, Purpose, and Teleology." *Philosophy of Science* 10: 18–24.
- Ruiz-Mirazo, K., and A. Moreno. 2004. "Basic autonomy as a Fundamental Step in the Synthesis of Life." *Artificial Life* 10: 235–259.
- Ruiz-Mirazo, K., J. Peretó, and A. Moreno. 2004. "A Universal Definition of Life: Autonomy and Open-Ended Evolution." *Origins of Life and Evolution of the Biosphere* 34: 323–346.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Internal Representations by Error Propagation." In *Parallel distributed processing: Explorations in the Microstructure of Cognition. Vol. 1. Foundations*, ed. D. E. Rumelhart and J. L. McClelland. Cambridge, Mass.: MIT Press.
- Schacter, D. L., and E. Tulving. 1994. "What are the Memory Systems of 1994?" In *Memory Systems 1994*, ed. D. L. Schacter and E. Tulving. Cambridge, Mass.: MIT Press. 1–38.
- Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27: 379–423, 623–656.
- Skinner, B. F. 1971. *Beyond Freedom and Dignity*. New York: Knopf.
- Sperling, G. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs* 74 (498): 1–29.
- Tulving, E. 1999. "Study of Memory: Processes and Systems." In *Memory: Systems, Process, or Function*, ed. J. K. Foster and M. Jelicic. Oxford: Oxford University Press. 11–30.
- Tulving, E., S. Kapur, F. I. Craik, M. Moscovitch, and S. Houle. 1994. "Hemispheric Encoding/Retrieval Asymmetry in Episodic Memory: Positron Emission Tomography Findings." *Proceedings of the National Academy of Sciences (USA)* 91: 2016–2020.
- Turing, A. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society, Second Series*. 42: 230–265.
- van Essen, D. C., and J. L. Gallant. 1994. "Neural Mechanisms of Form and Motion Processing in the Primate Visual System." *Neuron* 13: 1–10.
- van Leeuwen, C., M. Steyvers, and M. Nooter. 1997. "Stability and Intermittency in Large-Scale Coupled Oscillator Models for Perceptual Segmentation." *Journal of Mathematical Psychology* 41: 319–344.
- Varela, F. J. 1979. *Principles of Biological Autonomy*. New York: Elsevier.
- Wiener, N. 1948. *Cybernetics: Or, Control and Communication in the Animal Machine*. New York: Wiley.