JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2005, **84,** 313–325 NUMBER 3 (NOVEMBER)

THE CHALLENGE OF CHARACTERIZING OPERATIONS IN THE MECHANISMS UNDERLYING BEHAVIOR

William Bechtel

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Neuroscience and cognitive science seek to explain behavioral regularities in terms of underlying mechanisms. An important element of a mechanistic explanation is a characterization of the operations of the parts of the mechanism. The challenge in characterizing such operations is illustrated by an example from the history of physiological chemistry in which some investigators tried to characterize the internal operations in the same terms as the overall physiological system while others appealed to elemental chemistry. In order for biochemistry to become successful, researchers had to identify a new level of operations involving operations over molecular groups. Existing attempts at mechanistic explanation of behavior are in a situation comparable to earlier approaches to physiological chemistry, drawing their inspiration either from overall psychology activities or from low-level neural processes. Successful mechanistic explanations of behavior require the discovery of the appropriate component operations. Such discovery is a daunting challenge but one on which success will be beneficial to both behavioral scientists and cognitive and neuroscientists.

Key words: mechanistic explanation, operations, laws, levels of organization, connectionism, symbolic theories

---

People and other animals behave, and a major objective of the behavioral sciences is to characterize that behavior and identify the circumstances that bring it about and the consequences that change it. The goal often is to specify laws that relate such variables to behavior. Some investigators, however, are not interested in just discovering those laws but in explaining them in terms of ongoing processes occurring inside the organism. That is, they seek to understand the mechanisms operative within the organism that explain why, given particular environmental circumstances, specific behaviors result. This was a major objective of those psychologists who in the 1950s and 1960s created cognitive psychology (Miller, 1956; Miller, Galanter, & Pribram, 1960; Neisser, 1967). It also was the goal of practitioners of other disciplines who in the 1970s came together with cognitive psychology under the designation *cognitive science* (Bechtel, Abrahamsen, & Graham, 1998) or in the 1980s and 1990s under the label *cognitive neuroscience* (Bechtel, 2001a).

I thank William Timberlake, the Special Editor for this issue, as well as Adele Abrahamsen, Colin Allen, Joel Myerson, and Leonard Green, for encouragement and helpful comments on earlier drafts of this paper.

Contact address: Department of Philosophy-0119, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119 (e-mail: bill@mechanism.ucsd.edu).

doi: 10.1901/jeab.2005.103-04

Using terminology that I will develop below, these investigators were trying to discover the mechanisms responsible for producing the behavior. A major challenge facing such investigators is to develop appropriate concepts for characterizing the operations within these mechanisms (*operation* is a technical term that will be explicated below). My contention in this paper is that, despite progress in some specific domains, an adequate characterization of the operations in psychological mechanisms still eludes investigators.

Lacking an adequate account of the operations within psychological mechanisms, cognitive psychologists and cognitive scientists often have characterized these operations as of the same type as operations performed by the whole organism. This is particularly true of the appeals to representations and operations on them that appear in many cognitive accounts. Such an approach to characterizing operations within a mechanism is not unique to those trying to characterize psychological mechanisms, but it is problematic. Operations within a mechanism occur at a lower level of organization than the behaving mechanism itself, and these operations are typically different than those performed by the whole mechanism. The types of operations occurring at a given level of organization must be discovered, and such discovery is frequently difficult. After I develop the conceptual framework of mechanism and mechanistic explanation and the related notion of level of organization in the following section, I will develop an example from 19th century physiological chemistry that is illustrative of the challenge and what is required to surmount it. I then will focus on the particular challenge in developing accounts of the operations in psychological mechanisms.

Investigators whose primary interest is in discovering laws characterizing the behavioral level itself may question the relevance to them of efforts to discover psychological mechanisms. In many cases, however, an understanding of the internal operation of the mechanism suggests environmental factors to which the whole mechanism is sensitive, suggestions that then require behavioral investigation to evaluate. In other disciplines there often is reciprocal feedback between accounts of the whole mechanism's interactions with its environment and models of its internal operation. Understanding the environmental conditions under which organisms acquire a disease guides researchers investigating disease mechanisms within the organism, but information about those mechanisms also are suggestive of environmental contingencies that can mitigate or exacerbate the disease. Inquiries into public, environmental, and occupational health and internal pathology mutually support each other. I will return to the issue of how behavioral investigations and investigations of internal psychological mechanisms can support each other in the final section of the paper.

MECHANISMS, MECHANISTIC EXPLANATIONS, AND LEVELS OF ORGANIZATION

Most accounts of the *scientific method* characterize science as the quest to discover and test *laws* that characterize the phenomenon in a particular domain. Laws have similarly played a central role in what has been the received philosophical account of science (Suppe, 1977). According to the *deductive-nomological* (D-N) model, scientific explanation involves the derivation of a statement of the phenomenon to be explained from statements of laws and initial conditions (Hempel, 1965). Such a conception of scientific explanation provides a good characterization of the attempts of behaviorists in psychology who sought, for example, to explain behavior by identifying laws relating contingencies of reinforcement to resulting behavior. Laws often provide a good way of relating behaving systems to the entities and events in their environment that affect their behavior. A major quest in the life sciences, however, is not just to relate entities to others in their environment, but also to understand why a given entity responds to environmental occurrences as it does. Scientists engaged in such quests commonly characterize themselves as seeking to understand the responsible mechanisms. Thus, in biology one finds discussion of the mechanisms of blood circulation, of thermal regulation, of cell division, and of protein synthesis.

Although what Robert Boyle termed *the mechanical philosophy* played an important role in the scientific revolution of the 17th and 18th centuries, it received little attention in 20th century philosophy of science. When advocates of the D-N model tried to account for how scientists explain why a given entity behaves as it does, they extended the same deductive framework. They proposed that the laws of a given discipline might be derived from the laws of a more basic science (together with bridge principles that would relate the different vocabularies of the two sciences and boundary conditions). They referred to such deductive relations between sciences as *reductions* (Nagel, 1961), but I will use the term *theory reduction* to differentiate this conception of reduction from the very different one provided by mechanistic explanations.

A different strategy is to try to explicate what scientists, especially those in the life sciences, have in mind when they refer to mechanisms. On my analysis, *a mechanism is an organized system of component parts and component operations. The mechanism's components and their organization produce its behavior, thereby instantiating a phenomenon* (Bechtel & Abrahamsen, 2005; for related accounts, see Bechtel & Richardson, 1993; Glennan, 1996, 2002; Machamer, Darden, & Craver, 2000). According to this view, a mechanism is a system operating in nature, and a mechanistic explanation is an epistemic product. To arrive at a mechanistic explanation, scientists must represent (sometimes verbally, but often visually in diagrams) the component parts and their operations and the ways in which they are organized.

The central feature of such mechanistic explanations is that they *decompose* a system that produces a behavior into component parts and component operations. The parts and operations into which a mechanism is decomposed are closely related: the relevant parts are those that perform operations and hence are *working parts*. However, it is important to distinguish parts understood structurally from operations understood functionally since understanding a mechanism requires both the structural and functional perspectives. Moreover, different investigatory techniques are required to establish structural and functional properties of components, and a given group of researchers may be able to secure evidence only about one or the other. As a result, researchers often confront the challenge of relating parts with operations (an activity I refer to elsewhere as *localization*).

Many of the components of a mechanism are themselves mechanisms—they perform operations in virtue of their parts (now subparts of the original mechanism) performing operations of their own. This mereological relation gives rise to a clear sense of levels—parts are at a lower level than the mechanism they comprise. Although this makes mechanistic explanations inherently reductionistic, the focus in mechanistic explanation is not exclusively downwards. Mechanisms always operate in contexts and these can affect the behavior of the mechanism itself. Moreover, insofar as there is appropriate systemic organization at that higher level, a given mechanism may be part of another mechanism that regulates its behavior. Further, with evolved mechanisms, interactions with the environment are crucial in selecting between alternative mechanisms. To address such issues in explaining the phenomenon of interest, investigators may need to take into account several higher levels of organization.

Since the mechanism is not just an aggregate of its parts (Wimsatt, 1986), but requires the component parts and operations to be organized so as to produce the behavior of the system, I follow Wimsatt (1976) in referring to levels as *levels of organization*. There have been numerous attempts to characterize levels as strata across nature by focusing on such features as the size of the component parts (P. S. Churchland & Sejnowski, 1992) or the frequency with which components interact (Wimsatt, 1976), but these fail to capture the clustering of entities into levels that figure in scientific inquiry. My recommendation is to forego the attempt to identify levels as strata across nature and focus only on what results from the attempt to differentiate the component parts and operations that figure in accounting for a given phenomenon (Craver, forthcoming; Craver & Bechtel, submitted).

When we identify levels in terms of causal interactions within a mechanism, entities that are structurally alike may appear at different levels. Protons for example interact with membranes in the chemiosmotic mechanism responsible for converting energy liberated in oxidative reactions in cells into a proton gradient that drives ATP synthesis. Protons also occur in the molecules that comprise the membrane, but these are at a lower level than the protons that are transported across the membrane and thus interact with it. There is not a level of protons, but levels corresponding to the entities that causally interact in a given mechanism. The result is a hierarchy of levels, but one that is characterized relative to the phenomenon an investigator initially set out to explain.

An important feature of the components at different levels of organization is that they typically carry out different types of operations than those at lower or higher levels. Somatic cells, for example, do such things as secrete enzymes and exchange materials with the blood, organelles of the cell do such things as synthesize proteins or extract energy from oxidation reactions, enzymes within the organelles catalyze particular reactions. This was implicitly recognized in the need for bridge principles to relate the vocabularies of two sets of laws in theory reduction accounts, but the reason different disciplines invoke different vocabularies was not identified. It is that the different disciplines focus on different kinds of operations.

It is worth emphasizing why it is that whole mechanisms can do things that their parts cannot. The secret, as engineers have long known, is to organize components appropriately so that their operations are orchestrated to produce something beyond what the components can do. It is for discovering such organization that engineers win acclaim and secure patents. Organization also is crucial in naturally occurring mechanisms—it is only as they are properly organized that the operations of the parts of a mechanism combine to generate a phenomenon that is beyond the capacity of any given part.

With this sketch of explanation in terms of mechanisms, I can comment briefly on the relation of mechanisms to laws. There are two distinct ways in which laws figure in accounts of mechanisms (Glennan, 1996). First, as noted above, laws or effects are characterizations of phenomena to be explained by mechanisms. A law identifies a regularity between values of different variables. If the relation is mediated by a mechanism, the mechanism can explain why the law holds. Thus, the mathematical relations identified within mathematical psychology often lend themselves to mechanistic explanation (Bechtel & Abrahamsen, in press). Second, some of the operations within a mechanism can themselves be characterized in terms of laws relating variables. In some cases, laws identify constraints on the ways in which different component parts can operate on others. For example, rate equations in biochemistry specify the rate at which an enzyme-catalyzed reaction can produce its product. Laws do not, however, specify the particular parts, operations, and organization in place in a particular mechanism and, in this respect, mechanistic explanations go beyond what laws provide.

In this paper my focus is on the decomposition of a mechanism into component parts and component operations. The challenge in constructing mechanistic explanations is that normally operating mechanisms do not reveal either their parts or operations. Not just any way of carving up the mechanism reveals the appropriate parts. The relevant parts are those that actually perform the operations in the mechanism. To consider an example, although neuroanatomists over several centuries sought to delineate parts of the brains of humans and other species in terms of the gyri and sulci produced by the folding of the cortex, and these still serve as useful landmarks when identifying where operations occur in the brain, they do not represent the working parts. Brodmann (1909/1994) differentiated brain regions using criteria such as types of neurons and thickness of cortical layers that he thought would map onto operations (although he lacked tools for actually localizing operations in brain regions). Modern brain mappers (Felleman & van Essen, 1991; van Essen & Gallant, 1994; see discussion in Mundale, 1998) use additional criteria such as connectivity and function to demarcate brain areas.

As challenging as it is to identify candidate working parts, it is even harder to identify the component operations. The problem in part is that identifying actual operations requires appropriate experimental interventions that can reveal evidence about them. But even more fundamental is to develop concepts to characterize the operations parts perform. I will illustrate the challenge and one way it has been resolved by developing a case history from biochemistry, a case that is particularly suggestive of the challenges facing researchers investigating psychological mechanisms.

A BIOCHEMICAL EXAMPLE OF THE CHALLENGE OF IDENTIFYING OPERATIONS

Interest in the chemical processes operative in living organisms has a long history, but the investigation was radically reshaped at the end of the 18th century when Lavoisier reconceptualized what counted as an element and hence a chemical building block of any substance. He determined that carbon, hydrogen, and oxygen are constituents of organic substances (Lavoisier, 1781). Berthollet (1780) identified nitrogen as another frequent component. With this foundation, investigators began trying to characterize physiological processes in terms of changes in elemental composition (see Holmes, 1963). For example, Lavoisier (1789) himself characterized fermentation as involving the oxygenation of carbon in part of a sugar molecule, producing carbon dioxide, at the expense of the deoxygenation of the remainder to yield alcohol. Shortly thereafter Louis Jacques Thénard (1803) and subsequently Joseph Louis Gay-Lussac (1810) worked out the general equation for fermentation, represented in modern symbolism as

$$C_6H_{12}O_6 \rightarrow 2CO_2 + 2C_2H_5OH.$$

Arrows in chemical equations such as the one above indicate that the chemists were not interested just in determining the elemental composition of organic substances but in characterizing organic processes in terms of operations involving changes in elemental composition. Since such reactions do not occur spontaneously in ordinary environments (i. e., those typically prevailing on the surface of the earth), something additional was required to make them happen. The chemist Jacob Berzelius (1836) named the responsible agent a *catalyst* and many chemists hoped that catalytic chemical changes could account for the reactions in living organisms. In this regard, Friedrich Wöhler's synthesis of urea was regarded as particularly significant. Wöhler expressed his enthusiasm for his accomplishment in a letter he wrote to Berzelius: ''I can no longer, as it were, hold back my chemical urine; and I have to let out that I can make urea without needing a kidney, whether of man or dog'' (quoted in Friedmann, 1997, p. 68).

In the first half of the 19^{th} century, the prospect of explaining the chemical reactions in living organisms in terms of changes in elemental composition seemed promising. In terms of elemental composition, William Prout (1827) classified the nutrients required by animals into three classes: saccharine (carbohydrates), oleaginous (fats), and albuminous (proteins). Prout also noted that there were only minor differences between the chemical composition of nutrients animals took in from plants and the compounds that comprised the fluids and solids of animal bodies. Perhaps the most celebrated chemist of the first half of the 19^{th} century, Justus Liebig drew upon this idea to formulate a central part of his synthetic and highly speculative account of the chemical operations occurring in animals in his *Animal Chemistry* (1842). Since animal tissue was largely composed of proteins, he proposed that animals simply incorporated protein into their tissues whereas they oxidized the carbohydrates and fats in their diet to generate heat. When insufficient oxygen was available for oxidizing carbohydrates, Liebig proposed that animals converted them to fat and stored them. He conjectured that when work occurred, the proteins incorporated into the animal body were broken down and waste products excreted. New proteins thus were continually required in animal diets to rebuild animal tissues. In this manner, Liebig articulated a general scheme for the chemical operations occurring in animals, which he filled in with detailed formulae.

The ambitious program of the organic chemists of the first half of the 19^{th} century soon encountered serious complications. Not surprisingly, given the limited empirical evidence upon which he built his theory, Liebig's (1842) proposals fared poorly as empirical results emerged. Some of this evidence remained at the level of the whole organism and involved feeding experiments in which researchers measured food intake of various food groups, resulting waste products, and energy expenditure, and demonstrated that these failed to conform to Liebig's hypothesis. Fick and Wislicenus (1866), for example, used themselves as subjects and prepared to climb Mt. Faulhorn in the Swiss Alps by consuming a non-protein diet. They also measured their urine before, during, and after the ascent. They calculated the energy expended on the climb and determined that it greatly exceeded the amount accounted for by the nitrogen waste in their urine. Contrary to Liebig, the energy they expended on the climb must have come from carbohydrates and fats (other feeding studies were conducted by Frankland, 1866; Smith, 1862). Equally serious for Liebig's project was the recognition that the chemical reactions in living organisms were more complex than he anticipated. Claude Bernard (1848), for example, sought to trace where glucose was consumed in animals and discovered that it was actually synthesized in the liver. This showed that animal metabolism could not be understood as a linear chain of catabolic reactions. Identifying the more complex pattern of chemical processes involved in living systems, however, was challenging.

In the second half of the 19^{th} century, fermentation assumed a central place in pointing to the limitations of attempts to explain physiological processes chemically. Until the investigations of Kützing (1837), Schwann (1837) and Cagniard-Latour (1838) that indicated that alcoholic fermentation fundamentally involved living organisms, most chemists had assumed that fermentation was an ordinary chemical reaction simply requiring a catalyst. Leading chemists reacted harshly to the claim that living organisms were involved since it seemed a step backwards in the attempt to explain physiological processes. Wöhler published excerpts of a paper by Turpin (1838) following up on Cagniard-Latour's research in *Annalen der Pharmacie* (a journal he and Liebig edited), and followed it with a heavy-handed satire entitled ''The demystified secret of alcoholic fermentation.'' It purported to present detailed observations made with a special microscope of little animals shaped liked distilling flasks that had complete digestive systems and eliminated alcohol from their intestinal tract after digesting sugar. But the linkage of fermentation with living organisms was further secured through the investigations of Pasteur, who concluded ''Fermentation is correlated to the vital processes of yeast'' (Pasteur, 1860, p. 323). As the chemists feared, this seemed to put fermentation beyond the reach of chemical explanation, and nearly forty years intervened before Eduard Buchner (1897) discovered, serendipitously, that fermentation could occur in press juice in which no whole cells remained and attributed it to a catalyst (catalysts existing within living systems now being designated *enzymes*) he named *zymase*.

Advances in organic chemistry also posed a challenge to the project of providing an account of physiological processes in terms of changes in elemental composition. Organic chemists in the later decades of the 19th century determined that chemical compounds were not just composed of atoms but were structured. A consequent was that not every chemical formula designating a combination of elements corresponded to actually occurring substances. This indicated the need to consider chemical structure in explaining physiological processes.

The challenge was how to do so. One sort of investigation organic chemists pursued was to decompose glucose with a number of alkalis in the attempt to identify compound structures, not elements, out of which it might be composed. Researchers identified several three-carbon sugars—methylglyoxal, glyceraldehyde, and dihydroxyacetone. Were these intermediates in the processes in yeast that transformed sugar to alcohol? To answer this question, investigators supplied them to a fermenting system (yeast or, after Buchner, cell-free extracts) to see whether they would generate alcohol. What is particularly interesting is how researchers characterized these investigations. They asked whether methylglyoxal, for example, would *ferment* as rapidly as sugar. Abandoning the attempt to explain the processes in elemental terms, they now could only use the same vocabulary as applied to the overall process to the possible component operations.

The challenge confronting those seeking to provide chemical explanations of basic physiological processes was to characterize the component operations (reactions) at an appropriate level of organization. Elemental composition was too low a level at which to characterize changes, while decomposing fermentation into fermentations simply invoked the vocabulary designed to explain the overall behavior to describe the operation of its components. It did not explain the process in terms of something more basic. Fortunately for these researchers, at about this same time a new framework became available. Organic chemists' efforts to determine the structure of organic compounds revealed that they were composed of groups of molecules such as amino (NH_3^+), carboxyl (COO^-), hydroxyl (OH^-), and phosphate (PO_4^{3-}) groups that were bound to a carbon ring backbone (Holmes, 1992). Reactions would involve whole groups being added, deleted, or moved on the backbone—such as deamination (removal of an amino group), carboxylation (addition of a carboxyl group), dehydroxylation (removal of an hydroxyl group), phosphorylation (addition of a phosphate group), etc. This provided the basis for conceptualizing types of reactions at a level above that of elemental composition and provided the resource biochemistry needed to begin working out the intermediate steps in numerous physiological processes.

The view of physiological processes as involving pathways of successive operations involving chemical groups, together with the proposal that these reactions were catalyzed by enzymes, provided the guiding assumptions of the newly emerging discipline of biochemistry. For example, one of the best-known biochemical pathways, the citric acid or Krebs cycle, consists of successive steps involving oxidations (removal of 2H groups, picked up by NAD^+ or FAD), hydrations and dehydrations (adding or removing H_2O groups), decarboxylations (removal of CO_2 groups), addition or removal of sulfhydryl-CoA groups, etc. (see Figure 1).

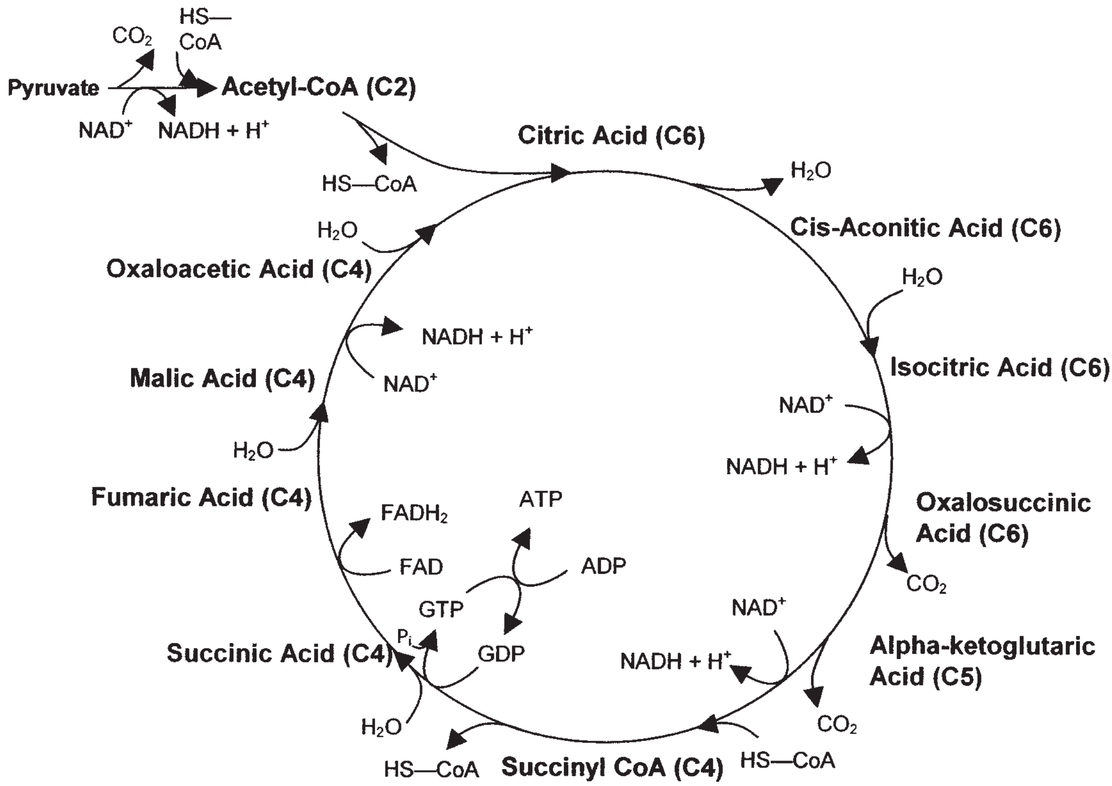

Fig. 1. Citric acid or Krebs cycle. In this prototypical biochemical pathway, each reaction (except for the condensation reaction between oxaloacetic acid and acetyl-CoA) involves operations of adding or removing groups of molecules (shown by arrows coming in or out of the overall cycle) from the previous substrate.

Ultimately, the biochemical level was not the only level at which researchers had to discover operations to develop a complete mechanistic account of bioenergetics. The investigatory techniques of biochemistry involved destroying cell structures to solubilize enzymes in a homogenate, fostering the *sac of enzymes* view of the cell. Some cell processes, such as the conversion of energy from foodstuffs into ATP, which provides temporary energy storage, depend on the structure of cell organelles as well. Understanding this level of organization required another new set of research techniques that figured prominently in the development of the new field of cell biology, in the 1940s and 1950s (see Bechtel, 2006). Understanding how the new structures figured in processes such as those of bioenergetics required conceptualizing yet further types of operations such as vectoral transport across membranes. The process of identifying unsuspected levels of organization situated between existing levels and conceptualizing the types of operations that occur there is a recurring step in the development of mechanistic explanation in the life sciences.

CHALLENGE: DETERMINING THE NATURE OF PSYCHOLOGICAL OPERATIONS

As I noted at the outset, when cognitive psychologists and cognitive scientists set out to discover the mechanisms responsible for behavior, they frequently characterized the operations in these mechanisms using concepts developed to describe the behaviors in which cognitive agents engage. This perspective is most clear in the symbolic or symbol manipulation approach to modeling cognitive activity. In it, psychological operations are viewed as transformations on symbol structures, where these symbol structures are construed as being much like sentences in a natural or a formal language. Fodor (1975) quite appropriately characterized these theorists as committed to "a language of thought." The operations in turn are much like those humans themselves perform when doing such tasks as writing a manuscript—typing words and phrases, reading them back, altering some, etc. The main difference is that these symbols are thought to be encoded in some way inside a person's brain, and the operations of reading and writing are internal operations, not operations on paper.

In this regard, it is interesting to note that Turing (1936; see also Post, 1936), in proposing the Turing machine as a computational device, was explicitly trying to model human computers—humans whose occupation was to carry out complex mathematical computations. Subsequently, the Turing machine often has been invoked by advocates of the symbolic account as the exemplar for the kind of device the mind is taken to be. In this instance, an activity performed by humans provided the model for operations occurring in their minds. It should be apparent that such invocation of symbol processing to explain how minds work is comparable to physiological chemists' invocation of fermentations as intermediate processes in alcoholic fermentation. The component operations within the posited mechanism are of the same sort as the behaviors of the mechanism itself.

Cognitive psychology is not just a theoretical enterprise hypothesizing internal operations; like physiological chemistry, its practitioners offer empirical evidence for their hypotheses. This evidence is often secured through behavioral measures such as reaction times (Donders, 1868). Early cognitive research in psycholinguistics provides an illustrative example. Psychologists extended Chomsky's (1957) proposals for generative grammar, developed initially simply to provide a compact account of the structure of language itself, to characterize the operations performed when people comprehend or construct sentences. Sentences whose grammatical analysis involved more transformations were hypothesized to require additional psychological operations, which would require additional time. Reaction time studies revealed that sentences requiring more transformations in the grammar did take longer to process than sentences requiring fewer operations, suggesting that the grammatical transformations were also psychologically real (Miller, 1962; for history and perspective, see Abrahamsen, 1987; Reber, 1987). Early research on memory exhibited a similar pattern. Sternberg (1966) compared different models of memory search, which all assumed that memory involved the storage of symbolic structures and mentally scanning them. These predicted different patterns of reaction times and he argued that the model that fit best characterized actual human psychological operations.

One of the most powerful tools for constructing artificial intelligence models, Newell and Simon's method of protocol analysis, made conceptualizing internal psychological operations on the basis of agent-level behaviors almost inevitable. They required subjects to talk aloud as they solved problems such as the Tower of Hanoi problem so as to elicit the steps the subjects employed in solving the problem. These operations then became the building blocks of their computational models, which were then further tested by data such as that provided by reaction time measures (Newell & Simon, 1972). The production-system architecture, which became the foundation for some of the most powerful computational models of human performance (Anderson & Lebiere, 1998; Rosenbloom, Laird, & Newell, 1993) developed out of this perspective. The fundamental idea of this architecture is that just as human agents have a variety of strategies that can be elicited by the problems they are trying to solve (and partial solutions already obtained), their minds are assumed to be equipped with productions that are executed when appropriate symbol strings are active in working memory.

It is possible that operations within psychological mechanisms do have the same character as those performed by human agents, but if so this is a very unusual case in the history of science. Typically, the operations within a mechanism that enable it to perform its behaviors are different in kind from those behaviors. The ability of mechanisms to perform behaviors different from those that their component parts perform is what makes mechanistic explanations so powerful. As noted above, organization is the key to achieving this. Although evolutionary arguments are subject to much abuse, a minimal appeal to evolution enables us to note that distinctive human behaviors largely originate through reorganization of components found in the brains of our close primate relatives. It is also operations performed in these other species, organized in novel ways, that permit human performance. It seems peculiar to propose that symbol-processing components would have evolved in species that had yet to develop the capacity to manipulate symbols.

If not from characterizations of the behavior of humans, where can investigators draw insights as to the nature of internal psychological operations? The prime alternative to which theorists have appealed is neuroscience. Such was the origin of the prime competitor to the symbol-processing paradigm in cognitive science. During the same period as the symbol-processing paradigm was developing, other scientists appealed to basic ideas about how brains work to construct an alternative perspective. In this alternative, parts of the mechanism (commonly called *units*) pass activations to each other and individual components become active when they receive the appropriate activation from other units to which they are connected (McCulloch & Pitts, 1943; Pitts & McCulloch, 1947; Rosenblatt, 1962). Although it encountered severe limitations in its first incarnation (Minsky & Papert, 1969), the approach reappeared in the 1980s under the banner of *parallel distributed processing* (PDP) or *connectionism* (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986; for an introduction designed to be accessible to non-specialists, see Bechtel & Abrahamsen, 2002).

Although connectionist accounts do not face the objection of using the behaviors to be explained as models for the operations appealed to in the mechanism explaining them, they exhibit the opposite shortcoming of appealing to what is likely to be too low a level of organization to characterize the operations. Recall that in the early 19th century many chemists attempted to explain physiological processes directly in terms of elemental composition. Although it is certainly true that changes in elemental composition of substrates occur in physiological processes, the relevant operations involved higher-level molecular units. Likewise, operations within psychological mechanisms involve neurons, but the operations themselves likely involve parts at a higher level than individual neurons.

In perceptual processing, neuroscience itself has made significant progress in identifying higher-level structures. For example, the component parts in contemporary accounts of visual processing are not individual neurons, but brain areas involving populations of neurons. Investigators characterize areas such as V1, V4, and MT as extracting different types of information from the input signal (edges of objects, shape and color, motion) and making it available to areas downstream for further analysis (van Essen & Gallant, 1994; see Bechtel, 2001b, for analysis and an historical account of the discovery of visual-processing mechanisms). Discovering mechanisms of vision was facilitated by both a fruitful technique (single-cell recording) and the fact that the visual system processes sensory input. Although single-cell recording actually records from individual neurons, it revealed that neurons in a particular area all processed similar types of information from different parts of the visual field. As well, within each region there was internal structure: neurons organized into columns involving layers of connected units that process information from the same part of the visual field. To determine what sort of information a given area extracted researchers could vary the stimuli and correlate inputs with responses. In many respects the kinds of information that visual areas extract are what one might expect from characterizing performance at the behavioral level—people see colors, shape, motion, etc. But the details are often surprising. The shapes detected, for example, are frequently not simple Cartesian shapes but rather more complex forms, and the motions are often nonlinear (van Essen & Gallant, 1994).

The fact that advances in discovering what information an area processed resulted from fortune and not from hypotheses being tested in experiments, suggests of how hard it is to figure out the component operations from behaviors. Hubel (1982) reports that he and Wiesel, for example, discovered edge-detecting cells in V1 when a slide stuck in the projector from which they had been projecting dark spots on light backgrounds and vice versa. Gross (1998) reports that he identified shape-detecting cells in inferotemporal cortex when, in frustration after projecting stimuli to which the cells did not respond, he waved his hand in front of the monkey's face.

Moving beyond vision to what are thought of as higher cognitive processes (reasoning, memory, language processing) is more challenging because investigators cannot so readily control the operations a person performs at a given time by controlling the stimulus. What has emerged as the dominant approach for linking psychological processes with brain activity is functional neuroimaging, in which investigators measure blood flow changes as subjects perform tasks. But, as Petersen and Fiez (1993) made very clear, the object in such research was not to localize tasks, although in early imaging studies, finding increased blood flow in only one or a small number of brain areas as subjects performed tasks, fueled such interpretations. As imaging techniques matured, neuroimaging has begun to identify multiple brain areas characterized as networks engaged in performing the task. But what does each area do? Here neuroimaging confronts the same problem I have been focusing on in this paper—characterizing the component operations.

Biochemistry was fortunate in that structural information about organic molecules provided it with information about higher-level structures on which enzymes operated. Cognitive science and cognitive neuroscience are unlikely to be able to take advantage in any direct way of information coming from the brain, making the challenge of discovering the nature of the component operations much greater. One problem is that cortical structures do not vary much. (This is in contrast to subcortical areas such as the hippocampus, where each subregion has a distinctive pattern of connectivity, with the particular connectivity pattern suggesting that one area may be performing a task like pattern generalization while another is performing pattern separation. See Rolls & Treves, 1998, who have drawn upon such clues to develop an analysis of its operation.)

I can foresee two strategies that may help guide the discovery of appropriately characterized cognitive operations. One is the discovery through techniques such as neuroimaging that the same brain areas are involved in multiple tasks, and then trying to assess what might be common requirements of the different tasks. The other, involving comparative psychology—discovery of the tasks in which related species use areas homologous to those in our brains—may likewise lead researchers to consider what operations contribute to both tasks (see Deacon, 1989, 1997, for probing suggestions with respect to the operations involved in human language processing). Ultimately, however, there is no simple algorithm for discovering the type of operations into which the behavioral system should be decomposed. There may be no alternative but for cognitive scientists to employ accounts of operations drawn from what are likely too high or too low a level while awaiting inspired theorizing. If I am right, though, such a theoretical advance is essential if cognitive science is to succeed in the search for mechanisms.

WHY WORRY ABOUT MECHANISMS?

If the challenge in discovering component operations is as great as I have proposed, one might question whether the quest to discover the mechanisms underlying behavior is worth pursuing. Perhaps it would be wiser for behavioral scientists to limit their focus to regularities discoverable in behavior and not concern themselves with the mechanisms that underlie them. In this final section I will focus on the value of accounts of mechanisms for understanding behavioral regularities, as well as the converse value of behavioral accounts for investigators seeking to understand internal mechanisms.

An understanding of the responsible mechanism, even a partial and flawed understanding, can serve as a valuable guide to developing and articulating further the account of the overall behavior. A hypothesis about the mechanism can suggest different contextual variables that affect behavior. For example, an understanding of the mechanism often will point to kinds of factors that can promote, alter, or disrupt it. These factors may be external to the mechanism, but without hypotheses about the mechanism, there may be little motivation to examine how they affect behavior. Nutrition research provides a good example. Although the first vitamin-deficiency diseases were identified prior to research linking vitamins with metabolic coenzymes, that discovery brought a major change in research on vitamins. Until then, vitamin-deficiency diseases were quite mysterious since dietary substances had been viewed as being burned (oxidized) to yield energy. Vitamins, however, were required in such small quantities that they could not be used in this way. Cofactors in enzyme-catalyzed reactions, however, are not consumed but reused repeatedly and hence are required in very small quantities. Once the vitamin-coenzyme link was established, any time an additional cofactor was discovered, that discovery motivated a search for whether that substance also might be a vitamin, and thus a requirement of an adequate diet (Bechtel, 1984).

Similarly, understanding the psychological mechanisms underlying behavior can guide researchers to new discoveries at the behavioral level. For example, the discovery of different *what* and *where* processing streams in the mechanisms underlying vision led Chen, Myerson, Hale, & Simon (2000) to look for and find behavior differences in tasks that emphasized one or the other processing stream. Without the clues from the mechanism, there was little reason to suspect such behavioral differences. Likewise, P. M. Churchland (in press) used information about opponent processing in the visual system to predict and demonstrate the human ability to see what he calls chimerical colors—colors beyond those found in the Munsell color spindle of colors usually experienced.

Just as an (even partial) understanding of the mechanism can contribute to the further development of behavioral level accounts, so, too, behavioral level accounts are crucial to the development of mechanistic accounts. To begin with, if scientists are theorizing about a mechanism to explain a particular kind of behavior, it is indispensable to begin with a good characterization of the behavior. Otherwise, they may produce a proposal for a possible mechanism that does not in fact exist, and whose behavior would not correspond to anything that actually happens. Moreover, once a mechanism is proposed, the evidence for or against it comes not just from investigations of internal operations but from whether it actually can account for factors that are known to affect the behavior. Liebig's (1842) proposal for animal chemistry (discussed earlier) failed not just because it falsely ruled out synthetic processes within organisms, but also because it falsely predicted dietary requirements of organisms. The development of a proposed mechanism does not obviate the need for behavioral investigation but in fact is a spur to discovery of the behavioral regularities predicted from the proposed mechanism.

I earlier characterized the search for mechanisms as reductionistic in that researchers decomposed mechanisms into their parts, which are entities at lower levels of organization. Sometimes reductionistic research is thought to undermine the importance of research at higher levels of organization—once we understand the mechanism at the lower level, it is thought that nothing remains for the higher level to provide (Bickle, 2003). But an understanding of mechanisms makes clear why such a view is mistaken. The behavior of a mechanism depends as much on engagement with things in its environment (including other mechanisms) as on the parts that comprise it. From the perspective of scientists, a mechanism is ultimately a multi-level integrator, providing a framework for relating information about the context in which a mechanism behaves and the internal operation of the mechanism.

CONCLUSION

I have focused on the challenge of characterizing the operations within the mechanisms that underlie psychological behavior. The challenge is not unique, as the discussion of the history of physiological chemistry shows. Operations within a mechanism are typically of a different sort than the behaviors of the mechanism—this is why scientists invoke mechanisms to explain behaviors. This was a major shortcoming of appeals to fermentations to explain fermentation and of most accounts in cognitive psychology, cognitive science, and cognitive neuroscience that model component operations on symbol-processing activities performed by people. But trying to proceed from neural processing confronts the same limitations as trying to explain physiological processes in terms of elemental chemistry. To account for the mechanisms underlying behavior, investigators need to discover a distinctive set of operations that relates to psychological mechanisms, just as biochemists had to discover the appropriate set of component operations underlying physiological processes. As difficult as this challenge is, it is an important one to surmount because of the valuable payoff to integrating behavioral accounts with those of underlying mechanisms.

REFERENCES

Abrahamsen, A. A. (1987). Bridging boundaries versus breaking boundaries: Psycholinguistics in perspective. *Synthese, 72,* 355–388.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Bechtel, W. (1984). Reconceptualization and interfield connections: The discovery of the link between vitamins and coenzymes. *Philosophy of Science, 51,* 265–292.

Bechtel, W. (2001a). Cognitive neuroscience: Relating neural mechanisms and cognition. In P. Machamer, P. McLaughlin, & R. Grush (Eds.), *Theory and method in the neurosciences* (pp. 81–111). Pittsburgh, PA: University of Pittsburgh Press.

Bechtel, W. (2001b). Decomposing and localizing vision: An exemplar for cognitive neuroscience. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 225–249). Oxford: Basil Blackwell.

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

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks.* (2nd 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., & Abrahamsen, A. (in press). Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R. Stainton (Ed.), *Contemporary debates in cognitive science.* Oxford: Basil Blackwell.

Bechtel, W., Abrahamsen, A., & Graham, G. (1998). The life of cognitive science. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science* (pp. 1–104). Oxford: Basil Blackwell.

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

Bernard, C. (1848). De l'origine du sucre dans l'économic animale. *Archives générales de médecine, 18,* 303–319.

Berthollet, C. L. (1780). Recherches sur la nature des substances animales et sur leurs rapports avec les substances végétales. *Mémoires de l'Académie royale des sciences,* 120–125.

Berzelius, J. J. (1836). Einige Kideen über bei der Bildung organischer Verbindungen in der lebenden Naturwirksame, aber bisher nicht bemerke Kraft. *Jahres-Bericht über die Fortschritte der Chemie, 15,* 237–245,

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account.* Dordrecht: Kluwer.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde.* Leipzig: J. A. Barth.

Buchner, E. (1897). Alkoholische Gärung ohne Hefezellen (Vorläufige Mittheilung). *Berichte der deutschen chemischen Gesellschaft, 30,* 117–124.

Cagniard-Latour, C. (1838). Memoire sur la fermentation vineuse. *Annales de chimie et de physique, 68,* 206–223.

Chen, J., Myerson, J., Hale, S., & Simon, A. (2000). Behavioral evidence for brain-based ability factors in visuospatial information processing. *Neuropsychologia, 38,* 380–387.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Churchland, P. M. (in press). Chimerical colors: Some novel predictions from cognitive neuroscience. *Philosophical Psychology.*

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain.* Cambridge, MA: MIT Press.

Craver, C. (forthcoming). *Explaining the brain: What a science of the mind-brain could be.*

Craver, C., & Bechtel, W. (2005). *Explaining top-down causation (away).* Manuscript submitted for publication.

Deacon, T. W. (1989). The neural circuitry underlying primate calls and human language. *Human Evolution, 4,* 367–401.

Deacon, T. W. (1997). *The symbolic species.* New York: Norton.

Donders, F. C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool: 1868–1869. *Tweede Reeks, 2,* 92–120.

Felleman, D. J., & van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fick, A. E., & Wislicenus, J. (1866). On the origin of muscular power. *Philosophical Magazine & Journal of Science London, 4th ser., 31,* 485–503.

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

Frankland, E. (1866). On the source of muscular power. *Royal Institution of Great Britain. Notices of the proceedings at the meetings of the members, 4,* 661–685.

Friedmann, H. (1997). From Friedrich Wöhler's urine to Eduard Buchner's alcohol. In A. Cornish-Bowden (Ed.), *New beer in an old bottle: Eduard Buchner and the growth of biochemical knowledge* (pp. 67–122). Valencia: Universitat de València.

Gay-Lussac, J. L. (1810). Extrait d'un mémoire sur la Fermentation. *Annales de chimie, 76,* 245–259.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44,* 50–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69,* S342–S353.

Gross, C. G. (1998). *Brain, vision, and memory.* Cambridge, MA: MIT Press.

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.

Holmes, F. L. (1963). Elementary analysis and the origins of physiological chemistry. *Isis, 54,* 50–81.

Holmes, F. L. (1992). *Between biology and medicine: The formation of intermediary metabolism.* Berkeley, CA: Office for History of Science and Technology, University of California at Berkeley.

Hubel, D. H. (1982). Evolution of ideas on the primary visual cortex, 1955–1978: A biased historical account. *Bioscience Reports, 2,* 435–469.

Kützing, F. T. (1837). Microscopische Untersuchungen über die Hefe und Essigmutter, nebst mehreren andern dazu gehörigen vegetabilischen Gebilden. *Journal für praktische Chemie, 11,* 385–409.

Lavoisier, A. L. (1781). Mémoire sur la formation de l'acide nommé air fixe ou acide crayeux, que je désignerai désormais sous le nom d'acide du charbon. *Mémoires de l'Académie royale des sciences,* 448–458.

Lavoisier, A. L. (1789). *Traité élémentaire de chimie, présenté dans un ordre nouveau et d'après les découvertes modernes.* Paris: Cuchet.

Liebig, J. (1842). *Animal chemistry: or organic chemistry in its application to physiology and pathology.* Cambridge: John Owen.

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

McClelland, J. L., & Rumelhart, D. E. (Eds.) (1986), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

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. (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. (1962). Some psychological studies of grammar. *American Psychologist, 17*, 748–762.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.* New York: Holt.

Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry.* Cambridge, MA: MIT Press.

Mundale, J. (1998). Brain mapping. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science* (pp. 121–128). Oxford: Basil Blackwell.

Nagel, E. (1961). *The structure of science.* New York: Harcourt, Brace.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Pasteur, L. (1860). Mémoire sur la fermentation alcoolique. *Annales de Chimie, 3e Ser, 58*, 323–426.

Petersen, S. E., & Fiez, J. A. (1993). The processing of single words studied with positron emission tomography. *Annual Review of Neuroscience, 16*, 509–530.

Pitts, W. H., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics, 9*, 127–147.

Post, E. L. (1936). Finite combinatorial processes - Formulation I. *Journal of Symbolic Logic, 1*, 103–105.

Prout, W. (1827). On the ultimate composition of simple alimentary substances; with some preliminary remarks on the analysis of organised bodies in general. *Philosophical Transactions of the Royal Society of London, 117*, 355–388.

Reber, A. S. (1987). The rise and (surprisingly rapid) fall of psycholinguistics. *Synthese, 72*, 325–339.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function.* Oxford: Oxford University Press.

Rosenblatt, F. (1962). *Principles of neurodynamics; perceptrons and the theory of brain mechanisms.* Washington: Spartan Books.

Rosenbloom, P. S., Laird, J. E., & Newell, A. (Eds.) (1993), *The Soar papers: Research on integrated intelligence.* Cambridge, MA: MIT Press.

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

Schwann, T. (1837). Vorläufige Mitteilung, betreffend Versuche über die Weingärung und Faulnis. *Poggendorf's Annalen der Physik und Chemie.* 41 184–193.

Smith, E. (1862). On the elimination of urea and urinary water. *Philosophical Transactions of the Royal Society, London, 151*, 747–834.

Sternberg, S. (1966, August 5). High-speed scanning in human memory. *Science, 153*, 652–654.

Suppe, F. (1977). The search for philosophical understanding of scientific theories. In F. Suppe (Ed.), *The structure of scientific theories* (pp. 3–241). Urbana: University of Illinois Press.

Thénard, L. J. (1803). Mémoire sur la Fermentation vineuse. *Annales de Chimie, 46*, 294–320.

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

Turpin, P. J. F. (1838). Mémoire sur la cause et les effets de la fermentation alcoolique et acéteuse. *Annales de chimie et de physique, 7*, 369–402.

van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*, 1–10.

Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. In G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the brain: A scientific and philosophical inquiry* (pp. 202–267). New York: Plenum Press.

Wimsatt, W. C. (1986). Forms of aggregativity. In A. Donagan, N. Perovich, & M. Wedin (Eds.), *Human nature and natural knowledge* (pp. 259–293). Dordrecht: Reidel.

Received September 14, 2004
Final acceptance July 8, 2005