Mechanism

Cory Wright & William Bechtel

1. Introduction

What is it to explain a psychological phenomenon (e.g., remembering names, comprehending words, solving Tower of Hanoi problems, remaining vigilant)? In philosophy, a traditional answer is that to explain a phenomenon is to show it to be the expected result of prior circumstances given a scientific law. Influenced by this perspective, behaviorists directed psychology toward the search for the laws of learning that explained all behavior as the consequence of particular conditioning regimens. Yet, with the rise of the cognitivist tradition, appeal to laws in the explanation of psychological phenomena has moved to the periphery in psychology proper, although discussion of laws remains commonplace in philosophical discourse about psychology. Examination of the explanatory discourse of psychologists reveals a shift in emphasis from laws to *mechanisms*—mechanisms of learning, memory, and attention, mechanisms of route navigation, mechanisms of drug addiction, mechanisms of development, etc. This raises a substantive philosophical issue: what is a mechanism, and how does discovering and specifying one figure in an explanation?

Although largely neglected in recent philosophical discourse about psychology, the search for mechanisms is one of the principal strategies for rendering the natural world intelligible through scientific investigation. It lay at the foundation of the scientific revolution of the 17th and 18th centuries and was enshrined in the mechanical philosophies advanced by Galileo, Descartes, and Boyle, among others. The dialectic between mechanistic and anti-mechanistic thinking played a decisive role in framing issues for theorizing about mental phenomena in the 19th century, and further shaped the genesis of psychology as a discipline in the 1880s and beyond. Indeed, the rise of cognitive psychology around the 1960s was guided by a particular, information processing, conception of mechanism.

In the next two sections we will briefly describe the development of the mechanical philosophy and its applications to mental phenomena, and will then turn toward a more analytical characterization of mechanism and mechanistic explanation. Understanding what mechanisms are and recognizing the role they play in psychology enables a number of traditional philosophical issues about psychology to be better characterized.

2. The rise of the mechanical philosophy and its application to the mind

Rene Descartes is a pivotal figure in the history of the sciences, and, arguably, his most influential contribution was the heralding of a mechanistic view of the natural world.[1] Whereas classical thinkers primarily viewed machines as devices operating *against nature* that satisfy human purposes (e.g., to lift heavy weights or launch projectiles in opposition to their natural downwards motion), Descartes proposed that natural systems were mechanical. He noticed

[1] Our discussion of Descartes' mechanical philosophy follows the analysis offered by Garber (2002).

mechanisms at work throughout the natural world—including the bodies and nervous systems of human and non-human animals (indeed, the human mind was virtually the only domain where he took them to be absent).

The mechanisms familiar to Descartes (e.g., clocks, which were undergoing rapid development the 17th century), typically produced their effects because of the shape, motion, and contact between their parts. So if natural systems are mechanical, then they could likewise be rendered explicable by appealing to the shape and motion of their parts: "I have described this earth and indeed the whole universe as if it were a machine: I have considered only the various shapes and movements of its parts" (Principia IV, p. 188). Two examples of physical phenomena—gravity and magnetism—will illuminate Descartes' appeal to mechanics. Explaining either phenomenon depends critically on the assumption that the physical universe is comprised of contiguous bodies such that no empty space or vacuum exists. Wherever space seems to be empty, as in the heavens, Descartes assumed that it was filled with a very fine material, the ether. Descartes maintained that, when an object moves, something else (such as the ether or another object) must immediately move into the space vacated. To explain gravity, then, Descartes appealed to the vortex created by the rapidly circulating ether, which forced objects downwards towards the center of the earth. In a similar manner, he proposed that the vortex surrounding the Sun served to hold the planets in their orbits. To explain magnetism, Descartes again invoked the model of vortex action, but this time implicated the motion of screw-threaded particles circulating around the magnet. These particles would screw themselves into corresponding threaded channels in a nearby metallic object such that the magnet and object would move together. Thus, it was the shape and motion of microscopic particles that he thought determined the behavior of macroscopic objects.

Descartes faced several challenges in developing such accounts of physical phenomena. In particular, the parts that he posited as constituting physical objects—minute corpuscles—were too tiny to be seen by the unaided eye. Yet he was undeterred by the fact that the properties of these particles therefore had to be inferred.

> I do not recognize any difference between artifacts and natural bodies except that the operations of artifacts are for the most part performed by mechanisms which are large enough to be easily perceivable by the senses—as indeed must be the case if they are to be capable of being manufactured by human beings. The effects produced by nature, by contrast, almost always depend on structures which are so minute that they completely elude our senses. (Descartes, 1644, Part IV, §203)

Descartes proposed to infer the properties of these corpuscles by a kind of reverse engineering, remarking that:

> Men who are experienced in dealing with machinery can take a particular machine whose function they know and, by looking at some of its parts, easily form a conjecture about the design of the other parts, which they cannot see. In the same way I have attempted to consider the observable effects and parts of natural bodies and track down the imperceptible causes and particles which produce them. (Descartes, 1644, part IV, §203)

Descartes invoked mechanistic processes to explain biological phenomena as well. He was quite impressed with William Harvey's account of the circulation of blood, although he did not follow

Harvey in construing the heart as a pump. Instead, Descartes proposed that the heart served to heat the blood, thereby causing it to expand and dilate, so that the corpuscles of the blood could move out through the arteries until they cooled in the capillaries and returned to the heart through the veins. Taking the circulation of blood as his starting point, Descartes offered similar mechanistic accounts of the behavior of various organs of the body.

The idea that natural phenomena—including physiological processes—are produced by the activity of mechanisms was already a radical departure from the traditions based on Aristotelian science, which endorsed teleological explanation. Yet Descartes made a further, controversial move in developing his mechanical philosophy. He maintained that *all* behavior exhibited by animals was generated mechanically and so did not require positing purposes or goals. Of paramount inspiration for this additional move were his encounters with the hydraulically controlled statues in the Royal Gardens at St. Germain-en-Lai outside of Paris. The opening and closing of valves in the plumbing, which resulted from visitors stepping on critical tiles, caused these statues to move in anthropomorphic ways. Consequently, Descartes proposed that a very fine fluid, which he called *animal spirits*, likewise ran through the nerves in animal bodies, causing them to respond differentially to various sensory stimulations.

> In proportion as these [animal] spirits enter the cavities of the brain, they pass thence into the pores of its substance, and from these pores into the nerves; where, according as they enter, or even only tend to enter, more or less, into one than into another, they have the power of altering the figure of the muscles into which the nerves are inserted, and by this means of causing all the limbs to move. Thus, as you may have seen in the grottoes and the fountains in royal gardens, the force with which the water issues from its reservoir is sufficient to move various machines, and even to make them play instruments, or pronounce words according to the different disposition of the pipes which lead the water (Descartes, 1664, AT VI, 130)

Moreover, Descartes did not see any reason to distinguish human and non-human animals in this respect; any human behavior that was comparable to that of non-human animals was likewise the product of mechanisms operative in the physical body. But as humans do complicated things which animals do not, Descartes concluded that total mechanistic explanation was not possible. One such human activity is the construction and comprehension of novel sentences:

> We can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of a corporeal kind, which brings about a change in its organs; for instance, if it is touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. (Descartes, 1637, Part V)

A second activity for which he thought mechanistic explanation failed is the ability to reason with regard to any given topic. Although animals might exhibit intelligent behavior in particular domains, Descartes maintained that this particularity revealed that the behavior was therefore not due to reason, but to a cleverly designed mechanism. He compared an animal’s superior performance in a given domain to “a clock which is only composed of wheels and weights,” yet “is able to tell the hours and measure the time more correctly than we can do with all our wisdom.” Reason is a universal capacity, and if animals did act from reason, their greater capacity in one domain would result in greater capacity in all domains.

Thus, for Descartes, mechanisms lacked the flexibility needed to account for the variability and context-sensitivity of language use and reason. Instead of explaining these activities in terms of mechanisms, he attributed them to an immaterial mind, which he construed as a distinct substance. Whereas physical substance was defined by the primary attribute of being extended and therefore occupying space, Descartes construed mind in terms of the attribute of thinking: The mind is "a thing that thinks. What is that? A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions" (Descartes, 1658, 2). Because it is not physical, the mind does not occupy space and is not located in space.

Having made a sharp distinction between material bodies and the immaterial mind, Descartes faced the potentially embarrassing problem of explaining how the one could affect the other. Famously, he located the site of the mind's interaction with the body at the pineal gland. Since Descartes viewed the nerves as conduits for animal spirits, if the mind was to have impact, it had to affect the flow of animal spirits. The pineal gland's central location appeared to Descartes to be the point where such interaction would most readily be achieved, altering the flow of fluids through the ventricles of the brain through slight shifts in position.

Descartes' substance dualism was an unstable feature of his philosophy. Some of his followers, such as Julian Offray de La Mettrie (1748), argued for extending the mechanistic view to the human mind. But far more common was an anti-mechanistic attitude toward mental processes. Even some who found the problem of interaction to pose sufficiently serious problems to undercut Cartesian dualism nonetheless rejected a mechanistic construal of mental processes.

This attitude is well-illustrated at the beginning of the 19th century in Jean-Pierre-Marie Flourens' opposition to Franz Joseph Gall's proposal to distinguish a number of different mental functions and localize each in a different brain area based on cranial shape. Up to a point, Flourens supported the project of localizing functions in the brain. But he based his own inferences on experimental lesions (primarily in birds), rejecting the reliance on correlational methodology and cranial measures that subsequently inspired much of the negative commentary on Gall's phrenology. Flourens made the important discovery that coordinated movement is controlled in the cerebellum and also found support for Gall's overall claim that mental activity is localized in the cerebral hemispheres.[2] However, Flourens attacked Gall's key claim that mental activity should be divided into isolable parts, each seated in its own area of the brain. Arguing in part from his own lesion experiments, Flourens concluded that cognitive capacities were not differentially localized in the cerebral cortex; rather, the cortex was a unitary organ. He cited his finding that, to the extent that a lesion compromised one mental capacity, for example, perception, so too to the same extent would reason, memory, judgment, will, and so forth be compromised. For Flourens, this pointed to a Cartesian view that the mechanistic program of decomposing a system into parts with distinctive functions ends at the mind. It is noteworthy that

[2] In fact, identifying mental abilities with the brain was one of the few features of Gall's views about which Flourens had anything positive to say, although he emphasizes that Gall could note claim propriety over the view: "the proposition that the brain is the exclusive seat of the soul is not a new proposition, and hence does not originate with Gall. It belonged to science before it appeared in his Doctrine. The merit of Gall, and it is by no means a slender merit, consists in his having understood better than any of his predecessors the whole of its importance, and in having devoted himself to its demonstration. It existed in science before Gall appeared—it may be said to reign there ever since his appearance" (Flourens, 1846, pp. 27-28).

Flourens dedicated his *Examen de la phrenology* to Descartes: “I frequently quote Descartes: I even go further; for I dedicate my work to his memory. I am writing in opposition to a bad philosophy, while I am endeavouring to recall a sound one.” (1846, p. xiv)

As the 19th century progressed, an alternative tradition of localizing cognitive processes in the brain took root. But the guiding conception of cognitive activities was very different than Gall’s phrenology, drawing inspiration not, as Gall did, from differences between individuals (as well as between species) in mental activities, but from the associationist tradition arising from John Locke. Although most 17th and18th century associationists such as Locke and David Hartley rejected any attempt to link associationist psychology to the brain, an entrée for doing so was provided by Charles Bell and François Magendie’s discovery in the 1820s that the posterior spinal nerves are sensory while the anterior nerves are motor. The fact that the sensory inputs arrived at the brain via a different pathway than that carrying the specification of motor activity generated by their associations} suggested that the intervening brain constituted a mechanism for making associations. This opening was identified by Alexander Bain, who made his objectives clear in a letter to John Stuart Mill in 1851:

> I have been closely engaged on my Psychology, ever since I came here. I have just finished rough drafting the first division of the synthetic half of the work, that, namely, which includes the Sensations, Appetites and Instincts. All through this portion I keep up a constant reference to the material structure of the parts concerned, it being my purpose to exhaust in this division the physiological basis of mental phenomena…. And although I neither can, nor at present desire to carry Anatomical explanation into the Intellect, I think at the state of the previous part of the subject will enable Intellect and Emotion to be treated to great advantage and in a manner altogether different from anything that has hitherto appeared. (quoted in Young, 1970, p. 103)

Despite Bain’s stated objective, his own publications failed to advance the links between the associationist tradition and sensory and motor processing in the brain.[3] His students, especially David Ferrier and John Hughlings Jackson, however, pursued precisely that connection by using weak electrical currents to probe the brain and by analyzing deficits resulting from brain injury. Jackson, for example, comments that,

> To Prof. Bain I owe much. From him I derived the notion that the anatomical substrata of words are motor (articulatory) processes. (This, I must mention, is a much more limited view than he takes.) This hypothesis has been of very great importance to me, not only specially because it gives the best anatomico-physiological explanation of the phenomena of Aphasia *when all varieties of this affection are taken into consideration,* but because it helped me very much in endeavouring to show that the 'organ of mind' contains processes representing movements, and that, therefore, there was nothing unreasonable in supposing that excessive discharge of convolutions should produce that clotted mass of movements which we call spasm (Jackson, 1931, pp. 167-8).

Ferrier and Jackson were not the first to develop a link between a type of mental process and the brain. In 1861, Paul Broca established that an area in the frontal cortex was involved in speech.

[3] Although Bain does not pursue the physiological component, he clearly construes the mind as a mechanism: “The science of mind, properly so called, unfolds the mechanism of our common mental constitutions. Adverting but slightly in the first instance to the differences between one man and another, it endeavours to give a full account of the internal mechanism that we all possess alike—of the sensations and emotions, intellectual faculties and volitions, of which we are every one of us conscious” (Bain, 1861, p. 29).

Broca made this connection working with a patient, Monsieur Leborgne, who lost the capacity for articulate speech. (Leborgne is better known by his pseudonym in the research literature, *Tan* –one of the few words he uttered.) After Leborgne's death, Broca conducted an autopsy, and even though the brain damage was by then massive, Broca argued that it began in the frontal area that came to bear his name. Like Gall, Broca approached mental capacities from a faculty perspective, but subsequent work on language deficits by Carl Wernicke (1874) instead adopted an associationist perspective. Wernicke construed the cortex as realizing associations between sensory and motor areas, with particular types of associations realized in their own distinctive brain regions. In Wernicke's model of reading, acoustic or visual images of words were connected to motor images that controlled either speech or manual action.

In the early 20th century even the degree of localization Wernicke endorsed was challenged by researchers who adopted a very holistic conception of associations. Such an anti-localizationist view is exemplified by Karl Lashley (1950; 1948), who argued, much in the spirit of Flourens, that beyond the primary sensory and motor projection areas, cortex was non-specific and acted in a holistic manner to implement associations between sensory and motor areas. He coined the term *association cortex*, which was in common use in neuroscience through the mid-20th century. This changed in the 1960s to 1980s, when researchers adopting a localization perspective were able to make dramatic progress in showing that extensive brain areas anterior to primary visual cortex were involved in visual processing. By correlating activity in different areas with different kinds of stimulus characteristics, they were able to identify each area's specialization. Thus, areas that Lashley had identified as general association areas gradually became identified with processing of specific types of visual information (see (Bechtel, 2001b; van Essen & Gallant, 1994). We will return to this developing mechanistic understanding of how neural processes figure in mental activity in a subsequent section.

Associationism was rooted in more than one field (epistemology and psychology) and also influenced more than one field. In addition to its influence on neuroscience, associationism contributed to the rise of behaviorism within psychology in the United States in the early 20th century. Commitment to a positivistic philosophy of science led behaviorists to be suspicious of any appeals to psychological processes occurring in the head that could not be objectively observed. In particular, James Watson (1913) and subsequent behaviorists were skeptical of the proposals regarding mental processing advanced by Edward Titchner (1907) and his followers, who used introspection as their guide to mental operations. In place of appeals to mental processes, behaviorists sought laws relating behavior to objectively observable variables—stimuli in S-R psychology, reinforcers in Skinner's accounts of operant conditioning. Behaviorists construed organisms, including humans, as learning mechanisms, and the strictest of them limited their laws to the regularities in input-output relationships that could be observed when these mechanisms functioned. They did not deny that there were internal processes occurring in the head, but rather, denied that psychology could or needed to provide an account of these processes. Such was the task of physiology. Moreover, behaviorists invoked Hempel's theoretician's dilemma (1958) to argue that if intervening processes were caused by external variables and themselves caused behavior, one could develop laws adequate to account for all behavior purely in terms of the casual external variables.

By the mid-20th century, Descartes' mechanical philosophy was far more influential than his dualism, but the working conception of mechanism was still quite impoverished. In particular, most conceptions of mechanisms focused on sequential operations. Within physiology, investigators working on mechanisms within living systems had come to recognize that the component processes were often organized in cycles, not linearly, and theorists such as Claude Bernard (1865) and Walter Cannon (1929) had begun to appreciate the significance of more complex modes of organization for physiological regulation. Except for the cybernetic movement which flourished especially in the period 1945-1955 (Wiener, 1948), theorists continued to think in terms of relatively simply, linearly organized machines of the sort Descartes envisaged. But a new kind of machine—the digital computer with a random access internal memory—was capable of more complex patterns of behavior. It quickly supplanted the Cartesian mechanisms that had occupied traditional thinking about psychological phenomena.

3. Information processing machines and their application to psychology

Three hundred years after Descartes, linguists began to confront a version of the problem that led him to argue against mechanism. Starting with the efforts of Ferdinand de Saussure in the late 19th century, structural linguists developed useful proposals regarding the basic components of language. To the extent they concerned themselves with how those components were combined, however, they found available mechanisms inadequate. In the 1950s Noam Chomsky proposed new mechanisms and applied automata theory to the task of characterizing the power and limitations of different kinds of grammars. (To do this, he had to view a grammar as a kind of automaton specialized to the task of generating sentences.) The weakest kind of automaton Chomaky considered was a a *finite state device*, in which the generation of a sentence would involve a sequential transition from state to state. For example, the initial state might offer a choice among nouns with which to begin a sentence. Depending upon which noun was chosen, a specific set of choices would open up for the next word—perhaps a set of verbs. Transitions from state to state would continue until a complete sentence had been generated in this way. Behaviorist accounts of language tended to be of this type.

In arguments reminiscent of Descartes, Noam Chomsky (1965) contended that finite state grammars cannot adequately characterize natural languages An example of the problem for a finite state automaton is that a natural language allows a potentially unlimited number of embedded clauses to intervene between a noun and verb that need to agree in number. There are ways to build in agreement across clauses, but the more clauses, the more unwieldy becomes the grammar, and there is no way to get agreement across an indefinitely large number of such clauses. Chomsky proposed that natural languages required grammars utilizing phrase-structure rules (which build tree structures) and transformational rules (which alter the tree structures). He also argued that a transformational grammar was equivalent in power to a Turing machine, a more complex sort of mechanism than Descartes could have anticipated.

A Turing machine is an abstractly characterized device (an automaton) proposed by Alan Turing. Prior to the actual construction of electronic computing machines in the 1940s, the term 'computer' referred to humans whose occupation was to perform complex calculations by hand. In advancing his characterization of a computing device, Turing (1936; see also Post, 1936) drew upon procedures executed by these human computers. Thus, in a Turing machine, a finite state device is coupled to a potentially infinite memory in the form of a tape on which symbols (typically just 0 and 1) are written. (Thus, the tape provides a memory that a finite state device lacks.) The finite state device has a read head that can read the symbol on one square of the tape and then may replace it by writing a different symbol or may move left or right one square. Which operation it performs depends on which of a finite number of states the device is in and which symbol it reads from the tape. Although such a device does not sound particularly impressive, Turing demonstrated that, theoretically, for each computable function there exits a Turing machine that can compute it. Moreover, he established that, by encoding the description of each specific Turing machine on a tape, it is possible to devise a universal Turing machine that can simulate any given Turing machine and so compute any computable function.

Construction of actual computing devices began during World War II, although the first to be completed—ENIAC (Electronic Numerical Integrator and Calculator)—was not operational until 1946. John von Neumann designed the basic architecture that still bears his name while ENIAC was under development, but his crucial idea of a stored program was not implemented until ENIAC's successor, EDVAC (Electronic Discrete Variable Computer). By the time EDVAC itself was fully operational in 1952, the first commercially produced computer, UNIVAC I (Universal Automatic Computer) had been delivered to the Census Bureau, and the computer revolution was underway.

Although primitive and slow by contemporary standards, in their day these earliest computers were impressive in their speed of computation. But what was more theoretically significant for psychology than their speed was that they could be viewed as symbol manipulation devices. This inspired researchers to ask whether the same devices might perform other cognitive activities that were generally taken to require thinking and intelligence. While much of the popular attention focused on attempts to create programs that could play well-defined games such as chess, pioneers such as Alan Newell and Herb Simon (1972) were enticed by the idea of mechanizing human problem solving. Newell and Simon's approach paralleled Turing's original work insofar as they drew upon the procedures that they believed humans explicitly follow in solving problems. Accordingly, one of their methods was to collect protocols by asking subjects to continuously describe what they were thinking while solving a problem, and then to devise programs that would employ similar procedures.

Newell and Simon construed themselves as making contributions to both computer science and psychology. Within computer science, the project they pioneered was designated *artificial intelligence*. But within psychology their work represented just one strand of the development of the tradition in cognitive psychology that construed the mind as an information processing mechanism. An independent thread derived from the mathematical theory of information. In the late 1930s, Claude Shannon—in a master's thesis entitled *A Symbolic Analysis of Relay and Switching Circuits*—employed Boolean operations to analyze and optimize digital circuits that would later be used in computers. He then took a position at Bell Laboratories, focusing on the transmission of information over channels (such as a phone line). In the course of that research, Shannon (1948; Shannon & Weaver, 1949) introduced the concept of a bit (binary unit) as the basic unit of information, and characterized the information capacity of a channel in terms of the ability of a recipient at the end of a channel to differentiate the state at the source of the channel.

A particularly influential consequence of this research in psychology was that Shannon's analysis of redundancy in a signal, which resulted when a given item in a signal would constrain the possibilities for another item (as the sequence of letters in 'mailbo' in an English text constrains the next letter). This result provided a basis for George Miller, who had done his dissertation research during the war on the capacity to jam speech signals, to demonstrate that certain messages were harder to jam than others. Miller and Selfridge (1950) further developed applications of information theory in a list learning experiment, explaining that the more closely word lists resembled English sentences (i.e., the greater their redundancy), the more words a subject could remember.

As we have noted, until this time American psychology had been dominated for forty years by behaviorist learning theory, which rejected attempts to explain behavior in terms of internal mental processes proposed on the basis of introspection. Information theory and the development of the computer provided a basis for thinking about internal processes in a much more constrained manner than introspectionism had offered. Very precise models of internal processes could be proposed and their predictions tested against observable behavioral data. This new movement within American psychology, known as information processing theory, changed the landscape of psychology.

Miller was one of most influential researchers to develop information processing models of cognition. His 1956 paper, "The magical number seven, plus or minus two: some limits on our capacity for processing information" became a classic. For a variety of activities, such as remembering distinct items for a short period, distinguishing phonemes from one another, and making absolute distinctions amongst items, Miller showed that significant changes in processing occurred when more than a few items (7 +/- 2) were involved. In addition to using behavioral data to establish limits on cognitive processing mechanisms in this and earlier work, Miller also collaborated with Eugene Galanter and Karl Pribram (1960) to provide one of the first suggestions of the structure of an information processing mechanism. The investigators' goal was to develop a framework in which to account for mental activities such as the execution of a plan. One challenge in executing a plan is to know when it is appropriate to initiate a behavior and when to end it. Miller, Galanter, and Pribram proposed a basic cognitive mechanism they called a TOTE unit: Test-Operate-Test-Exit. The idea is that when a test operation indicates that the conditions for an operation are met, it is performed, and continues to be performed until the conditions for it are no longer satisfied. One of the particular powerful features of the proposed scheme was that TOTE units could be embedded within other TOTE units (see figure 1). One of the applications they envisioned (but did not work out in detail) was a realization of Chomsky's phrase structure and transformational grammars.

<Insert Figure 1 here>

As noted above, information processing psychologists primarily appeal to behavioral data to constrain their models of cognitive mechanisms. Error patterns and reaction time are two of the key behavioral measures invoked. For example, Saul Sternberg (1966) used reaction time data to choose between candidate mechanisms for human memory retrieval. Measuring the time subjects required to determine whether a given digit was on a just-memorized list, he found a linear relationship between the number of items on the list and how long it took to respond affirmatively or negatively to the test item. This ruled out a process of parallel access to the whole list in memory. More surprisingly, positive responses took as long as negative responses. If subjects performed a self-terminating search, stopping once they had found an item, positive responses should have taken less time. Since this was not the case, Sternberg posited a memory retrieval mechanism that incorporated exhaustive search in its design.

Most psychologists working within the information processing approach to cognition believed that the information processing mechanisms they were investigating were realized in the brain. However, they lacked tools for making the appropriate connections to brain processing. Although techniques for studying neurons using micro-electrodes have been widely used in non-human animal studies since the 1940s (recording techniques) and even the 1870s (stimulation techniques), and have figured prominently in systems-level neuroscience research on such processes as visual perception during this same period (see (Bechtel, 2001b), ethical considerations limit the use of such techniques in humans. For human studies, neuropsychologists have obtained extensive behavioral data on individuals with naturally occurring lesions as a means of identifying the brain components involved in different cognitive mechanisms. However, until they began to collaborate with cognitive psychologists, neuropsychologists were limited in their capacity to relate the brain regions they identified to actual mechanisms responsible for generating behavior (see (Feinberg & Farah, 2000). Scalp recordings of electrical activity in the brain enabled researchers to trace some of the temporal features of information processing—especially when such recordings were time-synced to stimuli in order to measure *evoked response potentials*—but offered little ability to localize the brain processes spatially. Thus, it was not until the advent of functional neuroimaging with PET and MRI that researchers interested in the mechanisms underlying cognitive activity could link the component operations to brain processes (see (Posner & Raichle, 1994). The introduction of neuroimaging coincided with the flowering of cognitive neuroscience as a research field involving the collaborative efforts of psychologists and neuroscientists in localizing information processing mechanisms (see (Bechtel, 2001a).

Practioners of both neuroscience and psychological science are now well-embarked on the project of explaining mental activity mechanistically and are effectively integrating their investigations. Descartes' concerns about the inadequacy of mechanism to explain cognition have been assuaged, and the framework of information processing has provided a powerful vehicle for developing models of mechanisms responsible for cognitive behavior—one that is continuously exemplified in the statements of contemporary researchers:

> Nervous systems are information-processing machines, and in order to understand how they enable an organism to learn and remember, to see and problem-solve, to care for the young and recognize danger, it is essential to understand the machine itself, both at the level of the basic elements that make up the machine and at the level of organization of elements. (Churchland 1986, p. 36)

With this overview of the historical route by which psychology became mechanistic, we now turn to a more analytical discussion of what a mechanism is and how a mechanical philosophy proffers a new perspective on some traditional issues in the philosophy of psychology.

4. Contemporary conceptions of mechanism

Some contemporary philosophers of science have regarded the increased sophistication of mechanistic approaches as concomitant with our best ways of understanding how reality is discovered and characterized; Salmon (1984, p. 260), for one, proclaims, "The underlying causal mechanisms hold the key to our understanding the world." Yet, while empirical researchers frequently refer to and incessantly search for a massive array of particular mechanisms, until recently philosophers have shown little interest in what a mechanism is and how mechanisms might figure in explanations. Philosophers interested in the biological sciences (especially physiology, cell and molecular biology, and neurobiology)—sciences in which more traditional accounts of explanation that appeal to laws seem to yield little traction—have led the way in trying to articulate a satisfactory conception of mechanism and mechanistic explanation. William Bechtel and Robert Richardson (1993) were among the first to offer such an account. They suggested that, "A machine is a composite of interrelated parts, each performing its own functions, that are combined in such a way that each contributes to producing a behavior of the system. A mechanistic explanation identifies these parts and their organization, showing how the behavior of the machine is a consequence of the parts and their organization" (1993, p. 17). They developed and explored the consequences of this mechanistic approach by examining research in biochemical energetics, molecular genetics, and the cognitive neuroscience of memory.

In the decade since, variations on this conception of mechanism and mechanistic explanation have been advanced by several philosophers. Stuart Glennan, for example, endorses the above conception, but explicitly attempts to make room for the possibility and import of laws: "A mechanism underlying a behavior is a complex system which produces that behavior by the interaction of a number of parts according to direct causal laws" (1996, p. 52). The definition proposed by Peter Machamer, Lindley Darden, and Carl Craver—namely, that mechanisms are "entities and activities organized such that they are productive of regular changes from start or set-up conditions to finish or termination conditions"—emphasizes the dual import of entities and activities, structure and function. They quip: "There are no activities without entities, and entities do not do anything without activities" (Machamer, Darden, & Craver, 2000, p. 3). Additionally, Jim Woodward (2002, pp. 374-5) characterizes mechanisms as modular systems whose independent parts are subject to manipulation and control and behave according to counterfactual-supporting regularities that are invariant under interventions.

While there certainly are subtle differences between these various conceptions of mechanisms, their overall coherence reflects a growing consensus on the proper formulation. Indeed, the intersection of these and other similar conceptions is the view that mechanisms are composite, hierarchical systems whose activity produces a target phenomenon _ and the regularities associated with it. For example, such a system might generate ATP from glucose, recall an episodic memory, or take different routes to a landmark. Systems may perform their activities in isolation or in transaction with other mechanisms, and they may be active at one time while inactive at other times.[4] As composite systems, mechanisms are composed of *component parts*

[4] Note that the inactivity of mechanisms does not render them explanatorily superfluous, as a mechanism's inactivity or inhibition may be just as important a factor in bringing about the phenomena and its associated regularities as another mechanism's activity.

and their properties.[5] Each component part performs some *operation* and interacts with other parts of the mechanism (often by acting on products of the operation of those parts or producing products that they will act on), such that the coordinated operation of the parts constitutes the activity of the mechanism. The tight causal interactions among the operating component parts are what produce the target phenomenon _ and, as first articulated by Kauffman (1971), component parts and their operations are individuated according to the causal influence that they exert on mechanistic activity.

In addition to the structural and temporal properties of the component parts and the functional properties of their operations, *organization* is critical to a mechanism. The relevant organizational and architectural properties—including location, orientation, polarity, cardinality and ordinality, connection, frequency and duration—enable the parts to work together effectively and perform the activity of the mechanism. The imposition of organization on components often produces more or less stable assemblies whose architecture then fixes the activities that can be performed. Elucidating organizational properties is crucial for any mechanistic explanation, but it is especially important when the organization involves non-linearity, cyclic processes, etc. As the results of complexity theory have demonstrated, surprising behavior often results from such modes of organization.

A mechanism's spatiotemporal organization is also, in part, what makes a composite system *hierarchical*. It has sometimes proven fruitful to conceive of organization in terms of levels, such that investigation and explanation of a mechanism's activity is understood as taking place at a higher level than an investigation or explanation of the constituency that composes it (see §6 below). At a higher level still, that very same mechanism may be a component part or subsystem in another, larger composite system. Its mechanistic activity would then constitute the operation of a component part).[6] Consequently, mechanisms—as composite, hierarchical systems—are multi-level; but just what these levels are has been a point of contention for mechanists.

5. Mechanistic explanation

5.1 Laws and the ontic/epistemic distinction

For much of the 20th century, philosophers eschewed talk of both causation and mechanism, instead construing explanation nomologically in terms of laws. According to the traditional deductive-nomological (D-N) model, explanation takes the form of a deductive argument in which an event description is logically deduced from a set of statements of general laws in conjunction with a set of initial conditions (Hempel & Oppenheim, 1948). Salmon (1984) was

[5] There are some entities that are occasionally referred to as 'simple machines', such as the wedge and wheel. Insofar as these entities are actually active (as determined by their shape alone) and lack operative parts, they may be exceptions to such a condition; yet, the mechanisms of interest to biology and psychology, though, are all necessarily composite systems.

[6] Such claims need not commit mechanists to the view that mechanisms stand in mereological relations *ad infinitum*. Whether a given mechanism is a component of a higher-level mechanism depends upon whether it is part of an organized system at the higher level. Going the other direction, mechanists need not commit to where or whether mechanistic explanation 'bottoms out'. Whether a researcher pursues explanations that require ascending to higher levels vs. descending to lower ones, depends upon his or her explanatory goals.

one of the first to dissent from the hegemony of law-based views of explanation.[7] Although Salmon spoke of 'causal/mechanical explanations', his principal focus was on causation rather than mechanisms *per se*. Salmon's characterization of the differences between nomological and causal/mechanical explanation continues to influence contemporary discussions—but also confuses them in an important respect.

Salmon characterized his causal/mechanical account of explanation as *ontic* and nomological accounts as *epistemic*. The motivation for this distinction is that, traditionally, nomological accounts have identified explanation with *argumentation*. As such, the explanandum—a set of statements about the phenomenon _ to be explained—is a logical consequence of the explanans—a set of statements about the relevant antecedent conditions whereupon _ is produced together with a generalization (law) stating that, when those conditions prevail, _ occurs. Salmon objects to rendering explanation in such terms: "An epistemic conception takes scientific explanations to be arguments…, but explanations are not the sorts of things that can be entirely explicated in semantical terms" (Salmon, 1984, pp. 239, 273). Instead, explaining a given explanandum "obviously involves the exhibition of causal mechanisms" leading to the occurrence of that explanandum (p. 268). While Salmon does note that explanations of an event take the form of fitting that event into a pattern of regularities described by causal laws, by "exhibiting it as occupying its place in the discernable patterns of the world" (pp. 17–18), he is quick to add that adequate explanations must track the mechanisms responsible for events.

> To provide an explanation of a particular event is to identify the cause and, in many cases at least, to exhibit the causal relation between this event and the event-to-be-explained…. Causal processes, causal interactions, and causal laws provide the mechanisms by which the world works; to understand *why* these things happen, we need to see *how* they are produced by these mechanisms. (pp. 121–4, 132)

So, it is because causal/mechanical explanations track mechanisms in the world that Salmon construes his account as ontic. Peter Railton (1978) articulates a similar shift in conception, suggesting that whatever lawlike generalizations range over regularities, they must be supplemented with information about the mechanisms producing those regularities:

> The goal of understanding the world is a theoretical goal, and if the world is a machine—a vast arrangement of nomic connections—then our theory ought to give us some insight into the structure and workings of the mechanism, above and beyond the capability of predicting and controlling its outcomes…. Knowing enough to subsume an event under the right kind of laws is not, therefore, tantamount to knowing the how or why of it. What is being urged is that D-N explanations making use of true, general, causal laws may legitimately be regarded as unsatisfactory unless we can back them up with an account of the mechanism(s) at work. (Railton, 1978, p. 208)

Machamer *et al.* take a more aggressive approach than either Railton or Salmon, upholding this emphasis on tracking mechanisms as the measure of explanatory adequacy while depreciating the import of lawlike generalizations with wide scope and global applicability.

[7] An earlier, but unfortunately less well-known, conception of mechanistic explanation is found in Harré's (1963) essay—some twenty years prior to Salmon's seminal work. Harré does not analytically define the concept of mechanism, but his preliminary framework partially anticipates the recasting of mechanistic explanation in terms of models as advocated in §5 below.

> We should not be tempted to follow Hume and later logical empiricists into thinking that the intelligibility of activities (or mechanisms) is reducible to their regularity. Descriptions of mechanisms render the end stage intelligible by showing how it is produced by bottom out entities and activities. To explain is not merely to redescribe one regularity as a series of several. Rather, explanation involves revealing this productive relation. It is not the regularities that explain but the activities that sustain the regularities. There is no logical story to be told…. (Machamer *et al.* 2000, pp. 21–2)

The remarks of Railton, Salmon, and Machamer *et al.* point to typical motivations for this shift in conceptions in the following two respects. On one hand, the abandonment of epistemic conceptions is negatively motivated by sundry problems with the principles (e.g., subsumption of psychological phenomena under laws of nature, explanatory unification) and models (e.g., D-N model, classical reduction) of epistemic conceptions. Many of these problems are conceptual snags leftover from failed attempts to carry out various logical empiricist and positivist programs. The negative motivation for this abandonment can be understood, in part, as an attempt to distance mechanistic approaches from such programs. On the other hand, this abandonment is also positively motivated by the idea that an adequate conception of mechanistic explanation should emphasize the production of local, individual phenomena by the activity of the composite system.

Accordingly, on most ontic conceptions, an explanation counts as mechanistic for a phenomenon _ only when it identifies a hierarchical, composite system whose activities produce _ and whose component parts are organized in certain ways and perform certain operations so as to be constitutive of the systemic activities producing _—thereby situating _ among a nexus of natural regularities (Salmon, 1984, pp. 260, 268; Bechtel, 2002, p. 232; Bechtel & Richardson, 1993, p. 17-18; Glennan, 1996, p. 61; Railton, 1998, p. 752; Machamer et al., 2000, pp. 3, 22; Woodward, 2002, p. 373). A common way of framing this conception is in terms of giving an answer to a how-question. Accordingly, Thagard (2003) writes,

> [T]he primary explanations in biochemistry answer how-questions rather than why-questions. How questions… are best answered by specifying one or more mechanisms understood as organized entities and activities. … Thus answering a how-question is not a matter of assembling discrete arguments that can provide the answer to individual why-questions, but rather requires specification of a complex mechanism consisting of many parts and interconnections. (Thagard, 2003, p. 251; Cummins, 2000)

To grasp the importance of mechanists' appeal to the mechanisms themselves as the explanation, consider two further examples that have figured prominently in the recent philosophical literature: electro-chemical synaptic transmission (Machamer et al., 2000, p. 8-13; Bickle, 2003, pp. 62-50) and stereoptic color and depth perception in state-space (Churchland & Sejnowski, 1992). To explain the phenomenon of electro-chemical synaptic transmission, one *demonstrates* or *reveals* the operations and organization at the level of component parts, and the activities performed at the level of the composite whole—e.g., the synthesis, transport, and vesicle storage of agonist/antagonist neurotransmitters and neuromodulators, their release and diffusion across the synaptic cleft, the process of binding with presynaptic autoreceptors and postsynaptic receptors, reuptake, depolarization, etc. Or again, the phenomenon of stereoptic color and depth perception is explained by demonstrating or revealing the internal structure, activities, and organization of the component parts that both constitute the visual system, and bring about the perceptual activities that produce that phenomenon.

5.2 Deconstructing the distinction between epistemic and ontic conceptions of mechanistic explanation

The current literature is filled with explications of the basic ontic conception whereby the explanans just is a mechanism and the explanandum is the phenomenon produced by the activity of that mechanism. As such, all references to non-ontic entities (e.g., propositions, inference, truth or reference) are omitted; the component parts and their operations and organization are themselves what do the explaining; and they do so in virtue of simply *being* the explanans. Accordingly, explananda are explained *through the mechanisms* that produce them. Hence, in discussing long-term potentiation, Machamer *et al.* (2000) write, "It is through these activities of these entities that we understand how depolarization occurs" (p. 13). Similarly, in discussing biochemical pathways, Thagard (2003) writes, "What explains are not regularities, but the activities that sustain regularities. Thus biochemical pathways explain by showing how changes within a cell take place as the result of the chemical activities of the molecules that constitute the cell" (p. 238).[8]

Now, although there is a clear contrast between invoking laws versus mechanisms in explanation, the epistemic/ontic distinction turns out to be an unfortunate way to draw the contrast. One reason is that, while mechanists desire an account that pitches the explanation of psychological phenomena *purely* in terms of the mechanisms that produce them—thereby shedding the overbearing dependence on logical, semantic, grammatical, and inferential concerns—they simultaneously help themselves to the resources of abandoned epistemic conceptions.

To see that epistemic conceptions tend to be kicked out the front door whilst being ushered in the back, one should ask, do the component parts and their operations and organization figure in our understanding of how and why depolarization occurs? Well, yes, in a flat-footed sense—without any of these things to implicate, mechanistic explanations would be without content. But, in another sense, what our understanding *literally* proceeds 'through' is a network of linguistically- or graphically-expressed operations on representations. After all, scientists typically explain by *marshalling a narrative*—i.e., telling a story about why the explanandum is a consequence (material or otherwise) of antecedent conditions. The view that mechanistic explanations are given by indicating mechanisms (qua explanans) must be understood as metonymic—i.e., as an emblematic stand-in for a richer, more accurate explication of what mechanistic explanation actually consists in. Presynaptic autoreceptors, sodium potassium pumps, and ligand-gated ion channels are simply inapposite candidates for an explanans; the relevant parts, operations, and organization minimally need to be captured and codified in a structural or functional *representation* of some sort. And unless these suggestions are understood as metonymical, accounts of mechanistic explanation will misconstrue explanatory practice and pervert the

[8] To be fair, Thagard (2003, pp. 237, 251-2) nicely raises the need for cognitive psychological research on mental representations of mechanisms, and advances 'a cognitive view of theories' whereby representations and operations thereon serve as a vehicle of mechanistic explanation. However, his reliance on the ontic/epistemic distinction encumbers the further development of these ideas.

meaning of the term. Giving an 'explanation,' after all, refers to a practice that cognizers engage in to make the world more intelligible; the non-cognizant world does not itself so engage. One way to appreciate this point is to recognize that mechanisms are active or inactive whether or not anyone appeals to them in an explanation. Their mere existence does not suffice for explanation; a phenomenon is produced by the mechanism, but it is not explained until a cognizer contributes his or her explanatory labor.[9]

To see more clearly what is at stake, consider Glennan's (1996) example of the mechanics of a toilet tank. Suppose that you are among the millions of people throughout the world who have had little-to-no experience with the recurrent flushing of a toilet and the regulation of water-level in its tank, and have no knowledge of indoor plumbing more generally. If someone desires to explain to you how and why toilets are able to do those activities, then merely depressing the lever that initiates flushing, or taking off the lid and letting you see the internal goings-on of the tank, would be entirely deficient. Similarly, if the phenomena of salutatory conduction or memory consolidation were cognitively abstruse, something more would be needed in addition to merely *presenting* or setting out the requisite mechanisms for observation—for even if all aspects of a mechanism are observable, there is no guarantee that one will have the "instant flash of insight" that accompanies self-explanation of how the phenomenon is produced.

A further reason for why the epistemic/ontic distinction poorly captures what is distinctive of mechanistic explanations is that mechanists consistently use certain concepts (viz., 'demonstrate', 'reveal', 'lay bare', 'indicate', 'exhibit', 'display') which are left as semantic or conceptual primitives—they are not clarified, characterized, or given meaning over and above the already intuitive 'explain'. Uninformative treatments of these cognates are then themselves used to articulate the nature of mechanistic explanation. These concepts, however, again conceal an epistemic perspective, one that needs to be elucidated if we are to understand the sense in which mechanistic explanations are explanatory.

For instance, on some ontic conceptions, it is suggested that explanation consists in an indicative act—literally, a demonstrative gesturing or pointing at the component parts' operations and organization that together constitute the mechanistic activity producing the explanandum.[10] On the basis of such a construal, Scriven, for instance, levels the following complaint against epistemic conceptions that construe explanation as a type of argumentation: "Hempel's models could not accommodate the case in which one 'explains' by gestures to a Yugoslav garage mechanic what is wrong with a car" (quoted in Salmon 1984, p. 10). But what work is a gesture or an indication doing? If a gesture or indication consists in a relation between a cognizer and an explanans—i.e., the mechanism that is supposed to do the explaining—as ontic accounts seem to advocate, then the above problem with metonymy simply recurs rather than gets resolved: conditions of the world do not themselves explain, and no amount or complexity of indication (in this ontic sense) will make that the case. If, on the other hand, such acts are construed as bringing

[9] There is yet a further reason why the mechanism itself is not the appropriate vehicle of explanation. Many mechanistic explanations that are offered turn out to be wrong, either fundamentally or in details. If the mechanism itself sufficed for explanation, erroneous explanation would not be possible.

[10] While there are phenomena where it is possible to point to the requisite mechanisms, there are others where it is not clear what acts of indication could possibly amount to—e.g., Higgs-Boson particles, creativity, or riotous mob behavior. Since these cases involve appeal to mechanisms as well, mechanistic explanation must crucially involve something over and above the bare presentation of the mechanism itself.

it about that mechanistic activity is represented and understood, then an epistemic perspective is being surreptitiously implicated. To realize what is being concealed with references to gestures, or to ‘indicating’, ‘demonstrating’, ‘revealing’, ‘laying bare’, ‘exhibiting’, ‘displaying’, etc., notice that what the Yugoslav garage mechanic considers to be an explanation could only be what she or he finds cognitively salient about the situation. Indication, in Scriven’s sense, would only be explanatorily helpful if it were made against the background of large corpus of conceptual knowledge inquiry. And even if cognitive salience-conferring behaviors were sufficient for explanation, no such demonstrative gesturing or pointing could alone unequivocally specify what had been indicated, since myriad structures, functions, and organization would be consistent with any such gesture.

We conclude that construing mechanistic explanation ontically misplaces *the space of explanation*. Instead of understanding explanation as taking place within the “space of reasons”—i.e., as the argumentative expression of the reasons for the production of psychological phenomena, as operations on *re*presentations of mechanistic activity—ontic conceptions place explanation out in the space of mechanisms as they present themselves to us. This is tantamount to a category mistake.

5.3 Recasting a proper conception of mechanistic explanation

Having articulated why the view that nomological explanations are epistemic while mechanistic explanations are ontic is naïve, let’s consider what a proper construal of mechanistic explanation would entail.

While mechanists have made much of Salmon’s ontic/epistemic distinction, they actually vacillate between appealing to mechanisms themselves and identifying the explanans of mechanistic explanations with sets of descriptions of mechanisms. Hence, Machamer *et al.* (2000, p. 3; see also Craver, 2001, p. 68) aver that, “Giving a description of mechanism for a phenomenon is to explain that phenomena and its production,” and Glennan (1996, p. 61; see also 2002, p. 347) writes, “A description of the internal structure of the mechanism explains [its] behavior.”[11] This vacillation reflects a recognition that a necessary condition on mechanistic explanation is that the structure, function, and organization of mechanisms needs to be captured and codified representationally.

The importance of the representational aspect of mechanistic explanation can be acknowledged, though, and the contrast with the nomological account maintained—what is represented are mechanisms, not sets of laws. We can also recognize that linguistic representations do not exhaust the range of possible representations. In many cases, graphical representations such as diagrams or figures are far more effective representations of mechanisms than linguistic descriptions. Because of their ability to represent objects in two or three dimensions, graphical representations are able to capture important elements of the spatial organization of a mechanism. Since one can also reserve a dimension for time, or use arrows to represent

[11] Such remarks might seem to suggest that descriptions of mechanisms are not just coincident with, or derivative from, explanations—they *are* explanations. But explanations are not merely lists of descriptions of mechanisms or sets thereof; they include *inferential* operations on them. To grasp this, simply consider the semantics of the explanatory connective ‘because’, or what it is that arrows in box-and-arrow diagrams represent).

succession relations, graphical representations can also capture important aspects of the temporal organization of mechanisms. Whereas linguistic representations can only capture one component at a time, graphical representations can identify multiple components and their relations.

Just as mechanistic explanation extends representation beyond the linguistic, it also expands the way that representations are related to phenomena. Instead of deductive argument, one must understand how the mechanism produces the target phenomenon. One strategy is to use imagination to put one's representation of the mechanism into motion so as to *see* how that phenomenon is generated. This is not an area in which there has yet been much work by philosophers. Cognitive psychologists have begun to make some headway in characterizing the processes by which scientists and science students develop the ability to simulate mentally the behavior of simple mechanisms such as springs (Hegarty, 2002; Clement, 2003). There is related work by philosophers (Nersessian, 1999, 2002) and psychologists (Ippolito & Tweney, 1995) on simulating experiments which may be useful to developing a more robust account of how cognitive agents come to understand the relation between the components of a mechanism and what it does. (The idea that what a person is doing is simulating an event points to a link with the activity of computer simulation in psychology. Like mental simulations, computer simulations show that some phenomenon is what would result from the conditions specified in the simulation.)

Mechanistic explanation is an epistemic practice. There are norms governing such a practice—namely, that explaining a target phenomenon requires an understanding of the systemic activities that locally produce it, which in turn requires revealing the mechanism's internal structure, function, and organization. The 'understanding', though, is constituted by both representations of mechanistic activity and inferential operations on those representations, and which, *a fortiori*, places the nature of explanation within the argumentative, epistemic 'space of reasons'.[12]

Having resituated mechanistic explanation within an epistemic context, we need to consider briefly two traditional epistemic concepts that arise in talk of explanation—models and laws. The concept of *model* is interspersed throughout the scientific literature, including in psychology, In general, the model is a structure in which a set of entities stand in specified relations to one another. The model stands in for the actual systems that researchers are trying to understand and is invoked in reasoning and theorizing about the actual system (see Harré, 1963). Beyond these commonalities, there are many disparate senses of model and ways in which models are used. In model theory, for example, a model is a set of entities, often abstract, that satisfy a set of axioms. Proponents of the structuralist and semantic views of theory (van Fraassen, 1989; Giere, 1988, 1999) construe a model as a set of abstract (nonphysical) objects that conform to the theories advanced in a science which then stand in isomorphic or approximately isomorphic relations to the actual objects in the world to which the theory is supposed to apply. In the context of mechanistic explanation, the objects in a model correspond to the parts of a mechanism, and their structure conforms, not to a theory, but rather to the mechanism's constituency and interactivity.

[12] The use of 'argumentative' need not be construed as to refer only to syntactic operations on propositions. Indeed, such narrow construals are part of the reason why epistemic conceptions have been dragged through the mud.

Mechanistic models may be abstract, or may be implemented physically. For example, an engineer may build a scale model to experiment with before building a device. Scientists may also build such models, but towards the goal of understanding the mechanism being modeled. The best known example is the physical model of DNA that Watson and Crick constructed in discovering its double helix structure. A more subtle example is that a model of a cognitive activity, such as natural language understanding, may take the form of a computer program. The program itself is an abstract mechanistic model, but implementing it on a particular computer gives it a physical realization in which the consequences of its design are more readily discovered.

Yet another sense of model arises from the fact that in the biological and behavioral sciences, researchers may select a particular model system (organism) on which to conduct research intended to apply to a much broader range of organisms or to humans. Mouse models of navigation and spatial memory are a prominent example in behavioral neuroscience. A model system brings out another important feature of models—they typically simplify the target mechanism, abstracting from features of the system that are not taken to be essential to the generation of the phenomenon.

Lastly, consider laws. Mechanists often emphasize the contrast between nomological explanations that give pride of place to laws and mechanistic explanations. As such, the motivation underlying the rejection of an epistemic conception can be understood not so much as an aversion to argumentatively representing the production of psychological phenomena as it is an aversion to doing so by deducing descriptions of phenomena from laws. Cummins (2000, pp. 118-22) nicely articulates one of the main reasons for rejecting the nomological conception of explanation: the explanation of psychological phenomena is not a matter of subsumption under law because psychological laws are simply 'effects', and effects are simply explananda—not explanans. He writes,

> In psychology, such laws…are almost always conceived of, and even called, effects. We have the Garcia effect, the spacing effect, the McGurk effect, and many, many more. Each of these is a fairly well-confirmed law or regularity (or set of them). But no one thinks that the McGurk effect explains the data it subsumes. No one not in the grip of the D-N model would suppose that one could *explain* why someone hears a consonant like the speaking mouth appears to make by appeal to the McGurk effect. That just *is* the McGurk effect. (Cummins 2000, p. 119)

Cummins correctly adduces that the practice of explanation by subsuming phenomena under laws is rare in psychology, and even when it is invoked, what is understood by 'law' tends to be a description of the target of (mechanistic) explanation.

It would, however, be a mistake to suggest that mechanists are simply opposed to the appeal to laws in explanations; on the contrary, they certainly include analyses of the significance of laws in their approaches, *where appropriate* (e.g., Salmon 1984; Glennan 1996, 2002; Hardcastle 1996). Bechtel & Richardson (1993, p. 232) write that, "The explanatory task begins and ends with models; we question the hegemony of laws in explanation, not their existence." Laws are sometimes needed to help characterize the regularities in the behavior of components of a mechanism and thus can play a supplementary role in mechanistic explanation. What does the major explanatory work is the identification of the components and their operations as well as the manner in which they are organized. This work is not performed by identifying laws.

6. Hierarchical mechanisms, levels of organization, and reduction

An issue that has long garnered the excitement and consternation of generations of philosophers is whether theories in the psychological sciences reduce to theories in the neurosciences. One of the distinctive features of the mechanistic approach is that it demands a fundamental reorientation of this issue of reduction and reductionism. Accordingly, in one sense, a mechanistic explanation is through and through reductionistic: it appeals to the component parts and their operations in explaining the activity of a mechanism. But in another sense, a mechanistic explanation is non-reductionistic: there is no sense in which explanations at a lower level replace, sequester, or refine the role of higher-level explanations, because mechanisms are hierarchical, multi-level structures that involve real and different activities being performed by the whole composite system and by its component parts. Rather than serving to reduce one level to another, mechanisms bridge levels. So, while reductive and mechanistic approaches can be closely aligned, they diverge in important respects. Before developing this contrast further, we first need to clarify the basic concept of a level.

6.1 Mechanistic levels of organization and analysis

Talk of 'levels' spans numerous disciplines—especially in the brain and neural sciences, cognitive science, and psychology—and it is hard to overstate the significance of the concept of a level in these disciplines. Just the same, levels-talk is virtually threadbare from overuse (Craver, forthcoming). One reason is that the various concepts of levels are rarely analyzed in any sustained, substantive detail despite there being a large litany of literature on the subject (for an attempt to rectify this problem, see Wilson & Craver, ch. XX). A second reason is that levels are ambiguously construed as both ontic levels of mechanistic organization and as epistemic levels of analysis.

A traditional concept suggests a neat hierarchical layering of entities into levels across phenomena. On this conception, scientific disciplines (viz., physics, chemistry, biology, psychology, sociology).can be viewed as distinguished in part by the level of phenomena in nature that are their target of study. This view is found in Oppenheim and Putnam's (1958) layer cake account of disciplines, and Simon's (1969) account of metaphysically discernible levels of organization. Appealing to a variety of evolutionary considerations, Simon argues that nature would have to build complex systems all at once—an implausible conclusion—were it not for the assembly of stable, semi-autonomous, modular parts. The use of assemblies and subassemblies to facilitate increasingly organized, complex systems allows for component parts to be differentially deployed and combined. It also means that impairment of a subassembly is less likely to be disruptive to the overall system. Simon also offers an explanation of the differential stability of assemblies at successive levels of organization: the bonding energies used to create structures are greatest at the lowest levels (e.g., with the atom) and weaker at higher levels (e.g., covalent bonds in macromolecules). This traditional construal of ontic levels of organization is further developed by Wimsatt (1976), who also argues that the most frequent causal interactions are among entities of the same scalar magnitude. Yet, Wimsatt notes that any hope of finding neatly delineated levels diminishes as one approaches entities the size of macroscopic objects, which interact across size scales; consequently, he introduces the concept of *perspectives* for "intriguingly quasi-subjective (or at least observer, technique or technology-relative) cuts on the phenomena characteristic of a system, which needn't be bound to given levels" (Wimsatt, 1994; see also Wimsatt, 1974, 1976).

Several philosophers have resisted any overtly realist construal of levels.[13] For instance, Hardcastle (1996, pp. 29-32) argues that the neatly divided, layer cake concept of levels is nothing more than a theoretical imposition, since what counts as a level can be fairly arbitrary and relative to a variety of factors (e.g., types of questions asked, methodology). Craver (2001, pp. 65–7) takes a similar stance in decrying such construals as typifying a problematic reification; in their stead, he recommends neutralizing some of the relevant metaphysical commitments by way of construing levels as 'perspectival' in the sense of involving different views on an entity's "activity in a hierarchically organized mechanism." He differentiates contextualized, isolated, and constitutive perspectives on a given component of a mechanism and on how these might be integrated in a complete account of a mechanism.

Yet a different construal of levels is more realist than those discussed just above, but not as global as those of Oppenheim and Putnam, Simon, and Wimsatt. On a given cycle of decomposing a mechanism it treats all relevant components as being at the same level (Craver & Bechtel, forthcoming; Craver, forthcoming). That is, the decomposition is local to the mechanism and no attempt is made at identifying the level each component would occupy in a global portrayal of levels across all of nature. For example, suppose a biological mechanism is being described in which sodium molecules cross a membrane. In a global portrayal the membrane would be at a higher level than the sodium molecules, since the membrane is itself composed of molecules. But in this more local construal of levels, the sodium molecules and the membrane are treated equally. Each is a component of the mechanism that is relevant to the activity in question. If an investigator pursues another cycle of decomposition, the components will themselves be analyzed into components at a lower level, but again this level is local to the analysis. Multiple cycles of analysis thus gives rise to a hierarchy of levels that is confined to a given mechanism. An investigator who had started by focusing on a different activity may have ended up with a different parsing of entities into levels. Levels on the mechanistic account are ontic in that they deal with real components and their operations, but they are perspectival in that they are defined with respect to a specific activity of the mechanism. (One advantage of eschewing global analysis into levels is that there is no need to answer such conundrums as whether a person's hippocampus is at the same level as a car's radiator.)

In order to stave off possible confusion, one should distinguish between the mechanistic concept of levels of components and the concept of epistemic levels of *analysis*. Perhaps the best known epistemic conception of levels in psychology and cognitive science is due to Marr (1982, p. 25), who distinguished three "different levels at which an information device must be understood before one can be said to have understood it completely":

[13] Perhaps the most radical arguments against the 'layered picture of the world' are those leveraged by Heil (2002), who ultimately suggests that talk of ontic levels of organization be dispensed with altogether.

> (i) the abstract computational theory of the device, in which the performance of the device is characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated,
> (ii) the choice of the representation of the input and output and the algorithm to be used to transform one into the other,
> (iii) the details of how the algorithm and representation are realized physically—the detailed computer architecture, so to speak.

The emphasis for Marr is on the different epistemic projects an investigator can perform and the kind of account required, not on part-whole relations within a mechanism. Thus, the characterization of the algorithm (ii) and of its physical realization (iii) may be characterizations of the same thing, and thus not span ontic levels. One might be focusing on the same component parts, yet describing them in terms of their spatiotemporal properties rather than in terms of the algorithm they implement. The mechanistic account, unlike Marr's, in mereological—it is not by describing something different but by decomposing it into its component parts and operations, that one reaches a lower level.

Unlike levels of analysis, mechanistic levels involve mereological relationships between the component parts and their operations and the mechanism itself. One important consequence of this account of levels is that component parts operating within a mechanism typically do different things than the composite system. For example, contrary to the standard identity claim that pain is C-fiber firing, the spiking of an individual A_ or the C-fiber in the anterior cingulate does not itself constitute the experience of pain. Pain is the activity of an overall pain mechanism, whereas firing is the activity of a single fiber within that mechanism

Merely indicating that the properties of component parts and their operations at one level of organization are distinct from those of the overall mechanism and its activity, however, is insufficient to capture an important feature often attributed to higher levels—namely, that "composite wholes are greater than the sum of their component parts." Capturing this feature requires that one take seriously the term *organization* in the phrase 'levels of organization', yet that term has not received detailed philosophical analysis. Perhaps the starting point *par excellence* comes from Wimsatt (1986; 1997), who has articulated several criteria over the years distinguishing between wholes that are *mere aggregates* of their parts, and wholes that are constituted by parts in some further way characteristic of organized systems. He suggests that in mere aggregates the parts (i) are intersubstitutable or (ii) can be reaggregated without altering the behavior, (iii) can be added or subtracted without only qualitative changes in behavior, and (iv) exhibit no co-operative or prohibitory interactions. Composite wholes that do not satisfy one or more of these criteria possess organizational properties that give them a more complicated, systemic character.

One of the most basic types of departures from mere aggregativity arises when component parts interact *sequentially* so that at least one component performs its operations on the product of the operation of the previous components. It is often important for efficient operation that the products of one operation are immediately available to the entity performing the second operation. One way to insure this is to situate the component parts spatially and temporally adjacent to each other, fixing them in position. In human engineered machines (e.g., cameras, cochlear implants), it is precisely the imposition of spatiotemporal order that renders parts into the sort of composite system identifiable as a machine. In biological systems, membranes often perform this function; they maintain the enzymes that catalyze a sequence of reactions in close proximity to one another in an organelle.

Particularly significant are deviations from aggegativity that result from going beyond sequential organization by allowing processes later in a sequence to feedback on those earlier in the sequence. Such feedback is often differentiated as negative or positive. In negative feedback, a product of a sequence of operations serves to inhibit one of the earlier operations. For example, the production of ATP from ADP in glycolysis and oxidative phosphorylation serves to inhibit earlier operations in these pathways, insuring that valuable foodstuffs are not metabolized until the ATP is utilized in energy demanding operations and must be regenerated. In positive feedback, a product of an operation might serve to increase the responsiveness of a component earlier in the process. Although positive feedback often results in runaway, uncontrolled behavior, in some cases (e.g., when two reactions each create a catalyst that promotes the other reaction) it results in the self-organization of composite systems. With positive or negative feedback, the causal interactions among component parts are a significant factor by which systems are able exhibit the sort of integrity that allows them to form coordinated, stable subassemblies.

6.2 Contrasts with philosophical accounts of reduction

Mechanisms integrate levels, but the integration they offer contrasts sharply with philosophical accounts of relating levels in terms of the reduction of pairwise theories (*intertheoretic reduction*). Generally, this approach characterizes each level as the locus of theories expressible as sets of axioms and postulates. The classical version of the intertheoretic reduction held that a higher-level theory was reduced to a lower-level theory in virtue of being derived from it, together with a specification of boundary conditions and bridge laws. The boundary conditions restricted the conditions under which the higher-level theories would be applicable to specific situations, whereas the bridge laws equated vocabulary in the higher-level theory with that of the lower-level theory.

The strict derivation condition in the classical version proved difficult to satisfy, and—beginning with the work of Schaffner (1967)—a variety of post-classical accounts of intertheoretic reduction have been developed (Hooker, 1981; Churchland, 1986; Kim, 1998; Bickle, 1998). These accounts exhibit numerous conceptual differences that we will overlook for the purposes of this essay (for an overview, see McCauley this volume, ch. XX). The key feature of these accounts is that they allow that lower-level theories might revise or refine the higher-level ones (depending on the situation, the lower-level theory might entail a replacement or elimination of the higher-level theory).

These post-classical accounts have been criticized on their own terms by a number of philosophers (Wright, 2000; Schouten & Looren de Jong, 1999; Richardson, 1999; Endicott, 2001, 1998). Our goal here is not the evaluation of either classical or post-classical accounts of reduction, but rather to examine the differences between such accounts and mechanist accounts of relations between levels. Perhaps the clearest difference is that mechanistic accounts do not start with separate theories at different levels which are then logically related to one another.

Instead, the accounts offered at each level are partial. An account at the level of the whole mechanism characterizes its engagement with other entities apart from itself. An organism's visual processing mechanism, for example, responds to different visual presentations, providing information to guide higher cognitive activities or action. At the level of component parts and operations, the account describes the operations inside the mechanism which enable it to perform the activity. For example, in the case of visual processing, a lower-level account describes how cells in different processing areas extract specific information from earlier processing areas. Neither account attempts to provide a complete account of everything that happens. And the relation between the accounts results from the ability of a cognizer to simulate how the coordinated performance of the lower-level operations achieves the higher-level activity. The result has the character of an integrated interfield theory (Darden & Maull, 1977) rather than a deductive relation between independent theories.

6.3 Reduction and emergence

Explanations of mechanisms have elements of both reduction and emergence. Mechanisms are inherently multi-level; the components occur and are investigated at one level, whereas the entire mechanism occurs and is investigated at a higher level. A mechanisms is, in some sense, an *emergent* structure that engages its environment, and its ability to do so is explained in part by appealing to its components (but only in part, since the conditions in its environment are also important parts of such explanation).

Hooker (1981, pp. 508-12) provides an instructive example (albeit one he advances in the context of providing an account of reduction). He imagines a series of electrical generators _ that individually exhibit fluctuations in the reliability <r_1,…, r_n> of their net output, but, when regarded collectively, exhibit a far more stable reliability f(r) of output. As a mechanistic system, the generators form a single generator, a 'virtual governor', which is able to do something that no generator on its own can. As such, the properties of the virtual governor just are those of the mechanistic system, and they are neither identical with the properties of any component generator, nor of sets of component generators. Each level of properties exhibits a unique integrity, as the activity of the virtual governor that produces f(r) occurs at a level of organization that is markedly different than the organization at the level where specific reliabilities <r_1,…, r_n> of component generators' net output are produced.

The phenomenon _—virtual government—is an activity of the whole system. But it is an activity that results from the operations of the various component generators in _. Consequently, it is important to 'look down' a level, and this can be done recursively as one continues to descend to lower levels to explain component activities. Yet, _ does not meet Wimsatt's above criteria (i)–(iv) for being a mere aggregate, and so there is a sense in which the activities of government emerge from the operations and organization of components (Bechtel 1995, pp. 147–9). As explanatory models of the causal mechanics involved circle back toward higher and higher levels in order to reconstruct the ultimate relation between explanans and explanandum, the significance of each component alone diminishes and one again focuses on the overall activity—virtual governance.

Accordingly, Wimsatt (1986) reminds us that issues of emergence at higher levels of organization, and of reduction at lower levels, are *inseparably entwined*: “Aggregativity and failures of aggregativity give a clear sense in which a system property may be reducible to properties of its parts and their relations and still be spoken of as emergent.” (Wimsatt, 1986, p. 259; see also Wimsatt, 1976, p. 206; Churchland & Sejnowski, 1992, pp. 2-3). In considering whether ‘having a virtual governor reliability of f(r)’ is a real property, and, if so, whether it is an irreducible one, Hooker similarly concludes that a real property *does* emerge from the structure of the system in explanatory sense—one that cannot be reduced to the reliability outputs of any of the component generators. He, however, waxes instrumentalistic on the question whether there is anything that exhibits virtual governance.[14] For a mechanist who recognizes entities at multiple levels of organization, though, the virtual governor itself is no less real as the component generators.

7. Discovering mechanisms

Thus far, we have focused on conceptions of mechanism, but have not emphasized how mechanisms are discovered. Within the positivist tradition that viewed laws as the primary explanatory vehicle, philosophers following Reichenbach (1938) drew a sharp distinction between the context of discovery and the context of justification. Justification was viewed as an appropriate topic for philosophical analysis since it was construed as involving logical relations, but discovery was viewed as a non-philosophical topic relegated to psychology. The reason for this was relatively easy to appreciate: if laws are not just inductive generations of observations, the process of constructing new laws does not seem to be guided by principles that can be abstractly formulated. Although a number of philosophers have entered into the discussion of discovery in the last twenty-five years (Nickles, 1980b, 1980a; Holland, Holyoak, Nisbett, & Thagard, 1987; Darden, 1990, 1991), the task of analyzing discovery looks more manageable from the perspective of mechanism since the conception of a mechanism itself suggests in large part what needs to be discovered—component parts, their operations, and the modes of organization.

7.1 Explanatory strategies at multiple levels

The multi-level nature of mechanisms can be couched in terms of a trichotomy of explanatory strategies—contextual, isolated, and constitutive (e.g., Craver 2001, pp. 62–8). Contextual strategies (+1 level) describe a mechanism performing operations as a component part in a higher-level, composite system, explain why that mechanism has developed or adapted, suggest how that mechanism is affected ecologically, and so forth. In many cases, contextual strategies are executed by assuming or holding constant some projected description of the organization and operations of component parts, and developing a model of the mechanism’s systemic activity to test against its actual behavior (Bechtel & Richardson 1993, p. 21; Hardcastle 1996, pp. 21–2).

[14] Hooker concludes that predication of “virtual governorship” is ontologically empty, extensionless. “There is no thing which is the virtual governor, so ‘it’ isn’t anywhere, and even the property of being virtually governed cannot be localized more closely than the system as a whole” (1981, p. 509). He extrapolates this conclusion about the mechanism of virtual governorship to the relationship between cognitive and neural states, suggesting that functionally-construed cognitive states recede or “disappear” at lower levels of analysis.

Isolated strategies (0 level) identify the mechanistic activity relevant to the production of _ without reference to ecological or evolutionary context, and without implicating lower-level structure and function. Laboratory studies, such as stimulating and recording spike trains from an isolated neuron or studying a human subject's responses to computer-generated stimuli are examples of this strategy. Setting up such studies requires making judgments about what environment conditions are and are not relevant for the activity and controlling them. Constitutive strategies (-1 level) describe the mechanism's component parts, their operations, and their organization, showing how the mechanism's constituency is responsible for its activity and so makes it more than merely aggregative system. By themselves, constitutive strategies are a form of reductive explanation that involves "taking the mechanism apart."

Ideally, a complete understanding of a given mechanism's systemic activity would make use of each explanatory strategy in order to reflect the hierarchical nature of the composite system: isolating the mechanism to get a handle on its relationship to the target phenomenon produced by it, contextualizing it in order to identify the context in which it operates, and "looking down" a level to identify the lower-level organization of component parts and operations (Craver 2002, p. 91). While looking down is thus not the only explanatory strategy employed in developing an understanding of a mechanism, it is the one most distinctive of the mechanist endeavor and hence the reason why mechanistic philosophers and psychologists typically applaud work on reduction and reductive explanation, but take issue with reductionism (Endicott, 2001, p. 378).

7.2 Decomposition and localization

In explaining any mechanistically produced phenomenon, one adopts the fallible, explanatory heuristics (as opposed to algorithms) of decomposition and localization (Bechtel & Richardson, 1993; Bechtel, 1994, 1995, 2002). *Decomposition* refers to taking apart or *dis*integrating the mechanism into either component parts or component operations, and different experimental procedures figure differently in doing so. *Localization* refers to mapping the component operations onto component parts.

Success in the two forms of decomposition—into component parts and component operations—often occurs on different timescales in different sciences. In the brain and neural sciences, neuroanatomists such as Korbinian Brodmann (1909/1994) developed cytoarchitectural procedures for differentiating brain areas. Brodmann clearly anticipated that areas with different cytoarchitectures would perform different operations, but he had no tools for identifying these operations. When the technique of recording the electrical activity from single cells enabled researchers to begin to determine what stimuli would drive cells in different brain regions, Brodmann's hope began to be realized. As applied to visual processing, for example, the approach allowed researchers to localize different steps in analysis of visual stimuli in different brain regions (van Essen & Gallant, 1994).

For most of its history, researchers in information psychology had no access to what brain components were operative, but they developed powerful strategies for identifying the information processing procedures that subjects were using. As we noted earlier, timing and accuracy of responses were the most common types of data to which these strategies were applied. The type of errors made in a task can often provide suggestions as to the ways subjects are performing the task. Daniel Kahneman and Amos Tversky, for example, used errors in a number of judgment tasks (e.g., do more English words have *r* as a first letter or third letter?) to discover reasoning strategy that produced the judgments (Kahneman, Slovic, & Tversky, 1982). Likewise, perceptual illusions such as the Muller-Lyer illusion and the moon illusion offer clues as to the information processing involved in perception. As these two examples illustrate, investigators cannot directly "read-off" the information processing operations that give rise to false judgments: psychologists need to engage in a kind of reverse engineering to propose what kind of information processing could generate such an error. Moreover, often, as the perceptual illusions illustrate, the proposed mechanisms remain controversial long after the phenomena which prompted the search for the mechanisms are well-established.

Neuropsychology provides yet another strategy for separating information processing operations based on the deficits exhibit by subjects with neural damage. Until the development of neuroimaging techniques, neuropsychologists generally had little information about the locus of brain damage, but they did develop a variety of tests to determine rather precisely the nature of the cognitive deficit. If they discover patients with a distinct deficit (e.g., naming animals), that provides a clue as to the existence of a specific cognitive operation that is compromised. If they can also find other patients who are normal in that capacity but exhibit a contrast deficit (e.g., in naming inanimate objects), the resulting *double dissociation* is taken as evidence that there are two separate operations performed in normal subjects.

Both cognitive psychology and neuropsychology provide techniques for decomposing cognitive function, but not for localizing these in brain regions. Starting in the 1980s with PET and in the 1990s with MRI, neuroscientists have tools for measuring blood flow in different brain regions, which cognitive neuroscientists treat as a proxy for neural activity. Since neural activity is always occurring in the brain, a strategy was needed to link it with cognitive operations. The strategy for doing so was adapted from Donder's original technique for determining reaction time: the blood flow recorded during performance of one task is subtracted from that recorded in a different task requiring additional cognitive processing (e.g., generating an appropriate verb rather than just reading a noun, as reported by (Petersen, Fox, Posner, Mintun, & Raichle, 1989; Petersen, Fox, Posner, Mintun, & Raichle, 1988). Like chronometric studies, these techniques require initial hypotheses about the component information processing steps. But they also have the potential for prompting revisions in these beginning assumptions if the studies identify additional active areas, prompting a inquiry into what operations they contribute. Thus, neuroimaging serves both to localize functional operations onto structural components but also to guide the search for additional functional operations.

7.3 Identifying modes of organization

The discovery of organization often lags behind the discovery of components and their operations. A common strategy when researchers start to take a mechanism apart is to identify a component within the mechanism as alone responsible for the activity of the mechanism. This approach, with Bechtel and Richardson termed *direct localization*, is evident in Broca's interpretation of the area to which he localized damage in Leborgne's brain as the locus of articulate speech and in much of the first generation research in neuroimaging. Typically, however, such research results in discovering more components of the mechanism as involved in the activity, resulting in what Bechtel and Richardson referred to as *complex localization*, where components are construed as operating serially in the generation of the activity. In many cases, however, mechanistic research gives rise to the discovery that the components are not just serially ordered, but figure in cycles and other more complex modes of organization comprising integrated systems.

When the organization being investigated remains relatively simple, it is possible to mentally simulate the activity in the mechanism step by step. As researchers discover complexity in the way the components of a mechanism are organized, however, this becomes more difficult. With complex feedback loops, the mechanism can begin to behave in unexpected ways. To understand such behavior, researchers often need to supplement their own ability to mental trace activity in a system with computer-based simulations. By supplying specifications for the activity of the various components and the manner in which they act upon each other, researchers can discover the consequences of organization.

8. Mechanistic explanation in psychology: Motivation and reward

We finish our discussion by giving an example that will illustrate many of the issues discussed in earlier sections of this chapter—one in which research is progressing because of increasingly sophisticated mechanistic explanations of a psychological phenomenon initially characterized in folk idioms. The example—motivation and reward—has received little discussion in philosophical accounts of psychology. This situation is doubly unfortunate. First, there is currently an intensive search for the mechanisms producing motivational states, and this research is ripe for philosophical analysis. Second, reticence about fully extending the mechanistic approach to psychology has been primarily motivated by the alleged inability to account for purposive, directional, or intentional behavior (Malcolm, 1968; von Eckardt, 2003). Accordingly, were mechanists able to show how a mechanism could account for motivation (a phenomena inextricably bound to purposive behavior if ever there was one!), it would help exorcise such reticence.

As is true of many psychological constructs, the concept of motivation entered scientific psychology as a variable that figured in explanations of behavior—in particular, goal directed behavior of animals in specific environments. To a first approximation, the concept can be characterized as a dispositional variable inferred when behavior is reinforced (a *reinforcer* being anything that will change the probability that immediately prior behaviors will recur). During the heyday of behaviorism, when the goal was to develop laws that describe functional relations between reinforcement and patterns of behavior, Clark Hull (1943) found it necessary to introduce intervening variables such as *drive* and *habit strength* into the laws he proposed. But what do these variables represent? Given the positivistic objectives of behaviorism, this was not a pressing question for Hull; it was sufficient to show how this variable related to the selection, initiation, performance, and reinforcement of behavior. For those psychologists who take the task of psychology to be to explain how the mind/brain produces behavior, however, the question of how motivation operates within the system is compelling.

Mechanistic research on motivation, though, confronts a daunting, if common problem: the term ‘motivation’ does not seem to denote a single, unified phenomenon. On the contrary, it notoriously seems to fragment into a variety of related phenomena at the crossroads of emotion, cognition, and action (similar problems arise with concepts such as attention and memory that have likewise been taken into empirical psychology from the folk idioms for describing mental life). This fragmentation of the construct is manifest in the large litany of conceptions of motivation. For instance, following the currents of Darwinism, some early theorists treated motivation as a matter of non-learned instincts (e.g., McDougall, 1908). Freudian psychoanalytic theory treated motivation as a “hidden force” lurking below the surface of conscious mental processes. Bernard Weiner’s (1986) attributional theory construes motivation as the outcome of causal ascriptions to achievement-related success or failure in psychosocial situations. In contrast, Kent Berridge (2003a; 2003b) construes motivational states as attributions of the ‘wanting’ component of reward that transform perceptual representations into desired incentives for action, and which are independent of hedonic affect.

Reflecting on its extremely wide berth, Brown (1961; see alsoWong, 2000, pp. 1-2) wrote, “The ubiquity of the concept of motivation, in one guise or another, is nevertheless surprising when we consider that its meaning is often scandalously vague….” (p. 24). In the half century since Brown’s remark, the juxtaposition of ubiquity and vagueness has left some researchers wondering about the utility of the concept of motivation, the extent of disunification in the set of phenomena that it picks out, and whether the concept is even necessary for the explanation of molar behavior (Wise, 1989). Consequently, if motivation is to be the focus of mechanistic research, more precise characterizations of the target phenomenon must be developed. One strategy is to try to extract the common factor of the different theories and hypotheses. Kleinginna and Kleinginna (1981) pursued this strategy by reviewing over a hundred such conceptions of motivation in the century since the rise of psychology as a scientific discipline and proposed.

> It may be useful for psychologists to limit motivation, perhaps to the energizing mechanisms that are directly connected to the final common pathway for motor responses. This restriction would exclude both receptor influences and muscular/glandular reactions, as well as most analysis, storage, and retrieval mechanisms…. This view would allow for most of the energizing and some of the directing functions that psychologists traditionally have associated with motivation. These processes may not always be highly localized in the brain and may depend on cortical control as well as on the traditional subcortical motivation circuits such as the lower limbic system structures. By restricting motivation in this manner, we do not overlook the fact that psychological processes are complex and involve continuous interactions among various systems. (1981, p. 272)

A different strategy is to focus on just one of the many phenomena that have been subsumed under the common term ‘motivation’. One approach that has been quite productive has been to focus on motivation in the context of reward and pleasure processes that have been localized in the brain, and develop an account of the mechanisms of motivation in that context (Bielajew & Harris, 1991).

The research developing this approach began with James Olds and Peter Milner’s serendipitous discovery of brain reward circuitry (Olds & Milner, 1964; Olds, 1965). They found that animals with electrodes inserted in certain areas of their brains will work extremely hard when their actions are followed by electric pulses. Further, the profile of reinforcement by self-stimulation and certain drugs of abuse closely corresponds to the response profiles of many natural reinforcers. Their discovery of what they termed 'pleasure centers' represents an attempt to directly localize a psychological phenomenon in a brain region. Since the initial research involved no decomposition of either component parts or their operations, it did not offer an account of a mechanism. But by identifying a component, it opened the path to subsequent research that could differentiate component parts and processes. Some of this research, using both Olds and Milner's original technique, as well as lesion and neuroimaging studies (Martin-Soelch et al., 2001), served primarily to identify a cluster of highly-interconnected anatomical structures broadly constitutive of the mesocorticolimbic system and its component subsystems. This system is composed of a large number of dopamine (DA) neurons whose cell bodies are located in the ventral tegmentum (VTA) and substantia niagra (SN), and which receive significant glutamatergic and GABAergic input. From these two structures, dopaminergic axons project through the median forebrain bundle (MFB) to innervate a number of other components—including the ventral striatum (VS), hippocampus (HC), extended amygdala (A), lateral hypothalamus (LH), and prefrontal cortex (PFC)—all of which are involved in the mediation of complex behavioral responses to reinforcing stimuli. The original localizationist claim has given rise to the targeting of a substantial number of brain regions, raising the question of what process each contributes. One that has figured centrally in subsequent research on motivation and reward is the nucleus accumbens (NAc), an area within the VS which receives major afferent DA projections from the VTA.

<Insert Figure 2 here>

The prominent role of DA in this system inspired the now-infamous dopamine hypothesis of reward. This hypothesis, most notably associated with Roy Wise (1989; 1982; 1985; see also Cabanac, 1992; Ikemoto & Panksepp, 1999; Koob & Le Moal, 1997; Martin-Soelch et al., 2001), suggests that dopaminergic transmission mediates reward and reinforcement, and is the basis for investigating the homeostatic and allostatic mechanisms of motivation. Furthermore, the DA hypothesis of reward has also been used to advance a claim about the pleasurable hedonic affect associated with rewarding and reinforcing stimuli: "Dopamine has often been called the 'brain's pleasure neurotransmitter', and activation of dopamine projections to accumbens and related structures has been viewed by many researchers as the neural 'common currency' for reward" (2003a, pp. 32-33). As a result, Berridge (2003b; see also Salamone & Correa, 2002) concludes that, "There is… evidence to suggest that several basic types of sensory pleasure, including food pleasure, drug pleasure, and sex pleasure, all share in common at least certain stages of their neural circuits" (p. 123).

The import of dopamine transmission through the MFB and NAc has also been demonstrated in numerous studies showing that the addictive properties of certain drugs of abuse are governed by long-term structural, functional, and organizational neuroadaptive changes in the mesocorticolimbic system (Franken, 2003; Koob & Le Moal, 1997, 2001). In cocaine and opioid addiction, the integrity of brain reward function and organization is compromised, and the mesocorticolimbic system adapts its activities to counteract the physiological changes introduced by drugs. Given that the system requires continuous feedback and evaluation of its activities in response to fluctuating environmental stimuli, new set points for brain reward thresholds are constantly being generated in response to increasingly excessive environmental demands on the internal milieu of the mechanism (Koob & Le Moal 2001). When consumption is drastically reduced or terminated after a period of chronic administration, the drug user exhibits a range of aversive withdrawal symptoms and anhedonic deficits, given that the mechanisms for regulating normal brain reward function have been "braked" in order to accommodate the increase in the potentiation of monoaminergic neurotransmission.

Furthermore, in addition to mediation of reward in the mesocorticolimbic system, dopaminergic transmission is also associated with sensorimotor control and locomotion—lending support to Kleinginna and Kleinginna's extracted conception of motivation as the mechanisms responsible for goal-directed behavior and the 'final common pathway' for motor responses. In recent years, however, the dopamine hypothesis of reward has become the focus of criticism, which is directing researchers to yet a more complex account of reward motivation. This criticism is a result of refinements in structural and functional understanding of the mechanism involved in producing reward—refinements which are themselves a result of more accurate attempts at localization and decomposition. Accordingly, the NAc has further been decomposed into a shell and a core, each of which exhibits different neuronal organization and operations. Some operations of mesocorticolimbic mechanisms crucial for dopaminergic transmission have been dissociated—suggesting that reward itself may not be a unitary phenomenon (Berridge, 2003b; Berridge & Robinson, 1998, 2003; Salamone & Correa, 2002; Wise, 1989). Reward processes have been decomposed into those governing the 'liking' properties of hedonic affect, and those governing the 'wanting' or incentive salience properties. In particular, Berridge & Robinson (1998) demonstrated that dopaminergic transmission is necessary for the latter, but not the former.

In conclusion, through the increasingly precise characterizations of motivation, and the ability to localize the component operations in the brain, researchers have developed basic outlines of how the target phenomena are produced, and are revising this outline as more evidence is developed about the contributions of specific neural components. The explanation of how reward, and the hedonic affect associated with it, is produced primarily appeals to the operations and organization of component parts of a series of composite systems and subsystems, and their causal interactions with each other; as such, this example confirms that the import of laws for understanding motivation and reward has limited utility. To be sure, explanations of those mechanisms are mediated epistemically by models, which include inferential operations on schematic representations of processes. The focus downwards to the level of the different components of the mechanism is characteristic of mechanistic research. But we have also emphasized that mechanisms are hierarchical systems and that they also engage their environment at higher levels. A focus on this engagement is also critical in the very characterization of what the components of the mechanism are doing, a point well illustrated by recent reflections of some of the central studies in this endeavor. Thus, Robbins and Everitt (1996) write, "Even leaving aside the complications of the subjective aspects of motivation and reward, it is probable that further advances in characterizing the neural mechanisms underlying these processes will depend on a better understanding of the psychological basis of goal-directed or instrumental behavior" (p. 228). In particular, while the neural substrates of motivation and reward are becoming increasingly well-understood, the authors note that there is an immediate need for systems level functional neuroimaging results which can place the specific components into the broader context of overall brain function. (p. 233). Similarly, Berridge & Robinson

(2003) surmise that, “[F]urther advances will require equal sophistication in parsing reward into its specific psychological components” (p. 507). Consequently, although decomposition and localization are crucial constitutive explanatory strategies, and are continuously applied in the reduction of composite systems into component parts and operations, contextual and isolated strategies are equally important—for no explanation of motivation or reward could possibly hope to be adequate without appealing to psychological and behavioral processes.

References

Bain, A. (1861). *On the study of character, including an estimate of phrenology*. London: Parker.

Bechtel, W. (1994). Levels of description and explanation in cognitive science. *Minds and Machines, 4*, 1-25.

Bechtel, W. (1995). Biological and social constraints on cognitive processes: The need for dynamical interactions between levels of inquiry. *Canadian Journal of Philosophy, Supplementary Volume 20*, 133-164.

Bechtel, W. (2001a). Cognitive neuroscience: Relating neural mechanisms and cognition. In R. Grush (Ed.), *Theory and method in the neurosciences*. Pittsburgh: University of Pittsburgh Press.

Bechtel, W. (2001b). Decomposing and localizing vision: An exemplar for cognitive neuroscience. In R. S. Stufflebeam (Ed.), *Philosophy and the neurosciences: A reader* (pp. 225-249). Oxford: Basil Blackwell.

Bechtel, W. (2002). Decomposing the mind-brain: A long-term pursuit. *Brain and Mind, 3*, 229-242.

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

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

Berridge, K. C. (2003a). Comparing the emotional brain of humans and other animals. In H. H. Goldsmith (Ed.), *Handbook of Affective Sciences (pp. 25-51). New York: Oxford University Press* (pp. 25-51). New York: Oxford.

Berridge, K. C. (2003b). Pleasures of the brain. *Brain and Cognition, 52*, 106-128.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*, 309-369.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences, 26*, 507-513.

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

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

Bielajew, C. H., & Harris, T. (1991). Self-stimulation: a rewarding decade. *Journal of Psychiatry and Neuroscience, 16*, 109-114.

Broca, P. (1861). Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphemie (perte de la parole). *Bulletin de la Société Anatomique, 6*, 343-357.

Brodmann, K. (1909/1994). *Vergleichende Lokalisationslehre der Grosshirnrinde* (L. J. Garvey, Trans.). Leipzig: J. A. Barth.

Brown, J. S. (1961). *The motivation of behavior*. New York: McGraw Hill.

Cabanac, M. (1992). Pleasure: The common currency. *Journal of Theoretical Biology, 155*, 173-200.

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

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* Cambridge, MA: MIT Press/Bradford Books.

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

Clement, J. J. (2003). Imagistic simulation in scientific model construction, *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Craver, C. (2001). Role, mechanisms, and hierarchy. *Philosophy of Science, 68*, 53-74.

Craver, C. (forthcoming). A field guide to levels.

Craver, C., & Bechtel, W. (forthcoming). Explaining top-down causation (away).

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

Darden, L. (1990). Diagnosing and fixing faults in theories. In P. Langley (Ed.), *Computational Models of Scientific Discovery and Theory Formation* (pp. 319-353). San Mateo, CA: Morgan Kaufmann.

Darden, L. (1991). *Theory change in science: Strategies from Mendelian genetics*. New York: Oxford University Press.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 43*, 44-64.

Descartes, R. (1637). *Discours de la méthode pour bien conduire sa raison & chercher la varité dans les sciences*. Leyden: I. Maire.

Descartes, R. (1644). *Principia philosophiae*. Amsterdam: Apud Ludovicum Elzevirium.

Descartes, R. (1658). *Meditationes de prima philosophia*. Amsterdam: J. Janssonium Juniorum.

Descartes, R. (1664). *Traite de l'Homme*. Paris: Angot.

Endicott, R. P. (1998). Collapse of the new wave. *Journal of Philosophy, 95*, 53-72.

Endicott, R. P. (2001). Post-structuralist angst: A critical notice of Bickle's Psychoneural reduction: The new wave. *Philosophy of Science, 68*, 377-393.

Feinberg, T. E., & Farah, M. J. (2000). A historical perspective on cognitive neuroscience. In T. E. Feinberg (Ed.), *Patient-based approaches to cognitive neuroscience* (pp. 3-20). Cambridge, MA: MIT Press.

Flourens, M. J. P. (1846). *Phrenology examined* (C. D. L. Meigs, Trans.). Philadelphia: Hogan and Thompson.

Franken, I. H. A. (2003). Drug craving and addiction: integrating psychological and neuropsychopharmacological approaches. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 27*, 563-579.

Garber, D. (2002). Descartes, mechanics, and the mechanical philosophy. *Midwest Studies in Philosophy, 26*, 185-204.

Giere, R. G. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. G. (1999). *Science without laws*. Chicago: University of Chicago Press.

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.

Hardcastle, V. G. (1996). *How to build a theory in cognitive science*. Albany, NY: SUNY Press.

Hegarty, M. (2002). Mental visualization and external visualizations. In C. Schunn (Ed.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Heil, J. (2002). Functionalism, realism, and levels of being. In U. M. Zeglen (Ed.), *Putnam: Pragmatism and realism* (pp. 128-142). New York: Routledge.

Hempel, C. G. (1958). The theoretician's dilemma. In H. Feigl & M. Scriven & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 2, pp. 37-98). Minneapolis, MN: University of Minnesota Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 137-175.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1987). *Induction: Processes of Inference, Learning and Discovery*. Cambridge, MA: MIT.

Hooker, C. A. (1981). Towards a general theory of reduction. *Dialogue, 20*, 38-59; 201-236; 496-529.

Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Research Reviews, 31*, 6-41.

Ippolito, M. F., & Tweney, R. D. (1995). The inception of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 433-462). Cambridge, MA: MIT Press.

Jackson, J. H. (1931). *Selected writings of John Hughlings Jackson* (Vol. 1). London: Hodder and Stoughton.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kauffman, S. (1971). *Articulation of parts explanations in biology and the rational serach for them. In Robert C. Bluck and robert S. Cohen, eds., PSA 1970. Dordrecht: Reidel.*

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

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of motivation definitions with a suggestion for a consensual definition. *Motivation and emotion, 5*, 263-291.

Koob, G. F., & Le Moal, M. (1997). Drug abuse: Hedonic homeostatic dysregulation. *Science, 278*, 52-58.

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology, 24*, 97-129.

La Mettrie, J. O. d. (1748). *L'homme machine*. Leyde: E. Luzac.

Lashley, K. S. (1948). The mechanism of vision: XVIII. Effects of destroying the visual "associative areas" of the monkey. *Genetic Psychology Monographs, 37*(1948), 107-166.

Lashley, K. S. (1950). In search of the engram, *Physiological mechanisms in animal behavior* (Vol. iv, pp. 454-482). New York: Academic.

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

Malcolm, N. (1968). The conceivability of mechanism. *Philosophical Review, 77*, 45-72.

Marr, D. C. (1982). *Vision: A computation investigation into the human representational system and processing of visual information*. San Francisco: Freeman.

Martin-Soelch, C., Leenders, K. L., Chevalley, A.-F., Missimer, J., Kunig, G., Magyar, S., Mino, A., & Schultz, W. (2001). Reward mechanisms in the brain and their role in dependence: evidence from neurophysiological and neuroimaging studies. *Brain Research Reviews, 36*, 139-149.

McDougall, W. (1908). *An introduction to social psychology*. London: Methuen.

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

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176-185.

Nersessian, N. (1999). Model-based reasoning in conceptual change. In L. Magnani & N. Nersessian & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 5-22). New York: Kluwer.

Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers & S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133-153). Cambridge: Cambridge University Press.

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

Nickles, T. (Ed.). (1980a). *Scientific discovery: Case studies*. Dordrecht: Reidel.

Nickles, T. (Ed.). (1980b). *Scientific discovery: Logic and rationality*. Dordrecht: D. Reidel Publishing Company.

Olds, J. (1965). Pleasure centers in the brain. *Scientific American, 195*, 105-116.

Olds, J., & Milner, P. (1964). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47*, 419-429.

Oppenheim, P., & Putnam, H. (1958). The unity of science as a working hypothesis. In H. Feigl & G. Maxwell (Eds.), *Concepts, theories, and the mind-body problem* (pp. 3-36). Minneapolis: University of Minnesota Press.

Petersen, S. E., Fox, P. J., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing single words. *Journal of Cognitive Neuroscience, 1*(2), 153-170.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(18 February), 585-588.

Posner, M. I., & Raichle, M. E. (1994). *Images of Mind*. San Francisco: Freeman.

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

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science, 45*, 206-226.

Railton, P. (1998). A deductive-nomological model of probabilistic explanation. In J. A. Cover (Ed.), *Philosophy of science: The central issues* (pp. 746-765). New York: W. W. Norton and Company.

Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.

Richardson, R. C. (1999). Cognitive science and neuroscience: New wave reductionism. *Philosophical Psychology, 12*, 297-307.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology, 6*, 228-236.

Salamone, J. D., & Correa, M. (2002). Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behavioral Brain Research, 137*, 3-25.

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

Schaffner, K. (1967). Approaches to reduction. *Philosophy of Science, 34*, 137-147.

Schouten, M., & Looren de Jong, H. (1999). Reduction, elimination, and levels: The case of the LTP-learning link. *Philosophical Psychology, 12*, 237-262.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-656.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

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

Thagard, P. (2003). Pathways to biomedical discovery. *Philosophy of Science, 70*, 235-254.

Titchner, E. (1907). *An outline of psychology* (Revised and enlarged ed.). New York: MacMillan.

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

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

van Fraassen, B. (1989). *Laws and symmetries*. Oxford: Oxford University Press.

von Eckardt, B. (2003). Mechanism and explanation in cognitive neuroscience. *Philosophy of Science, 70*.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*, 158-177.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer Verlag.

Wernicke, C. (1874). *Der aphasische Symptomenkomplex: eine psychologische Studie auf anatomischer Basis*. Breslau: Cohn and Weigert.

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

Wimsatt, W. C. (1974). Complexity and organization. In R. S. Cohen (Ed.), *PSA 1972* (pp. 67-86). Dordrecth: Reidel.

Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. In I. Savodnik (Ed.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry* (pp. 202-267). New York: Plenum Press.

Wimsatt, W. C. (1986). Forms of aggregativity. In M. Wedin (Ed.), *Human nature and natural knowledge* (pp. 259-293). Dordrecht: Reidel.

Wimsatt, W. C. (1994). The ontology of complex systems: Levels, perspectives, and causal thickets. *Canadian Journal of Philosophy, Supplemental Volume 20*, 207-274.

Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science, 64*, S372-S384.

Wise, R. A. (1982). Neuroleptics and operant behavior: The anhedonia hypothesis. *Behavioral and Brain Sciences, 5*, 39-87.

Wise, R. A. (1985). The anhedonia hypothesis: Mark III. *Behavioral and Brain Sciences, 8*, 178-186.

Wise, R. A. (1989). The brain and reward. In S. J. Cooper (Ed.), *The pharmacological basis of reward* (pp. 377-424). New York: Oxford University Press.

Wong, R. (2000). *Motivation: A biobehavioral approach*. Cambridge: Cambridge University Press.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science, 69*, S366-S377.

Wright, C. D. (2000). Eliminativist undercurrents in the new wave model of psychoneural reduction. *Journal of Mind and Behavior, 21*, 413-436.

Young, R. M. (1970). *Mind, brain, and adaptation in the 19th century: Cerebral localization and its biological context from Gall to Ferrier*. Oxford: Clarendon Press.