Forthcoming in *Topics in Cognitive Science*

Constructing a Philosophy of Science of Cognitive Science

William Bechtel
Department of Philosophy and Interdisciplinary Programs
in Cognitive Science and Science Studies
University of California, San Diego

Abstract

Although philosophy has often been an outlier in cognitive science to date, this paper describes two projects in naturalistic philosophy of mind and one in naturalistic philosophy of science that have been pursued during the past 30 years and that can make theoretical and methodological contributions to cognitive science. First, stances on the mind-body problem (identity theory, functionalism, and heuristic identity theory) are relevant to cognitive science as it negotiates its relation to neuroscience and cognitive neuroscience. Second, analyses of mental representation address both their vehicle and their content; new approaches to characterizing how representations have content are particularly relevant to understanding the relation of cognitive agents to their environments. Third, the recently formulated accounts of mechanistic explanation in philosophy of science both provide perspective on the explanatory project of cognitive science and may offer normative guidance to cognitive science (e.g., by providing perspective on how multiple disciplinary perspectives can be integrated in understanding a given mechanism).

1. Introduction

Most philosophers who have engaged in cognitive science have their roots in philosophy of mind (for discussion, see Bechtel, in press-c). A smaller number have their roots in philosophy of science; I place myself in this group, along with such philosophers as Paul Thagard and Nancy Nersessian. In this paper I will first discuss some of the main goals and endeavors of philosophy of science and then explore how it can offer an enhanced appreciation of explanations in cognitive science. Of particular interest is a new mechanistic philosophy of science that provides different answers to long-standing questions. In applying this perspective to cognitive science I address the role representations can play in mechanistic explanations, the challenge of identifying cognitive operations at the appropriate level, and a new framework for understanding interlevel relations, including a reinterpretation of reduction. Because normative prescriptions to cognitive science emerge in this process, I conclude by considering how they are warranted.

2. What is philosophy of science?

Philosophy of science traditionally has addressed itself to a variety of questions about the nature of science such as:

- What is an explanation?
- Should theoretical claims offered in explanations be evaluated for truth, or only for their utility in making true predictions?
- How does evidence confirm or falsify a proposed explanation?
- What is the relation between the explanations offered by different sciences?
- How do scientific claims differ from other knowledge claims?

Philosophers of science have employed a variety of different strategies in addressing these questions. Some have thought that philosophy possessed independent tools that could be used to advance a priori claims of what science ought to be like, whereas others have argued that it is limited to a more naturalistic endeavor in which the goal is to characterize actual science. Both a priori[1] and naturalistic approaches develop normative claims about science, but the norms of the a priori approach are put forward as apodictic (necessary) whereas the normative claims of the naturalistic approach are based on the history and claims of actual sciences (contingent).

Attempts to advance a priori accounts of science have taken a number of different forms. Most notably, logical positivism emerged in the early decades of the 20th century from a convergence of philosophers and philosophically interested scientists from Vienna, Berlin, and Prague. (For a comprehensive exposition and examination of logical positivism, see Suppe, 1974.) Starting with a commitment to empiricism—the idea that all knowledge is rooted in sensory experience—their innovation was a rigorous use of logic to link claims based on experience to explanatory hypotheses and theories. They viewed the resulting account of science as normative—that is, as articulating the standards to which good science ought to conform. Scientists might not present their claims in this way, but Carnap (1928) maintained that philosophers could rationally reconstruct the underlying logic. More recently Karl Popper (1965) criticized the positivists' execution of their project (arguing that no amount of positive evidence could ever confirm a theoretical claim), but shared their normative aspirations. He famously argued that scientists ought to advance plausible but falsifiable conjectures and focus their efforts on falsifying them.

The naturalistic alternative to an a priori view of science has its origins in a well-known book by Thomas Kuhn (1970). He argued that scientists in each mature area of science were guided by a theoretical and methodological framework he called a *paradigm*, which could neither be confirmed nor disconfirmed. The paradigm served as a basis for solving problems and when its successes became less frequent and failures accumulated, a scientific revolution might supplant it with a more promising paradigm. Kuhn also treated his account as a normative portrayal of what a mature science should be like. (He maintained, for example, that the behavioral and social sciences were immature sciences since they had yet to reach the stage in which a single paradigm would guide a particular area of research.) The normativity stemmed from his taking the physical sciences as mature and laying out the path that other sciences would need to follow.

Philosophers of science adopting a naturalistic perspective often present themselves as investigating the domain of science in the manner in which scientists investigate phenomena in their own domains of inquiry. In fact, though, philosophers' efforts generally have been limited to observational techniques. When philosophers have taken past science as their focus, what they have observed typically are the products of science, such as journal articles and textbook chapters. Occasionally they have followed historians of science into archives to identify less public records suggestive of how scientists pursued their projects, or have adopted their techniques of oral history with living scientists. When philosophers work in contemporary science they often supplement these approaches with formal or informal observation in scientific laboratories and by interacting with scientists in other ways. To understand debates in biological

[1] Quine (1969) labeled what I am calling the a priori approach *first philosophy*—the idea that philosophy could proceed prior to science in setting out what science ought to be like.

taxonomy, for example, David Hull (1988) became an active participant, an editor of *Systematic Zoology*, and President of the Society for Systematic Biology.

Naturalistic philosophers of science are not alone in making scientists and science itself the object of inquiry. In what is known as *science studies*, their work is brought into conjunction with that from other fields, especially history of science and sociology of science but also other cognitive science disciplines. The distinctive contribution of naturalistic philosophers of science to this broader endeavor is that they continue to focus on traditional epistemological questions about science, such as those listed at the beginning of this section, often seeking answers that are normative. In contrast, sociologists and increasingly historians have emphasized social factors—both those internal to the operation of a science and those involving the social and political settings of particular scientific inquiries—and have been disinclined to render normative verdicts. From the cognitive sciences, a few influential psychologists have shared philosophers' interest in the epistemic and cognitive features of science and deployed their own powerful tools—experiments on scientific reasoning (e.g., Tweney, Doherty, & Mynatt, 1981) or systematic observation of the scientific processes in actual laboratories (e.g., Dunbar, 1995). Others (e.g., Langley, Simon, Bradshaw, & Zytkow, 1987) have used computational modeling to explore the processes involved in scientific discovery.

Insofar as naturalistic philosophy of science repudiates the a priori tools of its predecessors, where does it acquire the conceptual resources to analyze science? Quine (1969), in advocating a naturalistic epistemology, proposed drawing upon the resources of (behaviorist) psychology for the tools to understand how people arrived at statements about the world from sensory experience. For him the reliance on psychology did not mean abandoning the project of epistemology (or philosophy of science) or the distinction between philosophy and science. Philosophers would still address questions that would not be addressed by psychologists—for example, the justification of people's statements—but would now turn their inquiry to questions of how these statements were produced and justified by actual people. Relaxing Quine's behaviorist principles, naturalistic philosophers of science can turn to cognitive science for new resources for addressing naturalized versions of the questions identified above. My own work, for example, has drawn importantly from Herbert Simon's (1977, 1980) conception of problem spaces and heuristic search, nearly decomposable systems, and hierarchical organization. Thagard (1988) drew inspiration from collaboration with computer scientist John Holland and psychologists Richard Nisbett and Keith Holyoak (Holland, Holyoak, Nisbett, & Thagard, 1986) in developing computational models of scientific reasoning. Nersessian (2008) has drawn upon research on distributed cognition by Edwin Hutchins (1995) among others in her work on understanding the construction of concepts in the development of new models in scientific inquiry.

In addition to the emergence of a naturalistic tradition, a major change in philosophy of science in the last decades of the 20th century is that instead of treating science as a monolith and advancing comprehensive accounts that applied to all sciences, philosophers of science began to specialize. Most notably, philosophy of biology arose as an enterprise distinct from philosophy of physics, and individual philosophers of biology or physics increasingly focused on specific subfields (e.g., evolutionary biology or quantum mechanics). As well, fields such as chemistry,

economics, psychology, and cognitive science began to draw attention.[2] In their initial forays into different sciences, philosophers tended to expect that their existing tools would generalize to the new science. The hope was that one could learn something new by deploying those tools on the new science.

Philosophers' initial confrontations with different sciences often generate considerable tension as ideas that seemed to have worked reasonably well elsewhere fail and new perspectives have to be crafted. For example, one prominent idea in the positivist's conception of science is that theories can be viewed as axiomatizable systems—as in Euclid's geometry, one could identify a set of postulates from which the axioms of the theory could be derived. Mary Williams (1970) attempted to provide such an axiomatization of evolutionary biology, but the result was viewed by other philosophers as not properly capturing the nature of evolutionary theorizing. Instead, its failure inspired the development of others ways of characterizing evolutionary theory (Hull, 1974; Lloyd, 1994). Likewise, when Richardson and I set out to analyze reductionistic research in the life sciences, we found that existing philosophical accounts in terms of laws (Nagel, 1961) failed to fit, prompting us to develop an alternative account in which we identified decomposition and localization as key strategies employed by scientists (Bechtel & Richardson, 1993). This developed into the broader project of helping to build a new mechanistic philosophy of science, to which I now turn.

3. From nomological to mechanistic explanations

One of the major legacies of the attempt by the positivists to use logic to articulate the structure of science is the deductive-nomological model of explanation (*nomos* = law; see Hempel, 1965, for an exposition). On this view, laws of nature are central to explanation. Individual events are explained by deriving statements describing them from laws plus initial conditions; thus, laws serve as the explanatory connection between initial conditions and ensuing events. This approach was attractive since it seemed to work reasonably well for some frequently invoked phenomena in physics. Students learn to explain the behavior of a pendulum, for example, by applying the mathematical law relating its period to both its length and its acceleration due to gravity. Given a problem stating the length of a given pendulum as initial conditions, they apply the law to obtain its period. However, the deductive-nomological model runs into problems in disciplines that possess few laws but many purported explanations—for instance, biology (as noted above), psychology, and other cognitive sciences. Cummins (2000) points out that "in psychology such laws as there are are almost always conceived of, and even called, effects" and argues that an effect does not explain, but rather describes the phenomenon in need of explanation. Ebbinghaus' spacing effect, for example, does not explain why learners exhibit better recall when their study time is spaced out rather than massed into one session; it simply characterizes the phenomenon.

[2] As philosophers of science engage particular sciences, they often are led not only to address philosophical questions about the science, but also to engage in the theoretical issues arising in the science. As philosophers of physics have specialized in specific domains of physics, they have addressed many of the same problems as theoretical physicists (see, for example, Malament 1977; 1987, on the structure of space-time). Likewise, philosophers of biology have been central contributors to debates over the units on which natural selection can occur, sometimes in collaboration with evolutionary biologists (e.g., Lloyd & Gould, 1993; Sober & Wilson, 1998). One context in which this has happened in cognitive science has been in the debates over whether connectionism can account for the putative systematicity of thinking (Fodor & Pylyshyn, 1988; Bechtel & Abrahamsen, 2002, Chapter 6).

Instead of invoking laws in the manner of the deductive-nomological model of explanation, biologists and psychologists typically seek to identify and describe the mechanism responsible for a phenomenon. This is not a new strategy; it figured centrally in the scientific revolution and was championed by Descartes, who endorsed mechanistic explanation for all phenomena except those unique to the human mind. (Phenomena exhibited by animals as well as humans were counted among those to be explained mechanistically.) For Descartes, a mechanistic explanation could only appeal to the shape and motion of the "corpuscles" out of which macroscopic objects were composed and direct impacts of one corpuscle on another; he famously attempted to explain magnetism as resulting from the motion of screw-shape of the particles of the magnet that pulled the object inwards as the particles moved. (In contrast, Newton advanced laws relating objects and allowed action at a distance in characterizing the interaction of physical objects, advancing no hypotheses about the mechanism involved.) In what came to be identified as the life sciences, mechanistic explanation became increasingly commonplace in the 18th and 19th centuries and ubiquitous by the 20th century. As it did so, it also expanded the range of properties of objects (e.g., to include chemical reactivity) that could be appealed to in mechanistic explanations.

Before Richardson and I made mechanism central to our account of explanation, it had drawn only limited philosophical attention (e.g., Salmon, 1984; Wimsatt, 1976). But since then a number of philosophers, drawing upon a variety of specific examples of explanation in cell and molecular biology and neuroscience have advanced more precise accounts of mechanistic explanation. Although the terminology varies somewhat across authors, the key tasks in offering a mechanistic explanation are the identification of the relevant parts of the mechanism, the determination of the operations they perform, and the provision of an account of how the parts and operations are organized such that, under specific contextual conditions, the mechanism realizes the phenomenon of interest (for representative discussions of mechanistic explanation, see Bechtel & Abrahamsen, 2005; Bechtel, 2008; Craver, 2007; Darden, 2006; Machamer, Darden, & Craver, 2000; Thagard, 2006).

Focusing on mechanisms rather than laws is not a simple substitution but rather transforms discussion of many basic issues in philosophy of science. First, explanations within the deductive-nomological (D-N) tradition relied on linguistic and mathematical representations, both for stating laws and initial conditions and for the derivations from them that yielded statements of phenomena. But mechanisms are often represented diagrammatically or even modeled physically. Second, reasoning about a mechanism often takes the form of simulating its operation (physically, mentally, or computationally), not drawing logical inferences. Third, D-N explanations are inherently general insofar as laws are universally quantified statements. But accounts of mechanisms are often developed by studying specific instances (e.g., using a model organism), and generalizing such accounts to other instances always involves identifying differences as well as similarities. Biologists often appeal to evolutionary conservation of mechanisms, but such conservation is only partial and researchers must also explore differences between instances (Bechtel, in press-b). Fourth, while philosophers were poorly equipped to investigate the discovery of laws, they were able to identify a variety of research strategies for discovering mechanisms, making scientific discovery once again a prominent topic in philosophy of science (Darden, 2006). Given these and other substantial differences (see Bechtel &

Abrahamsen, 2005, for further discussion), mechanistic explanation is increasingly recognized as radically different than nomological explanation.

4. What is distinctive about information processing mechanisms?

Given that talk of mechanisms is as ubiquitous in the cognitive sciences as it is in the biological sciences, it is natural to explore how well the accounts of mechanism developed for biological sciences apply in the cognitive sciences. When cognitive scientists propose mechanisms of perception, memory, problem solving, and so forth, they do seem to be presuming that the organized operation of parts of the mind-brain is responsible for these phenomena. The notion of mechanism that has been so fruitful in biology extends quite well to cognitive science up to this point—but a fundamental difference must then be confronted. Basic biological mechanisms move and transform physical substances so as to sustain life. (An example from cell biology is a chemiosmotic mechanism that moves hydrogen ions across a membrane so they are available to fuel ATP synthesis; an example from molecular biology is an RNA transcription mechanism that synthesizes proteins from amino acids.) Mechanisms in cognitive science, in contrast, are proposed to explain *cognitive* activities such as memory retrieval or problem solving by performing operations on *representations* that carry information about objects, events, and circumstances currently or previously encountered. Operating on representations is different than merely moving or transforming physical substances, in that representations serve an informational function: they relate a vehicle (the form of the representation) to a content (what it is about). As I use the term *information processing mechanism*, it is any system with parts that perform this function (even, as we will soon see, in a manner that is not cognitive or mental).

Cognitive scientists have given considerable attention to the question of how best to characterize the vehicle of representation. Depending upon their theoretical preferences in accounting for our ability to process information, the vehicle may be a language-like expression in mentalese, a pattern of activation in an artificial neural network, a brain state, and so forth. Cognitive scientists have directed much less attention to questions of the nature of content and how the vehicle relates to it. In the simplest construal, which is adequate to my purposes here, the content of a representation is the object, event, or circumstance that the vehicle represents. (Note, however, that more complex construals have been exhaustively debated by philosophers to deal with such problems as misrepresentation.[3])

My contention is that if we are to appreciate what is distinctive about information processing mechanisms, we need to account for how representations carry information about contents. I will discuss this claim in more depth and offer a framework inspired by control theory that provides a distinctive understanding the content-vehicle relation. In the process I will develop a deflationary approach according to which many systems other than minds traffic in representations. First,

[3] Since this problem was emphasized by Brentano (1874) in his account of intentionality, different approaches have been taken to misrepresentation. One strategy is to treat such contents as existing entirely in the mind; this is problematic since it fails to account for how representations enable us to know about real objects (Richardson, 1981). Another strategy is to differentiate the content of a representation from its target (Cummins, 1996), such that the mode of presentation of an object or event (content) does not necessarily correspond to the actual object or event (target). In this paper I will simply treat the content as the object, event, or circumstance that makes, or would make, the representation true—Sherlock Holmes is the person who would have made Conan Doyle's narratives true.

though, I will briefly place theories of human information processing in context and show how they came to focus nearly exclusively on vehicles, neglecting contents.

One important influence in formulating the idea of an information processing mechanism originated with philosophical work on logic. In the 19th century Boole characterized logic as embodying the rules of thought. As Frege and Russell developed symbolic logic in the early 20th century, they construed these rules of inference formally—that is, the rules applied to the forms of symbols without regard to their meaning. (In *p* and *q*, therefore *p*, any specific expressions can be inserted, as long as the first and third have the same form.) A similar conception of thinking is evident in early work on the theory of computation by Turing (1936) and Post (1936). They took as their model for computing devices human beings, referred to as "computers," who were hired to perform mathematical computations by applying memorized rules to written symbols. Turing and Post designed and offered proofs of the capacities of different classes of automata—abstract machines that would operate analogously to the human computers (and in principle could be implemented as electronic devices).

Perhaps the most potent influence on early information processing models, though, was Chomsky's generative grammar—most fundamentally, his idea that syntactic knowledge is best captured in a set of rules for an automaton. In principle, in systematically and exhaustively applying the rules, the automaton would generate the infinite set of well-formed sentences that constitute the language. Chomsky (1956), arguing that a finite state automaton was inadequate, instead proposed one that implemented rewrite rules to obtain representations of phrase structure and transformational rules to alter those structures. It is noteworthy that by concentrating on syntax as an autonomous module of linguistic knowledge, Chomsky focused on the forms of linguistic representations, not their content. Such abstraction from content, consistent with the foregoing work in logic and computation theory, had important consequences in the application of these ideas to cognitive science.

Chomsky's work in linguistics provided a model for psychologists attempting to overcome the strictures of behaviorism. Chomsky invited such influence by insisting that transformational grammar described linguistic competence—speakers' mentally represented knowledge of language—and hence was part of the project of psychology. He emboldened psychologists to posit mental representations, and offered guidance as to how they might look. Some went elsewhere to borrow particular representational formats (e.g., to symbolic logic), but others used Chomsky's innovations to transform psycholinguistics. Notably, Miller (1962) pioneered the use of reaction time patterns as behavioral evidence that the transformations posited in the grammar were actually performed when people processed particular sentences (a claim substantially revised as the relation between subsequent grammars and experiments grew more complex).

Chomskian psycholinguistics was a robust precursor of cognitive science, but so was the broader *human information processing* approach that emerged during the same period using similar research strategies. Sternberg (1966), for example, interpreted the pattern of reaction times in a mental search task as supporting exhaustive rather than self-terminating search for a target. Neisser's (1967) *Cognitive Psychology* provided one of the first systematic accounts of the new approach, and his own account of how he arrived at it illustrates the variety of influences that converged in characterizing the mind as an information processing machine:

> By 1964, it had come together in my head. In principle, I thought, one could follow the information inward from its first encounter with the sense organ all the way to its storage and eventual reconstruction in memory. The early stages of processing were necessarily holistic (an idea I borrowed from Gestalt psychology) and the later ones were based on repeated recoding (an idea borrowed, even more obviously, from George Miller). But the processing sequence was by no means fixed; at every point there was room for choice, strategy, executive routines, individual constructive activity. Noam Chomsky's linguistic arguments had shown that an activity could be rule governed and yet indefinitely free and creative. People were not much like computers (I had already sketched out some of the differences in a 1963 *Science* paper), but nevertheless the computer had made a crucial contribution to psychology: It had given us a new definition of our subject matter, a new set of metaphors, and a new assurance (Neisser, 1988, p. 86).

Once again, it is noteworthy that the focus is on procedures for manipulating representations as formal entities, not the content of those representations or any way in which the operations were sensitive to such content.

Computational modeling of cognitive function developed in the hands of Newell and Simon (1972) in a manner that parallels how Turing and Post developed their ideas of computation. By having participants talk aloud as they solved problems such as the Tower of Hanoi, Newell and Simon hoped to reveal the operations humans performed. They then implemented these in a computer program. The production system architecture they developed employed formal symbolic representations and rules for their manipulation. In contrast, a competing approach to conceptualizing an information processing device employed neuron-like units organized into networks by weighted connections. McCulloch and Pitts (1943) showed how an artificial neural network could implement logic functions, further evidence of the idea that logical operations characterized the functioning of the mind. Later neural network theorists broadened the conception of the behavior of neural networks to pattern recognition and association (Rosenblatt's 1962 perceptrons) and also to pattern transformation in service of a variety of sensorimotor and cognitive tasks (Feldman & Ballard's 1982 connectionist models; Rumelhart & McClelland's 1986 PDP models). By the 1980s, the patterns across layers of units were regarded as representations. A key aspect of PDP (parallel distributed processing) models is that representations are distributed over many units, distinguished from each other by the pattern of activation values across those units. A given unit therefore can contribute to many different representations. But the explanatory focus remained on the connection weights that networks developed during training, as these explained the transformations of the representations.

5. Restoring representational content to information processing

The information processing models that play an explanatory role in cognitive science, including neural networks, are regarded by their designers as a distinctive type of mechanism insofar as they manipulate representations. In working out this conception of representation, however, one immediately confronts a problem: processing within the mechanism is necessarily limited to the formal (sometimes called syntactic) aspect of representation—i.e., to vehicles, not content. This applies equally to classical information processing and neural network models. As critics have noted, it is the scientists who think of the operations in their models as working on content. In fact, the content has no effect on any particular operation or on the overall functioning of the mechanism. To better appreciate this, consider the possibility that a different scientist would

arrive at a mechanistic model of the same design for a task involving content that is structurally identical but otherwise quite different from the content of the original task. Within the mechanism the same parts would carry out the same operations over time, but each scientist would interpret this as modeling a different task in a different domain and, most salient here, attribute different content to the representations within the model.

That information processing models (as distinct from the mental activity they are intended to model) fail to account for content is compellingly illustrated in Searle's (1980) Chinese Room thought experiment. He directed his argument particularly against what he termed "strong AI"—the claim that a computer running an artificial intelligence (AI) program qualifies as an intelligent cognitive system. To assess the plausibility of this claim, he imagined himself playing the role of a computer programmed to execute a version of an AI program described in Schank and Abelson (1977). The program uses scripts to facilitate answering questions about stories—but in Searle's version everything except the computer program itself was in Chinese rather than English, to sharpen his intuitions. Imaginary Searle thus sits isolated in a room with two batches of Chinese writing that, could he read Chinese, he would realize were a story and the relevant script. Someone sends in a third, smaller batch of Chinese characters (in fact, a question about the story) and expects him to respond. Searle's only resource is sets of rules in English for correlating the strings (i.e., the program that he as the computer is to run). So, when the question batch comes in, he finds and applies the rules that enable him to send back out an appropriate batch of Chinese characters, which constitutes an answer. Eventually he does this so well that a Chinese observer would believe that in the room is someone who can understand and use Chinese. But Searle fails to understand: for him (and for the computer he replaced) every batch of Chinese writing consists of meaningless characters—a vehicle that conveys no content and hence is not actually a representation.[4]

Two possibilities present themselves at this point. One is to claim that information processing mechanisms are, in fact, no different from other biological mechanisms—they operate by transforming internal states much as a cell transforms glucose into carbon dioxide and water. Assigning content to representations is, at best, an interpretive activity of cognitive scientists in thinking and talking about the systems they design and of ordinary people in characterizing their own and other people's minds. There is, on this view, no intrinsic content to mental states (Dennett, 1971). The second possibility is to treat content as an essential aspect of representations (mental and otherwise) and give serious attention to the challenge of incorporating content into accounts of information processing. This second possibility is the one I will pursue in the remainder of this section.

Looking first to philosophy of mind for what it can contribute, we find three major approaches to content, none of which alone is adequate for making content integral to information processing

[4] Searle's own strategy for explaining how he knows the content of his thoughts is to appeal to the material constitution of minds which, he claims, possess the needed causal powers. Searle provides no account of what it is about the mind's material constitution that accounts for these causal powers. If he did he would succumb to his own challenge, since such an account could be employed to generate the functional account of the mind that he eschews. Let's assume that the mind's material constitution (the brain) is the subject of neuroscience. Neural investigations tend to proceed functionally: identifying what operation each part of the system performs and how they work together to realize the phenomenon of interest. If Searle performed these operations, he might appear to carry on a conversation in Chinese without knowing Chinese.

accounts. One approach (advanced by Dretske, 1981) appeals to how representations are caused to appear within the system by their referents (seeing a panda causes the representation PANDA to be produced in one's mind). Fodor's (1987) alternative appeals to the asymmetry in the representation's causal relation to its true content versus any other content it might misrepresent. The third approach (set out by Millikan, 1984) appeals to the history of selection that favored certain representations based on how they were used (consumed) within the organism: PANDA represents a panda because it had facilitated the organism's interactions with pandas in the past. The deficiency of these proposals for the current task is that all treat the content relation as external to the representation and the operations on it. On all three accounts, information processing operates on the representational vehicle, and its content does not seem to be playing any role.

For insight into how to develop an account of information processing in which content enters integrally into the analysis, we might more profitably turn to control theory. This is the subfield of mathematics and engineering devoted to understanding how to manipulate the parameters affecting the behavior of a system so to achieve a desired outcome. A controller is described in terms of how it interacts with the mechanism it controls (commonly referred to as the plant, e.g., a furnace). More elaborate controllers may employ a model of the plant (Grush, 1997, 2004). An effective controller must be causally connected with the plant it controls so as to receive frequently updated information about the plant's state and respond with appropriate actions.

The relevant system for analysis thus includes the plant as well as the controller. One can appreciate this by considering the controller that James Watt created for the steam engine, known as the Watt governor (left side of Figure 1).[5] The engine produced steam that flowed through a valve to provide power to machines via a crankshaft (not shown) to which a flywheel was attached. The problem was to ensure that the flow of steam would increase or decrease as needed to quickly correct any changes in the speed at which the flywheel turned, due to fluctuations in the amount of power needed by the machines. Watt devised a control mechanism in which a spindle with arms is attached to the flywheel and centrifugal force would extend the arms further the faster the flywheel turned. He then employed a linkage mechanism between the spindle arms and the steam valve. The right side of Figure 1 shows how the angle of the spindle arms carries information about the speed of the flywheel that is used to determine how far the steam valve opens or closes. Thinking about the governor in control terms does not change the fact that it is a representational vehicle in the mechanism that is operated on causally and causally affects the consumer. What it does do is widen the scope of the system that must be considered in the explanation: to understand it as a controller, one must understand its relation to the other components with which it was engineered to work. The diagram illustrates how a representational system is realized in the control system architecture (controller and plant). Note that the spindle arms are both controller and vehicle. Other components of the representational system are in the plant: the speed at which the flywheel operates (content) and the extent to which the valve (the consumer) is opened.

[5] Ironically, the Watt governor was introduced into discussions of cognitive science by Timothy van Gelder (1995) as showing how, within a dynamical systems perspective, one could explain cognitive activities without appealing to representations. In Bechtel (1998) I argued that the Watt governor in fact employed representations and that we could only understand how it controlled the steam engine by considering these representations.

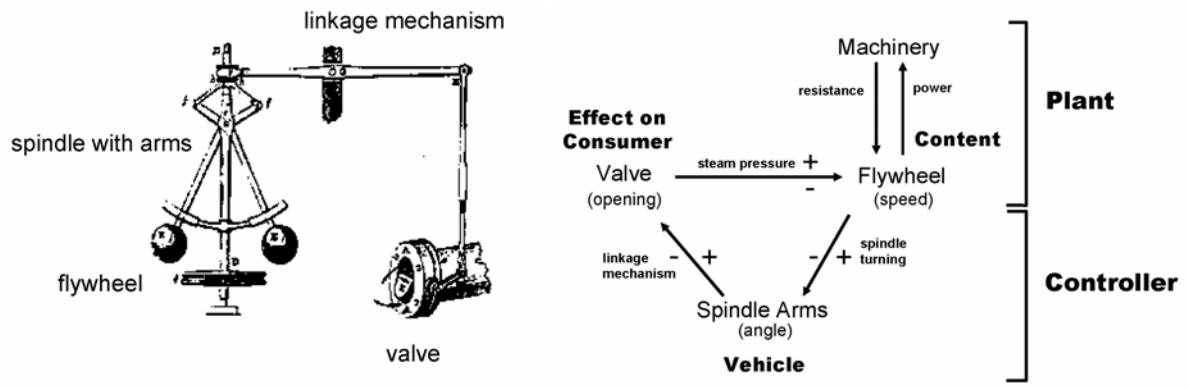

Figure 1. On the left, Watt's governor for the steam engine (adapted from Farley, 1827); on the right, a schematic diagram showing how a representational system (vehicle linked to content and consumer) is realized in the control system architecture.

Having described how certain physical systems can be understood as representational—those that incorporate a control system—we turn to the question of what kinds of biological systems might be understood using the same framework. As systems that maintain themselves far from thermodynamic equilibrium, living organisms continually recruit matter and energy from the environment, utilize these to build and repair their component parts, and dispel waste (Ruiz-Mirazo & Moreno, 2004). Component mechanisms of this kind (i.e., the organism's plant) are inherently active, but the operations of some are incompatible and must be controlled. This is typically accomplished by inhibiting a mechanism when it is not required and releasing it from inhibition when it is required. For example, the mechanism for burning sugar to capture energy in ATP is inhibited when ATP levels are high, and released when they are low. A control component within a living organism thus must be coupled not only to the mechanism it regulates, but also to sensors that detect when the system needs what that component mechanism provides. Through these couplings, control systems represent relevant conditions in the organism and use that information to continuously adjust the activity of one or more mechanisms. Put another way, the controller is a physical component of the organism. Insofar as it receives information from the sensor, it serves as vehicle for that content. And insofar as it both receives and uses the information to act on other components, it serves as controller.

The account so far shows only how processes in the controller could represent conditions within the organism's body. But organisms are also dependent on coordinating with their environment if they are to maintain themselves, and this provides an opportunity for extending the control theory perspective. For example, an *Escherichia coli* (*E. coli*) bacterium navigates up sucrose gradients in its environment to procure energy. The information processing challenge it confronts is to determine the direction of the gradient using only sensory receptors that are not sufficiently dispersed in space to detect differences in chemical concentrations (Mandik, Collins, & Vereschagin, 2007). The alternative strategy it invokes it to detect changes in concentration over time. This requires a form of memory that is realized by a control system that regulates the basic plant linking sensory receptors to motor units. In the basic plant, when a chemical attractant such as sucrose (or a repellant such as citrate) binds to a surface receptor, a sequence of phosphorylation reactions ensues. The phosphorylated substrates cause the bacterium's flagella to rotate counterclockwise and so act a single unit that propels the bacterium in a relatively straight path. Otherwise, the flagella rotate clockwise, detangle, and cease to propel the

bacterium; the bacterium then tumbles freely. Control is achieved via a methylation pathway that modulates the response of the receptor by increasing the number of methyl groups on glutamate side chains on the receptor in response to high concentrations of the attractant in the immediate past. The control pathway thus modifies the operation of the chemotaxic pathway so that it responds not to absolute level of an attractant or repellant in the local environment, but to changes in the concentrations of attractants or repellants (Falke, Bass, Butler, Chervitz, & Danielson, 1997; van Duijn, Keijzer, & Franken, 2006).

As another example of the critical function of control systems that enable organisms to regulate their activities in relation to the environment, consider the cyanobacterium *Chiamydomonas*. It must switch between photosynthesis during the day and nitrogen fixation at night as it must perform both reactions but the oxygen produced in photosynthesis is inimical to the mechanism of nitrogen fixation. A cyanobacterium controls this switching even when kept in total darkness by employing an endogenous clock that employs an approximately 24-hour oscillation between a phosphorylated and unphosphorylated state of Kai proteins to control behavior. When exposed to day-night cycles, this oscillation is entrained to the light period so that states in the oscillator represent times of day even when light is not available (Kondo, 2007).

In the examples presented in the two previous paragraphs, internal states of components of the organism carry information (sometimes false) about conditions external to the organism, and this information is essential for coordinating its behavior with the flux of external conditions. They indicate that the close coupling of organisms to their environments can be accommodated in the framework I am constructing. We saw earlier that to understand information processing mechanisms as control systems entails conceptualizing them as situated within a larger system, within which they exercise control. In the Watt governor example, that system was a physical plant such as a steam engine. In extending the framework to biology, the physical plant was upgraded to one or more biological mechanisms within an inherently active organism (e.g., the mechanism for burning sugar to capture energy in ATP). And now we see that the larger system in which the controller performs its task includes the environment with which the organism engages. Moreover, in each of these contexts—from the Watt governor to the segregation of photosynthesis and nitrogen fixation—we have seen that construing information processing mechanisms as control systems also ensures that they are fully representational. That is, when the representational vehicle is a controller, it is not merely formal but rather is tightly coupled to content that is the basis for controlling the operations of a consumer within the larger system. Having come this far, can we now apply the same framework to human information processing? The representations involved are more complex and less directly coupled to an individual's immediate environment than those just discussed, making it a nontrivial project to extend the control theory perspective to cognitive science. I must limit myself here to encouraging the pursuit of this project, and noting that one promising avenue is Barsalou's (1999) proposal to ground concepts, including abstract concepts, in sensorimotor processes (see Bechtel, 2008, chapter 5 for elaboration).

6. What are the operations in cognitive mechanisms?

Identifying the operations performed by the component parts of a mechanism is a crucial part of explaining how it is able to generate the phenomenon of interest. As a science develops, it commonly compiles a catalog of parts and operations that, once identified, can be drawn upon in

constructing mechanistic explanations of ever more phenomena in its domain. Though entries may occasionally be added (or removed), the catalog at a given time provides the basis for most explanatory pursuits. Cognitive science hosts multiple catalogs at present, and distinct groups of researchers tend to coalesce around each. Philosophers of cognitive science have tended to focus on two of these catalogs: (1) discrete symbol manipulations (used in classical cognitive models) and (2) quantitative computations in networks (used in neural network models). Cognitive scientists drawing on one of these catalogs (or on those of cognitive linguistics, Bayesian inference, etc.) have achieved a measure of success. It is my contention, however, that cognitive science may still be in quest of a catalog of operations that will provide a firm foundation over time.

As I noted, Turing and Post were partly inspired in developing their account of computation by the activities human "computers" employed (applying rules to numbers written on paper). Symbolic accounts assume that the operations within humans also involve applying rules to stored symbols. Typically, though, the operations within a mechanism are different from the phenomenon produced by the mechanism. Within a neuron, for example, neurotransmitters perform such operations as diffusing across a synapse and binding to a receptor; but the neuron itself generates action potentials.

The point of organizing component parts and operations into a mechanism is to accomplish something that cannot be performed by the individual components. Hence, assuming a homunculus with the same capacities as the agent in which it is posited to reside clearly produces no explanatory gain. The recognition that it is problematic to assume that operations within a mechanism perform the same type of operations as the mechanism itself may be a major reason many find problematic Fodor's (1975) proposal of a language of thought to explain language and thought.

Neural network models avoid this problem, but by virtue of being composed of numerous abstract neurons perhaps reach too far in the opposite direction. Neurons are certainly components of brains, but they (like atoms and quarks) are at too low a level to provide the operations involved in mental activities. In the most neurally realistic networks (those not using localist encodings), the representations involved in simulations of cognitive operations are distributed patterns over many units. If these representations are important to the functioning of the mechanism, then the operations described should deal with entire distributed patterns. If it proves impossible to identify a catalog of such operations and the operations to be considered are those involved in generating action potentials in individual neurons, then the representations are epiphenomenal from a mechanistic perspective.

The situation cognitive science confronts is not unique. In the late 19th century physiological chemistry confronted a similar challenge in trying to explain physiological processes. Researchers attempting to explain fermentation, for example, either sought a set of constituent "fermentations" (a conceptualization borrowed from the very level they were trying to explain) or moved to a level of operations that was too low to yield an insightful account—the addition or removal of atoms from molecules. It was not until the discovery of biochemical groups (e.g., phosphate groups) that the relevant operations for explaining physiological processes could be identified (e.g., phosphorylation and dephosphorylation). My suggestion is that contemporary cognitive science is in a predicament similar to that of physiological chemistry in the 19th

century. New insights are required to identify operations at a level below that of whole cognitive agents and above that of individual neurons—that is, operations appropriate for characterizing the workings of cognitive mechanisms (Bechtel, 2008, chapter 3).

One limitation cognitive science has faced in identifying the operations in cognitive systems is that it emerged when there were no available tools for identifying the brain structures performing those operations, a situation that has changed with the development of neuroimaging. Although neuroimaging is most often regarded as a tool for determining which brain areas are specifically involved in a particular cognitive activity, most salient here is that this technique makes possible new heuristic strategies for identifying cognitive operations (Bechtel & Richardson, in press). For example, when two apparently different cognitive tasks are found to employ a common brain area, this may provoke researchers to figure out what component operation(s) they share. Conversely, a given task typically activates a neural circuit involving multiple brain areas, not a single area. Researchers who interpret this as the neural realization of coordinated cognitive operations can be guided by the pattern of activity in seeking not only to identify operations performed by each area, but integrate them into a system that generates a phenomenon of interest. When imaging results are combined with lesion data (including that generated by temporary functional lesions induced with transcranial magnetic stimulation), researchers can acquire additional clues to the operations. Moreover, when these brain areas can be identified in other animals, researchers can take advantage of other techniques, including single cell recording, neural stimulation, and surgical lesioning, to generate yet more clues to the operations each performs. Access to brain areas does not solve the problem of identifying operations, but it expands the resources cognitive scientists can employ (Bechtel & McCauley, 1999).

7. Rethinking reduction: Cognitive science and the brain

Even suggesting that cognitive scientists consult neuroscience for assistance can provoke strong dissent by those who view it as surrendering a distinctive autonomy for cognitive science. To some, it amounts to claiming that cognitive science should be reduced to neuroscience, which will then provide the true account of cognitive phenomena. I would suggest this is a misperception originating in a faulty conception of reduction, one for which philosophers of science are partially responsible. Inspired by the positivists' portrayal of science, philosophers have long treated reduction as the derivation of laws of the reduced science from those of a more basic science (Nagel, 1961). On this view, a law of biology would be explained by deriving it from law(s) of chemistry, and a law of psychology would be explained by deriving it from law(s) of neuroscience. Physics was viewed as residing at the foundation of this whole explanatory edifice, serving to unify all sciences. Given this conception of reduction, various philosophers argued either that psychology, and cognitive science more generally, should be reduced to neuroscience and ultimately more basic sciences (P. M. Churchland, 1989; P. S. Churchland, 1986; Bickle, 1998) or that psychology and cognitive science could not be reduced and should remain autonomous from neuroscience (Fodor, 1974). (See McCauley, 2007, for an in-depth discussion of reduction within cognitive science.)

Construing mechanisms, not laws, as the vehicles of explanation gives rise to a very different conception of reduction. I regard mechanistic explanation as inherently reductionistic in that it appeals to the parts and operations within a mechanism to explain the phenomena produced by it. Thus, to explain a physiological process such as respiration a biochemist identifies the chemical

reactions and reagents involved. This first step in understanding a mechanism involves reduction, so construed, but it is not the only step. A mechanism also must be appropriately organized and it must be situated with respect to other objects and events in its environment (which often will activate or modulate the functioning of the mechanism). Accordingly, mechanistic inquiry must not only look down to the component parts and operations of the mechanism (reduction), but also look back up at how they are organized so as to interact appropriately (recomposition) and look out at how they are situated in their environment (emplacement); see Bechtel (in press-a). The mechanism is a bridge between the level of the parts and the level at which the mechanism as a whole engages its environment.

The notion of levels often is invoked in discussions of reduction, but a plethora of conceptions of levels are involved. Many of these turn out to be problematic Analysis of reduction in mechanistic terms requires and supports a very minimal and local notion of level in which the constituent parts and operations of a mechanism are at a lower level than the mechanism itself (Craver, 2007). This conception of level does not support levels extending across nature, in that no grounds exist for determining whether the parts of two mechanisms that interact with each other are at the same or different levels. At each level there are causal interactions among operations,[6] and such interactions at several different levels may be important to the explanatory project (Bechtel, 2008; Craver & Bechtel, 2007). Hence, this conception of level is at odds with the popular idea that there is a lowest level at which all these causal processes can be characterized (Kim, 1998). The causal engagements in which the mechanism participates are made possible by what its parts are doing *and* how they are organized so as to work together, thereby enabling the mechanism as a whole to do something. An inquiry into how the mechanism realizes a particular phenomenon typically involves a small number of levels. Starting from the characterization of the phenomenon and identification of the responsible mechanism, investigators decompose the mechanism into its parts (and possibly into their parts), and also situate it in a larger system—either an ordinary environment or a higher-level mechanism in which it is itself a component part. Only when all this is provided, have researchers answered the question of how the phenomenon was brought about.

Recently cognitive science has been confronted by challenges both from those advocating refocusing attention on the brain and those calling for attention to the embodied and situated aspects of cognition. The implication of the account of mechanistic explanation I have outlined is that these ought not to be viewed as challenges to cognitive science or as exclusive alternatives; both represent constructive avenues for advancing inquiry in cognitive science. For cognitive scientists interested in episodic memory, for example, both the molecular processes involved in long-term potentiation in hippocampal neurons and the social situation of the individual acquiring the memory can provide relevant information, These do not supplant the primary interest of cognitive scientists in characterizing the cognitive processes involved in memory encoding, however. As suggested above, the targeted cognitive level resides above the level at which the responsiveness of post-synaptic cells is altered and below the level at which the cognitive agent engages with its environment.

[6] There also is a tendency in discussions of reduction to talk of causal interactions between levels. Craver and I argued that such talk is problematic. Given that we restricted causation to within levels, we advocated using a constitution relation to account for relations between levels. In this way, the phenomena of bottom-up and top-down causation can be accounted for without facing the problems raised by positing interlevel causation.

8. Developing norms from a naturalistic perspective

The naturalistic approach to philosophy of science takes as its foundation science as we find it, not an a priori account of what science ought to be. Its aim is to understand science itself as it actually functions. And yet, as the last three sections illustrate, naturalistic philosophers of science do advance norms. To reprise, I have argued for adopting a control theoretic framework so as to understand the distinctive role of informational content in information processing mechanisms; advocated ways in which cognitive science might finally identify operations adequate to its task; and advanced the claim that mechanistic research is not solely reductionistic, needing to attend also to organization of components and to the environment. It is through the prescriptive nature of these normative principles that philosophy of science can most directly influence cognitive science; hence, I conclude by reflecting on the basis for advancing such normative prescriptions.

The account of mechanistic explanation itself provides one basis for hypothetical normative claims. If a group of scientists aims to produce mechanistic explanations and the operation of a mechanism depends on its component parts and operations, its organization, and conditions in its environment, then the scientists should attend to all of these. To reject this normative injunction, cognitive scientists would need either to repudiate the objective of mechanistic explanation or to reject the characterization of mechanistic explanation that has arisen from scientific practice. Like scientific claims, naturalistic claims about science might turn out to be incorrect, but this must be established by showing where the account falls short.

The characterization of mechanistic explanation also provided part of the basis for doubting that cognitive science has yet found the appropriate level of operations for mechanistic explanations of cognition. Such operations need to be at a lower level than the mechanism performing the task of interest but not so low as to miss the appropriate causal interactions. This doubt also was motivated by a comparison with another science—biochemistry—that had confronted a similar challenge. Indeed, although the search for mechanism would not necessarily follow the same path in different sciences, one can learn from an examination of other disciplines some of the possibilities and challenges for developing mechanistic explanations. Philosophers of science often are best positioned to provide such a comparative perspective and to extract insights that point to useful norms.

Finally, the tension between the espoused objective of a mechanistic explanation (here, to serve as an account of information processing) and the nature of a mechanistic account provides for a third source of normative implications. The appeal to representations as constituent parts of any information processing mechanism often is presented as a distinctive characteristic of cognitive explanations. But, as I have tried to elucidate, the standard treatment of information processing systems does not account for the content of representations. This tension can be reduced by reconceptualizing information processing in a control theory framework, which requires relating the controller to the plant being controlled and, when relevant, to the environment. In this case, the normative implication represents a suggestion as to how a tension in the project of cognitive science can be overcome by extending the understanding of the system in which information processing occurs.

Naturalistic philosophy of science takes science as its object of study. A philosophy of science of cognitive science makes cognitive science itself the object of study. I have argued, though, that it can also provide normative guidance to cognitive science and thereby contribute to it.

Acknowledgements

I thank Adele Abrahamsen, Andy Brook, Pete Mandik, and Robert McCauley for very helpful discussion and comments on an earlier draft of this paper.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science, 22*, 295-318.

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

Bechtel, W. (in press-a). The downs and ups of mechanistic research: Circadian rhythm research as an exemplar. *Erkenntnis*.

Bechtel, W. (in press-b). Generalization and discovery through conserved mechanisms: Cross species research on circadian oscillators. *Philosophy of Science*.

Bechtel, W. (in press-c). Is philosophy a cognitive science disciplines? *Topics in Cognitive Science*.

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks* (Second ed.). Oxford: Blackwell.

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

Bechtel, W., & McCauley, R. N. (1999). Heuristic identity theory (or back to the future): The mind-body problem against the background of research strategies in cognitive neuroscience. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 67-72). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Bechtel, W., & Richardson, R. C. (in press). Neuroimaging as a tool for functionally decomposing cognitive processes. In S. J. Hanson & M. Bunzl (Eds.), *Foundational issues of neuroimaging*.

Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge, MA: MIT Press.

Brentano, F. (1874). *Psychology from an empirical standpoint* (A. C. Pancurello, D. B. Terrell & L. L. McAlister, Trans.). New York: Humanities.

Carnap, R. (1928). *Der logische Aufbau der Welt*. Berlin: Weltkreis.

Chomsky, N. (1956). Three models for the description of language. *Transactions on Information Theory, 2* (3), 113-124.

Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge, MA: MIT Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified theory of mind-brain*. Cambridge, MA: MIT Press.

Craver, C. (2007). *Explaining the brain: What a science of the mind-brain could be*. New York: Oxford University Press.

Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547-563.

Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.

Cummins, R. (2000). "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117-144). Cambridge, MA: MIT Press.

Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: Cambridge University Press.

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy, 68*, 87-106.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press/Bradford Books.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *Mechanisms of insight* (pp. 365-395). Cambridge, MA: MIT press.

Falke, J. J., Bass, R. B., Butler, S. L., Chervitz, S. A., & Danielson, M. A. (1997). The two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual Review of Cell and Developmental Biology, 13* (1), 457-512.

Farley, J. (1827). *A treatise on the steam engine: Historical, practical, and descriptive*. London: Longman, Rees, Orme, Brown, and Green.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6*, 205-254.

Fodor, J. A. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese, 28*, 97-115.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology, 10*, 5-24.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27*, 377-396.

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.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning and discovery*. Cambridge, MA: MIT.

Hull, D. L. (1974). *The philosophy of biological science*. Englewood Cliffs, NJ: Prentice-Hall.

Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.

Kondo, T. (2007). A cyanobacterial circadian clock based on the Kai oscillator. *Cold Spring Harbor Symposia on Quantitative Biology, 72*, 47-55.
Kuhn, T. S. (1970). *The structure of scientific revolutions* (Second ed.). Chicago: University of Chicago Press.
Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative process*. Cambridge: MIT Press.
Lloyd, E. A. (1994). *The structure and confirmation of evolutionary theory*. Princeton, NJ: Princeton University Press.
Lloyd, E. A., & Gould, S. J. (1993). Species selection on variability. *Proceedings of the National Academy of Sciences of the United States of America, 90* (2), 595-599.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.
Malament, D. B. (1977). The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics, 18* (7), 1399-1404.
Malament, D. B. (1987). A note about closed timelike curves in Gödel space-time. *Journal of Mathematical Physics, 28* (10), 2427-2430.
Mandik, P., Collins, M., & Vereschagin, A. (2007). Evolving artificial minds and brains. In A. C. Schalley & D. Khlentzos (Eds.), *Mental States. Volume 1: Evolution, function, nature* (pp. 75-94). Amsterdam: John Benjamins.
McCauley, R. N. (2007). Reduction. In P. N. Thagard (Ed.), *Philosophy of psychology and cognitive science*. New York: Elsevier.
McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 7*, 115-133.
Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist, 17*, 748-762.
Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace.
Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
Neisser, U. (1988). Cognitive recollections. In W. Hirst (Ed.), *The making of cognitive science: Essays in honor of George A. Miller* (pp. 81-88). New York: Cambridge.
Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Popper, K. R. (1965). *Conjectures and refutations: The growth of scientific knowledge* (Second ed.). New York: Harper and Row.
Post, E. L. (1936). Finite combinatorial processes - Formulation I. *Journal of Symbolic Logic, 1*, 103-105.
Quine, W. V. O. (1969). Epistemology naturalized. In W. V. O. Quine (Ed.), *Ontological relativity and other essays*. New York: Columbia University Press.
Richardson, R. C. (1981). Internal representation: Prologue to a theory of intentionality. *Philosophical Topics, 12*, 171-211.
Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Washington: Spartan Books.
Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*, 235-259.

Rumelhart, D. E., & McClelland, J. L. (1986). *Explorations in the microstructure of cognition. Volume 1. Foundations*. Cambridge, MA: Bradford Books, MIT Press.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417-424.

Simon, H. A. (1977). *Models of discovery and other topics in the methods of science*. Dordrecht: Reidel.

Simon, H. A. (1980). *The sciences of the artificial. Second Edition.* Cambridge, MA: MIT Press.

Sober, E., & Wilson, D. S. (1998). *Onto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.

Suppe, F. (1974). *The structure of scientific theories*. Urbana: University of Illinois Press.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press/Bradford Books.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, second series, 42*, 230-265.

Tweney, R. D., Doherty, M., E., & Mynatt, C. R. (Eds.). (1981). *On scientific thinking*. New York: Columbia University Press.

van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination. *Adaptive Behavior, 14* (2), 157-170.

van Gelder, T. (1995). What might cognition be, if not computation. *The Journal of Philosophy, 92*, 345-381.

Williams, M. B. (1970). Deducing the consequences of evolution: A mathematical model. *Journal of Theoretical Biology, 29*, 343-385.

Wimsatt, W. C. (1976). Reductive explanation: A functional account. In R. S. Cohen, C. A. Hooker, A. C. Michalos & J. van Evra (Eds.), *PSA-1974* (pp. 671-710). Dordrecht: Reidel.