MECHANISMS AND PSYCHOLOGICAL EXPLANATION

Cory Wright and William Bechtel

1 INTRODUCTION

What is it to explain a psychological phenomenon (e.g., a person remembering a name, navigating through campus, understanding humor)? 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. Although discussion of laws remains commonplace in philosophical accounts of psychological practice, appeal to laws in the explanation of psychological phenomena has become increasingly peripheral in psychology proper — especially with the rise of the cognitivist tradition. Examination of the explanatory discourse of psychologists reveals a shift in emphasis from laws to *mechanisms* — mechanisms of motivation and drug addition, mechanisms of motor development, auditory recognition mechanisms, 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 17^{th} and 18^{th} 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 19^{th} 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 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 casts a number of traditional philosophical issues about psychology in a very different light than in most philosophical discussions. We will discuss some of these in subsequent sections.

Handbook of the Philosophy of Science. Philosophy of Psychology and Cognitive Science
Volume editor: Paul Thagard
General editors: Dov M. Gabbay, Paul Thagard and John Woods

2 THE RISE OF THE MECHANICAL PHILOSOPHY AND ITS APPLICATION TO THE MIND

2.1 Cartesian roots

René 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 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 17^{th} 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" [1644, IV, §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 also 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 — namely, 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:

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

> 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 proposed several 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 serves to heat the blood, thereby causing it to expand and dilate, so that the corpuscles of the blood can move out through the arteries until they cool in the capillaries and return 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 the activities 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 — following a tradition harking back to Galen — 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, Part 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. Of course, humans do perform some activities which non-human animals do not; for some of these — such as the construction and comprehension of novel sentences — Descartes could not conceive of a mechanism, and so concluded that mechanistic explanation of all human activities was not possible:

> 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, they would fail to behave intelligently in others. Descartes maintained that this particularity revealed that their apparently intelligent 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" [1637, Part V] Descartes seemed to be reasoning that reason, as found in humans, is a capacity with universal applicability to any subject, and that, 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. So, instead of explaining these activities in terms of mechanisms, he attributed them to an immaterial mind, i.e., a distinct substance construed 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, Meditation 2]. Hence, unlike material bodies, which were defined in terms of being extended, a Cartesian mind was not something that occupied, or could be 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, the mind had to affect the flow of animal spirits if it were to have any impact. For Descartes it was the pineal gland's central location that made it a good candidate for the locus of interaction, for there it could alter the flow of fluids through the ventricles of the brain through slight shifts in position.

2.2 Mechanizing thought

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 conception of mental processes.

This attitude is well-illustrated at the beginning of the 19^{th} 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 correlations 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. Deploying evidence 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, (e.g., perception), so too to the same extent would other capacities 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 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

[2]Identifying mental abilities with the brain was one of the few features of Gall's views about which Flourens had anything positive to say (though he emphasized that Gall could not 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, 27-8].

writing in opposition to a bad philosophy, while I am endeavouring to recall a sound one" [1846, xiv].

As the 19^{th} century progressed, an alternative tradition of localizing psychological processes in the brain took root. Yet, the guiding conception of mental 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. In many respects, the associationist tradition is, at root, a mechanistic tradition, since it construes thinking as involving an assembling or associating of ideas. Although most 17^{th} and 18^{th} century associationists, such as John Locke and David Hartley, rejected any attempt to link associationist psychology to the brain, an entrée for doing so was provided by Charles Bell's and François Magendie's lesion experiments on dogs in the 1820s, which culminated in the discovery that the posterior spinal nerves are sensory while the anterior nerves are motor. The fact that the sensory inputs arrive at the brain via a different pathway than that carrying the specification of motor activity suggested that the intervening brain constituted a mechanism for making associations. This suggestion was embraced 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, 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, commented that,

> To Prof. Bain I owe much. From him I derived the notion that the anatomical substrata of words are motor (articulatory) processes.

[3]Although Bain did not pursue the physiological component, he clearly construed 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, 29].

> (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, 167-68]

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. 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 sounds 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 20^{th} 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 [1948; 1950], 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-20^{th} century. This changed in the 1960s–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 and Gallant, 1994]. We will return to a detailed example of how neural processes figure in mental activity at the end of the chapter.

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 20^{th} 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, John Watson [1913] and subsequent behaviorists were skeptical of the introspectionist proposals regarding mental processing advanced by Edward Titchner [1907] and his followers. In place of appeals to mental processes, behaviorists sought laws relating behavior to objectively observable variables — stimuli in S-R psychology, reinforcers in Burrhus Frederic 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. This was, instead, a task for physiology. Philosopher Carl Hempel's [1958] theoretician's dilemma captures the essence of the behaviorist's motivation for discounting internal mental events in explaining behavior. According to the dilemma, even if there are intervening processes caused by external variables which then cause behavior, these could be discounted in any explanation. Laws adequate to account for any behavior could be stated purely in terms of the causal external variables.

By the mid-20^{th} 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], however, theorists continued to think primarily 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 MECHANISMS AND THEIR APPLICATION TO PSYCHOLOGY

Theoretical ideas that contributed to the development of the digital computer were in place well before the technology needed for its construction was available. For instance, in 1777, Lord Charles Stanhope invented a device, 'the Demonstrator', which was purportedly capable of solving traditional and numerical syllogisms and elementary problems of probability [Harley, 1879]. In 1805, Joseph-Marie Jacquard introduced the idea of removable punch cards to specify the pattern a loom would weave. Charles Babbage drew upon this invention in the 1840s in his design for an

analytical engine which was to be a steam-driven computational device [Morrison and Morrison, 1961]. Although he was not able to build the analytical engine, he did engage in fruitful collaboration with Lady Lovelace (Ada Augusta Byron), who worked out ideas for programming Babbage's machine. Stanhope's machine and Jacquard's mechanical application of instructions were improved upon by William Stanley Jevons's various logic machines, such as his 'Logical Abacus', which incorporated sundry keys, levers, pegs, and pulleys into a device that anticipated modern calculators [Jevons, 1870]. Other prototypes of logic machines by Henry G. Marquand [1885], Charles P. R. Macaulay, Annibale Pastore, and Charles Peirce [1887] soon followed.

Working at a theoretical level prior to the actual construction of electronic computing machines, Alan Turing abstractly characterized a device (an automaton that came to be known as a 'Turing machine') that would perform the same operations previously performed by 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. (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. The direction to perform an action also specify a new state into which the machine will enter. 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 late autumn 1945, and officially dedicated on February 15, 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 the construction of 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. (Arithmetic computation is just one form of symbol manipulation, and the characterization of a Turing machine as

writing and reading symbols from a tape reveals that it is fundamentally a symbol manipulator that might be employed in arithmetic computation.) 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 Herbert 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 the steps in their reasoning 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 in the development of the tradition in cognitive psychology that construed the mind as an information-processing mechanism. The characterization of the process of symbol manipulation as information processing stems from an independent intellectual 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 and 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 for psychology was 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 conducted 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 [Bechtel *et al.*, 1998].

Miller was one of most influential researchers to develop information-processing models of cognition. 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 [1956] 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. Their goal was to develop a framework that accounted 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 *et al.* 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, as diagrammed in Figure 1.

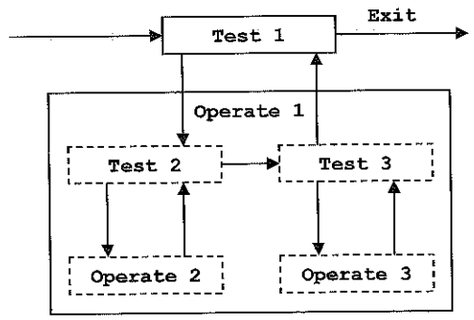

Figure 1. A TOTE unit in which two additional TOTE units are embedded. The upper level TOTE unit performs Test 1. If it is passed, the unit is exited; if not, Operation 1 is performed. This operation requires two additional tests, Test 2 and Test 3, each of which specifies an operation that is performed until that test is passed.

Artificial intelligence and psychology were not the only disciplines affected by the information-processing perspective. Linguistics was another. Starting with the efforts of Ferdinand de Saussure in the late 19^{th} century, structural linguists had 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 began to tackle this problem. He viewed a grammar for constructing possible sentences as a specialized automaton, and explored what kind would be necessary to generate all and only the well-formed sentences of a natural language such as English. One such automaton considered was a *finite state device*, in which the generation of a sentence would involve a sequential transition from state to state in the device. 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. Behaviorist accounts of language tended to be of this type, but Chomsky [1965] contended that finite state grammars cannot adequately characterize natural languages. An example of the problem for a finite state device 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 in a grammar that can be processed by a finite state device; 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 ultimately 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. Among the consequences of Chomsky's attempts to devise grammars and implement them in information-processing devices was that he enticed a number of psychologists to utilize his grammars in their analyses of the mental processes through which humans construct and comprehend sentences. One of these was Miller; already in his work with Galanter and Pribram he envisioned (but did not work out in detail) how a system of TOTE units could realize Chomsky's phrase structure and transformational grammars. Subsequently, he attempted to demonstrate the psychological reality of psychological transformations using reaction time data [Miller, 1962]. Chomsky's continued revisions of his grammars frustrated further attempts by psychologists to make such direct use of his grammars in understanding language processing [McCauley, 1987; Reber, 1987], although the efforts of other psychologists to reformat Chomskian grammars in more psychologically useful ways has spurred a successful lineage of research in psycholinguistics [Abrahamsen, 1987].

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. Already in the mid-19^{th} century, Franciscus Cornelis Donders [1868] developed the idea of subtracting the time required to perform one task from the time required to perform a task requiring an additional operation to determine the time required for that additional operation. Saul Sternberg [1966] employed 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, Sternberg 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, he 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 microelectrodes 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 [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 who employed the information-processing perspective, neuropsychologists were limited in their capacity to relate the brain regions they identified to actual mechanisms responsible for generating behavior [Feinberg and 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 fMRI that researchers interested in the mechanisms underlying psychological processes could link the component operations to brain processes [Posner and 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 [Bechtel, 2001a].

The quest to explain mental activity mechanistically is well-established among practitioners of the biological, neural, and psychological sciences, who are now 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, 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; Wesley Salmon, for one, proclaimed, "The underlying causal mechanisms hold the key to our understanding the world" [1984, 260]. Yet, while empirical researchers frequently refer to and incessantly search for a massive array of particular mechanisms, philosophers have shown — until recently — little interest in what a mechanism is and the manner in which mechanisms might figure in explanations. Philosophers interested in the biological sciences — especially sciences such as physiology, cell and molecular biology, and neurobiology, where 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] offered one of the earliest explicit treatments, though their account was pitched in terms of *machines*.[4] They wrote,

> 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, 17]; see also [Bechtel and Abrahamsen, 2005, 423]

Bechtel and Richardson 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 have been advanced by several philosophers. Stuart Glennan, for example, has endorsed the above conception with the explicit qualification that it make room for the possibility and import of laws: "A mechanism underlying a behavior is a complex system which produces

[4]Bechtel and Richardson's focus on machines raises interesting questions pertaining to the minimal structure sufficient for something to be a machine — especially given that some entities, which are occasionally referred to as 'simple machines' (e.g., the wedge, the wheel), seem to engage in activities (as determined by their shape alone) but lack operative parts. We are inclined to distinguish machines — especially simple ones in the above sense — from mechanisms, at least of the sort that are of interest to biology and psychology.

that behavior by the interaction of a number of parts according to direct causal laws" [1996, 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" — emphasized 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 *et al.*, 2000, 3]. Additionally, Jim Woodward [2002, 374-75] characterized 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 many target phenomena and their associated regularities are the functioning of *composite hierarchical systems*. For example, such a system might encode nociception signals, recall an episodic memory, or plan alternate routes to a landmark. Systems may be active in isolation or in transaction with other things, many of which are themselves mechanisms, and they may be active at one time while inactive at other times.[5] As composite hierarchical systems, mechanisms are composed of *component parts* and their properties. 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 operations of parts is what constitutes or comprises the systemic activity of the mechanism. Stuart Kauffman [1971] proposed that component parts and their operations can be variably picked out according to the overall mechanistic activity to which they contribute causally, thereby making the characterization of the phenomenon critical to the identification of the mechanism.

Clearly though, not just any sequence of causal interactions will suffice, given that mechanisms are generally composed in such a way that they can endure for certain intervals. This raises many thorny metaphysical questions about how best to articulate the individuation and identity conditions for mechanisms — a non-trivial task for any mechanistic approach. For example, what are the necessary and sufficient conditions for a given sequence of causal interactions to be constitutive of a mechanism? How important is endurance, and when do gaps in temporal continuity cease being mere interruptions? To what extent does repair or particulate change in organization involve the creation of a new mechanism? Are there determinate limits on how much spatial or structural disconnectedness a mechanism can exhibit?

Articulating the individuation and identity conditions for mechanisms is an important task for future mechanistic approaches to tackle; in the erstwhile, we note

[5]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 a given phenomenon and its associated regularities as another mechanism's activity.

that many of these questions involve another important feature of mechanisms — namely, how they are configured. The ability to endure and resist degradation importantly suggests that, 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, *co*-operation, connection, feedback, frequency, duration, and so forth — enable the parts to work together effectively and perform the phenomenon Φ targeted for explanation, i.e., the presumable activity of a mechanism. The imposition of organization on components often produces more-or-less stable assemblies whose architecture then fixes the systemic 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* [Simon, 1969]. 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 (see §6.1 below).

5 MECHANISTIC EXPLANATION

5.1 Laws and the ontic/epistemic distinction

For much of the 20^{th} 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 and Oppenheim, 1948]. Salmon [1984] was one of the first to dissent from the hegemony of law-based views of explanation.[7] Although Salmon

[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.

[7]An earlier, but unfortunately less well-known, conception of mechanistic explanation is found in Harré's [1960] 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 §5.2 below.

spoke of 'causal/mechanical explanations', his principal focus was on causation rather than mechanisms of the sort just described. 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 nomological accounts have traditionally identified explanation with (deductive) *argumentation*. As such, the explanandum — a set of statements about the phenomenon Φ to be explained — is a logical consequence of the explanans — a set of statements about the relevant antecedent conditions whereupon Φ is produced together with a generalization (law) stating that, when those conditions prevail, Φ occurs. Salmon objected 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, 239, 273]. Instead, explaining a given explanandum "obviously involves the exhibition of causal mechanisms" leading to the occurrence of that explanandum [1984, 268]. While Salmon did 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" [1984, 17-18], he was 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. [1984, 121-24, 132]

So, it is because causal/mechanical explanations track mechanisms in the world that Salmon construed his account as ontic. Peter Railton [1978] articulated 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, 208]

Machamer *et al.* took 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.

> 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, 21–22]

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 conceptions (e.g., D-N model, classical reduction) advanced by epistemic conceptions of explanation. 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 explanation should emphasize the mechanical production of local, individual phenomena.

Accordingly, on most ontic conceptions, an explanation counts as mechanistic for a phenomenon Φ only when it identifies a composite hierarchical system whose activities account for Φ and whose component parts are organized in certain ways and perform certain operations so as to be constitutive of the systemic activities — thereby situating Φ among a nexus of natural regularities [Salmon, 1984, 260, 268; Bechtel, 2002, 232; Bechtel and Richardson, 1993, 17-18; Glennan, 1996, 61; Railton, 1998, 752; Machamer *et al.*, 2000, 3, 22; Woodward, 2002, 373]. A common way of framing this conception is in terms of giving an answer to a how-question. Accordingly, Paul Thagard 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, 251]; see also [Cummins, 2000]

To grasp the importance of mechanists' appeal to the mechanisms themselves as explanation, consider two further examples that have figured prominently in the recent philosophical literature: electro-chemical synaptic transmission [Machamer *et al.*, 2000, 8-13; Bickle, 2003, 50-62] and stereoptic color and depth perception in state-space [Churchland and 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 systemic 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, function, and organization of the component parts that both constitute the visual system, and comprise the production of those perceptual activities.[8]

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

The current literature on mechanistic explanation is filled with explications of the basic ontic conception wherein the explanans just is a mechanism or parts thereof and the explanandum is some phenomenon. When this view is taken literally, all references to non-ontic features (e.g., statements, inference, reference, truth-preservation) 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 or with the mechanisms that are responsible for them. Hence, in discussing long-term potentiation, Machamer *et al.* wrote, "It is through these activities of these entities that we understand how depolarization occurs" [2000, 13]. Similarly, in discussing biochemical pathways, Thagard wrote, "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" [2003, 238].[9]

[8]The interchangeable use of certain cognates (viz., 'demonstrate', 'reveal', 'lay bare', 'indicate', 'exhibit', 'display') seems particularly counterproductive; for they are typically left as semantic or conceptual primitives, and are not clarified, characterized, or given meaning over and above the already intuitive 'explain'.

[9]To be fair, Thagard [2003, 237, 251–52] calls 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. We certainly applaud the call, but go even further in jettisoning the reliance on the (naïve) ontic/epistemic distinction.

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 problem with construing the ontic conception literally is well-illustrated in Scriven's example of the Yugoslav garage mechanic. In attempting to minimize any epistemic components of explanation, defenders of the ontic conception sometimes suggest 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 that account for the explanandum. In critiquing Hempel's view that explanation requires argument, Scriven levels the following complaint: "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, 10]). But what work is a gesture or an indication doing? 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 a large corpus of conceptual knowledge. And even if cognitive salience-conferring behaviors were sufficient for explanation, no such demonstrative gesturing or pageantry could alone unequivocally specify what had been indicated, since myriad structures, functions, and organizations would be consistent with any such gesture.[10]

One important consequence is that the view that mechanistic explanations are given by invoking 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 any explanans; the relevant parts, operations, and organization minimally need to be captured and codified in a structural or functional *representation* of some sort. Accordingly, generating scientific understanding through mechanistic explanation necessarily involves inferences and reasoning about those representations.

[10]The same point can be made in the context of or 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 attempts 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, memory consolidation, or creativity 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.

(Of course, the operations on *re*presentations of mechanistic activity involved do not reduce to syntactic operations on propositions.) And so, unless the ontic conception is understood as metonymical, such accounts of mechanistic explanation will misconstrue explanatory practice by leading into absurd de re/de dicto confusions. After all, *explaining* refers to a *ratiocinative practice* governed by certain norms 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; the systemic activity of a mechanism may be responsible for the presence of some psychological phenomenon Φ, but Φ is not explained until a cognizer contributes his or her explanatory labor.[11]

Now, there is clearly a contrast between the role of laws versus mechanisms in scientific understanding and explanation; but an ontic/epistemic distinction turns out to be an unfortunate way to draw that contrast because it invites trivializations and confusions. On one hand, characterizing explanation as ontic does not carry any additional commitments over and above what mechanists have already spelled out, so employing the distinction in the way Salmon intended merely reiterates the standard mechanist line on subverting the exclusive focus on laws and lawlike generalizations in the context of deductive arguments. Worse, taking ontic conceptions of mechanistic explanation at face value leads to absurdities, since mechanisms themselves are not the sorts of things that can be constituents of any explanans; and characterizing explanation as non-epistemic is clearly problematic insofar as explanation is through-and-through an epistemic practice of making the world more intelligible by providing a rational basis for inferring how some given psychological phenomena is or could be produced. On the other hand, it is entirely possible to give an account of explanation that is significantly epistemic without being overtly nomological (for related arguments, see Waskan, forthcoming). The upshot is that construing mechanistic explanation ontically and non-epistemically poorly captures what is distinctive of mechanistic explanations — i.e., the distinction is a false dichotomy.

Advocates of the ontic conception of mechanistic explanation cannot avoid appeal to epistemic conceptions; even when kicked out the front door, epistemic conceptions are quickly ushered in the back. One way in which this is manifest is in mechanists' vacillation between appealing to mechanisms themselves and identifying the explanans of mechanistic explanations with sets of descriptions of mechanisms. Hence, Machamer *et al.* aver that, "Giving a description of a mechanism for a phenomenon is to explain that phenomena and its production" [2000, 3]; see also [Craver, 2001, 68], while Glennan wrote, "A description of the

[11]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 stood as a relatum in an explanation, erroneous explanation would not be possible since there would have been no mechanism for the erroneous explanation to have identified.

internal structure of the mechanism explains [its] behavior" [1996, 61]; see also [2002, 347].[12] This vacillation reflects recognition that a necessary condition on mechanistic explanation is that the structure, function, and organization of mechanisms needs to be captured and codified representationally. And this recognition shows 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 cannot but simultaneously help themselves to certain resources of epistemic conceptions they abandon.

A thoroughly epistemic conception of mechanistic explanation can certainly maintain the contrast with nomological accounts by acknowledging the importance of representing mechanisms rather than laws. Such an epistemic conception can simultaneously recognize that semantic, logical, grammatical, and inferential concerns are not irrelevant while emphasizing that linguistic representations — much less those involving universal quantification — do not exhaust the range of possible representations. In many cases, graphical representations such as diagrams or figures are *much* 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 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 (see [Bechtel and Abrahamsen, 2005]). 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 *visualize* how that phenomenon is generated. This is not an area in which there has yet been much work by philosophers.[13] Cognitive psychologists have begun to make some headway in characterizing the processes by which scientists and science students develop the ability to mentally simulate the behavior of simple mechanisms such as springs [Hegarty, 2002; Clement, 2003]. There is related work by philosophers [Nersessian, 1999; 2002] and psychologists [Ippolito and Tweney,

[12]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 and simulatory operations on them. (Considerations of the semantics of the explanatory connective 'because', as well as what it is that arrows in box-and-arrow diagrams represent, help in grasping this point; see [Talmy, 2000].)

[13]An exception is Waskan [forthcoming], who develops complementary arguments to those offered here against a purely ontic conception of explanation and advances an account in terms of what he calls intrinsic cognitive models that are more like scale models than lingusitic descriptions.

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 in 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.)

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, psychology included. Models stand in for the actual systems that researchers are trying to understand and are invoked in reasoning and theorizing about the actual system (see [Harré, 1960]). Beyond these commonalities, there are many disparate senses of 'model' and contexts in which they 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 (non-physical) objects that conform to the theories advanced in a science. These then stand in (approximately) isomorphic relations to the actual objects in the world to which the theory supposedly applies. In the context of mechanistic explanation, the model's elements correspond to the parts of a mechanism, and their structure conforms, not to a theory, but rather to the mechanism's constituency and interactivity.

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 with the goal of understanding the mechanism being modeled. One of the best known examples is the physical model of DNA that James Watson and Francis Crick constructed in the course of 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 more generally — namely, that 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, reconsider 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 inferentially-based representation of the production of psychological phenomena as it is an aversion to doing so by deducing descriptions of phenomena from laws. Robert Cummins [2000, 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 wrote,

> 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, 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 the term '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 and Richardson wrote that, "The explanatory task begins and ends with models; we question the hegemony of laws in explanation, not their existence" [1993, 232]. 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 modeling 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 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's appeal to increasingly finer-grain component operations and parts in explaining the activity of a mechanism. But in another sense, a mechanistic explanation is non-reductionistic: explanations at a lower level do not replace, sequester, or exclusively preside over the refinement of higher-level explanations, because mechanisms are hierarchical, multi-level structures that involve real and different functions being performed by the whole composite system and by its component parts [Wright, forthcoming]. 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 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 LEVEL in these disciplines. Just the same, levels-talk is virtually threadbare from overuse [Craver, in press]. One reason is that the various conceptions 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 and Craver, this volume]. A second reason is that levels are ambiguously construed as both ontic levels of mechanistic organization and as epistemic levels of analysis [Bechtel, 1994].

A traditional conception suggests a neat hierarchical layering of entities into levels across phenomena. Here, scientific disciplines (viz., physics, chemistry, biology, psychology, sociology) are distinguished, in part, by the level of phenomena in nature that are their target of study. This conception 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 argued 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 offered 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 has been further developed by Wimsatt [1976], who also argued that the most frequent causal interactions are among entities of the same scalar magnitude. Yet, Wimsatt noted 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 introduced 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 traditional, overtly realist construal of levels. For instance, Craver [2001, 65-67] decries such accounts 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. The sense of perspectivalism involves using isolated, constitutive, and contextualized strategies for locating and differentiating components and their operations, as well as explaining how they work together to comprise mechanistic activity. The construal of levels as perspectival is also congruent with the idea that levels cannot be neatly delimited; for instance, Robert Wilson [2003] uses group and individual fitness in biological populations to propose that perspectival explanatory strategies naturally follow from the fact that levels of selection are actually fused or entwined, which precludes any determinate answers about causal efficacy or closure at a given level. Valerie Hardcastle [1996, 29-32] adopted a similar stance in arguing 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); taking this point to its logical conclusion, John Heil [2002] has urged that talk of ontic levels of organization be dispensed with altogether.

Between these two extremes of realism and anti-realism about levels, an intermediate construal of levels has emerged. On a given cycle of decomposing a mechanism, this intermediate construal treats all relevant components as being at the same level [Craver and Bechtel, in press; Craver, in press; Bechtel, 2006]. Such decomposition is local to each mechanism and no attempt is made at identifying the level each component would occupy in a global portrayal of levels across all of nature. So, while this construal takes levels seriously — i.e., there are levels, and they are integrated in mechanisms — it is more restricted than the traditional conceptions advocated by Oppenheim and Putnam, Simon, and Wimsatt. For example, suppose a biological mechanism is being described in which sodium molecules cross a membrane. The traditional construal would have the membrane being 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 fact that molecules are constituent parts of the membrane need not entail that they be treated as existing at a lower ontic level than the membrane across which they are transported; relative to the explanatory perspective and purpose of explaining sodium transport, the membrane and the sodium molecules can be construed as existing at the same level, since each is a component part of the mechanism relevant to the activity under investigation. 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 and bears the imprint of perspectival explanatory strategies. Multiple cycles of analysis thus give rise to a hierarchy of levels that is confined to a given mechanism, though the reality of the mechanism entails the reality of levels confined to it. An investigator who had started by focusing on a different activity may have ended up with a different parsing of entities into levels (see [Kauffman, 1971]). In sum, levels on the mechanistic account are real in that they deal the particularities of actual components and their operations, but they are perspectival in that they are defined with respect to specific foci on mechanistic activities.

One of the main ways of staving off confusion has been to distinguish between the mechanistic conception of levels of components and the conception of epistemic levels of *analysis*. Perhaps the best known epistemic conception of levels in psychology and cognitive science is due to David Marr, who distinguished three "different levels at which an information device must be understood before one can be said to have understood it completely" [1982, 25]:

1. 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,
2. the choice of the representation of the input and output and the algorithm to be used to transform one into the other,
3. the details of how the algorithm and representation are realized physically — the detailed computer architecture, so to speak.

The emphasis for Marr was on the different epistemic projects an investigator can pursue and the kind of account required, not on mereological 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, is mereological — working down to a lower level involves decomposing something into its component parts and operations, rather than merely describing it differently.

One important consequence of the mechanistic, mereological account of levels is that it makes it much clearer how component parts operating within a mechanism can perform different functions than the composite system. 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 notion of ORGANIZATION in play in phrases like 'levels of organization', yet that concept has not received sustained, 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 with 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. Biological mechanisms such as 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 aggregativity 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 (e.g., when two reactions each create a catalyst that promotes the other reaction), in some cases it results in the self-organization of composite systems [Kauffman, 1993]. With positive or negative feedback, the causal interactions among component parts are a significant factor by which systems are able to exhibit the sort of integrity that allows them to form coordinated, stable subassemblies.

Organization is especially important for understanding some of the more salient characteristics of living organisms such as their ability to develop and maintain themselves over time. Not able to rely on external agents to construct or repair them, they need to perform these operations for themselves. This means that they must be organized so as to secure matter and energy from their environments and channel them into the construction and repair of their own bodies. Such systems exhibit what Ruiz-Mirazo and Moreno [2004] identify as *basic autonomy*: they maintain themselves as identifiable functioning systems by constructing and reconstructing themselves and managing the exchange of nutrients and waste products with their environment (for further discussion, see Bechtel, in press). Beyond those mechanisms required for basic autonomy, living organisms may evolve additional mechanisms. Each of these, however, must be constructed and maintained

through the organism's own processes, imposing serious organizational constraints on living systems and the mechanisms, including cognitive mechanisms, comprising them. While individual mechanisms of biological organisms may perform specialized tasks, they are necessarily highly integrated with each other, especially those responsible for fundamental energetic and generative processes.

6.2 Contrasts with philosophical accounts of reduction

Mechanistic explanations relate levels, but the relation proposed contrasts sharply with philosophical accounts of *intertheoretic reduction* that relate levels in terms of the reduction of pairwise theories. Generally, this approach characterizes each level as the locus of theories expressible as sets of axioms and postulates. The classical version of intertheoretic reduction [Oppenheim and Putnam, 1958; Nagel, 1961] 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 principles. The boundary conditions restricted the conditions under which the higher-level theories would be applicable to specific situations, whereas the bridge principles 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 advanced [Hooker, 1981; Churchland, 1986; Kim, 1998; Bickle, 1998]. A key feature of these accounts is that they allow that lower-level theories might revise or refine (or, in certain circumstances, even entail replacement of or elimination of) the higher-level theories with which they are paired via a corrected image (we will overlook the nuances of these accounts for the purposes of this essay; (for an overview, see [McCauley this volume; Bechtel and Hamilton, in press]). These post-classical accounts have also been encumbered by certain difficulties [Endicott, 1998; 2001; Schouten and Looren de Jong, 1999; Richardson, 1999; Wright, 2000].

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 simply that mechanistic accounts do not start with separate theories at different levels that are subsequently related to each other with the formal apparatus of set theory or reconstructed in their idealized forms. Mechanistic explanations at each level are partial and constructed piecemeal with a focus toward actual experimental investigation, without overarching concerns that they be fit into grand, large-scale scientific theories; hence, there is no desideratum to provide a complete account of everything that happens. Further, the relation between different explanations at different levels results from the ability of a cognizer to simulate how the coordinated performance of the lower-level operations achieves the higher-level activity [Wright, forthcoming]. The result has the character of an interfield theory that identifies causal or mereological relations between phenomena described in different theories

[Darden and Maull, 1977] rather than a deductive relation between independent theories [Bechtel and Hamilton, in press].

Moreover, models of mechanisms have elements of both reduction and emergence. Mechanisms are inherently multi-level; the components and their operations occur and are investigated at one level, whereas the entire mechanism and its activity occur and are investigated at a higher level. Consequently, mechanistic explanation proceeds, in part, by modeling components of a system and how they are organized. At the level of component parts and operations, the explanation describes the operations internal to the mechanism that comprise and enable it to perform the overall 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. Of course, an organism's visual processing mechanism responds to different visual arrays in order to provide information that guides higher mental activities or action. Procedurally, then, modeling components, their operations, and organization is only part of the overall explanation, since the conditions in its environment are also important parts of such explanation. A mechanism is, in one sense, an *emergent* structure that engages its environment, and explanations at the level of the whole mechanism must therefore characterize its engagement with other entities apart from itself.

Hooker [1981, 508-12] provides an instructive example (albeit one he advances in the context of providing an account of reduction). He imagines a network of electrical generators that individually exhibit fluctuations in the reliability $\langle r_1, \ldots, r_n \rangle$ 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, systemic-level generator — what Hooker calls a 'virtual governor' — that is able to do something which no generator on its own can. As such, the properties of the virtual governor just are those of the whole mechanistic system, and they are neither identical with the properties of any component generator, nor of sets of component generators. Each level of organization exhibits a unique integrity. The activity of the virtual governor that produces $f(r)$ occurs at a level of organization that is markedly different than the level of organization at which each component generator produces a specific reliability r_i.

The phenomenon Φ to be explained — 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 the network. Consequently, it is important to "look down" a level — something which can be done recursively as one continues to descend to lower levels to explain how activities at a given level are comprised of component operations. Yet, the network 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, 147–49]. 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 systemic activity — virtual

governance.

Accordingly, Wimsatt's [1986] reminder that issues of emergence at higher levels of organization, and of reduction at lower levels, are *inseparably entwined* is important to heed: "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, 259]; see also [Wimsatt, 1976, 206; Churchland and Sejnowski, 1992, 2-3]. In considering whether 'having a virtual governor reliability of $f(r)$' predicates 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 an explanatory sense — one that cannot be reduced to the reliability outputs of any of the component generators. For his part though, Hooker waxes instrumentalistic on the question whether there is anything that exhibits virtual governance.[14] For mechanists who recognize entities at multiple levels of organization, though, a systemic-level governor itself is no less real than the component generators.

7 DISCOVERING MECHANISMS

Thus far, we have focused on conceptions of mechanism, but have only alluded to 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, while 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 generalizations 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 and have advanced constructive proposals [Nickles, 1980a; 1980b; Holland *et al.*, 1987; Darden, 1990; 1991], the task of analyzing discovery looks more manageable from the perspective of mechanism since the conception of a mechanism itself largely suggests what needs to be discovered — component parts, their operations, and the modes of organization.

[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, 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.

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, 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 and Richardson, 1993, 21; Hardcastle, 1996, 21-2]. Isolated strategies (0 level) identify the mechanistic activities that account for the presence of Φ 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 a 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: situating the mechanism to get a handle on its relationship to other phenomena, identifying the specific variables that affect its ability to realize the phenomenon in question, and "looking down" a level to identify the lower-level organization of component parts and operations [Craver, 2002, 91]; for a similar perspective emphasizing the importance of adopting pluralistic perspective, not just a downward one, see [McCauley, 1996; McCauley and Bechtel, 2001]. 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, 378]).

7.2 Decomposition and localization

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

parts (structural decomposition) or component operations (functional decomposition).

These two forms of decomposition typically require different experimental tools and techniques, and success in decomposing component parts and 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 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 and Gallant, 1994].

For most of its history, researchers in information-processing 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 suggest the reasoning strategy subjects employed both when they produced the right answer and when they made errors [Kahneman *et al.*, 1982]. Likewise, perceptual illusions such as the Muller-Lyer illusion and the moon illusion offer clues as to the information-processing operations 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.

Neuropsychological research involves yet another strategy for separating information-processing operations based on the deficits exhibited by subjects with neural damage. Until the development of neuroimaging techniques, neuropsychologists generally had little information about the locus of brain damage, but they developed a variety of tests to determine rather precisely the nature of the cognitive deficit in a patient. The discovery of patients exhibiting a distinct deficit (e.g., naming animals) provides a clue as to the specific cognitive operation that is compromised. If neuropsychologists 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 decom-

posing cognitive function, but not for localizing these in brain regions. Starting in the 1980s with PET and in the 1990s with fMRI, neuroscientists employed 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 Donders' original technique for determining reaction times for component mental operations: 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 [Petersen *et al.*, 1989; 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. Such discovery prompts inquiry into what operations these areas are performing. Thus, neuroimaging serves both to localize functional operations onto structural components as well as 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 systemic activity of the mechanism. This approach, which Bechtel and Richardson [1993] termed 'direct localization', is illustrated in both 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 are involved in the activity, prompting researchers to investigate the operations they perform. Once multiple components are identified as being involved in the production of a given mechanistic activity, researchers attempt to relate these as operating serially. Bechtel and Richardson referred to such serial models as yielding a 'complex localization'. 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 construct relatively simple explanatory models through mental simulation of the activity in the mechanism step-by-step. As more organizational complexity and *co*-operations are discovered, 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 offload some of the cognitive labor involved in constructing explanatory models of parallel complexity onto their (research) environment — e.g., by supplementing their own ability to mentally trace activity in a system with computer-based simulations. By supplying models of the operations of the various components and the manner in which they interact with each other, researchers can discover some of 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: motivation. Surprisingly, this example has not received much discussion in philosophical accounts of psychology (a noteworthy exception might be [Schroeder, 2004]). One of the main reasons involves difficulties in imagining how the invocation of mechanistic activity could suffice to explain such an abstract — if even not simply abstruse — psychological phenomenon such as motivation. After all, motivation potentially implicates such murky issues as life-long aspirations, psychodynamic personality traits, diachronic desires for intangible psychosocial rewards, and so forth. 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].

Showing how mechanistic models could account for motivational and related states would therefore be a major boon in exorcising such reticence. In fact, significant scientific progress has been made in characterizing the mechanisms responsible for producing motivation, much of which is due to increasingly sophisticated mechanistic explanations of a psychological phenomenon initially characterized in folk idioms. These explanations are through-and-through mediated epistemically by models (an important feature of which, again, is the range of inferential operations on schematic representations of processes). Such mechanistic explanations of motivation are particularly well suited for use in and development of analyses of topics traditionally thought to fall solely within the province of philosophy (the nature of desire, weakness of will and compulsion, reasons explanations, etc.).

8.1 Early motivation research

In psychological theorizing at the turn of the 20^{th} century, motivation was generally understood in terms of a large network of "hidden forces" that subsisted below the level of conscious mental processes but manifested themselves in human conduct. In views as disparate as William James's [1890/1950] and Sigmund Freud's [1922], attempts to differentiate among the class of hidden forces focused on varieties of instincts — latent, unlearned biological causes which compel an organism to behave in particular ways in the presence of certain environmental stimuli. Whereas James proposed a taxonomy of twelve such instincts (e.g., imitation, fear, love, curiosity, cleanliness),[15] Freud confined himself to two overarching

[15]Some of James' instincts were cross-classified with the 18 'hidden forces' posited William McDougall [1908], though the latter drew heavily on Darwinian evolutionary theory and ideas

ones — the life and death instincts — but then characterized them in terms of concepts like MOTIVATION and PLEASURE and AROUSAL in his theory of the unconscious. What he called the 'Id' is simply a wellspring of urges administered by hedonistic principles.

Despite a new mandate for empirical study, early psychologists' positing of 'hidden forces', and the analysis of such forces into elaborate classification schemes of (innate) instincts, did little to illuminate the nature of the states, processes, and properties that "move" animals to behave; moreover, the paucity of consensus about both the nature and number of instincts made it clear that the taxonomies were more descriptive than explanatory. Thus, early psychologists' theorizing about motivation made few advances on centuries of philosophical rumination — from Bacon and Hobbes to Bentham, Mill, and Nietzsche.[16]

As is true of many psychological constructs, the concept of MOTIVATION entered into more recognizably scientific accounts as a variable that figured in explanations of the goal-directed behavior of both human and nonhuman animals in specific environments. In lieu of positing hidden forces, instincts, or other unobservable psychobiological states, behaviorists sought to discover the laws relating behavioral regularities to objectively observable variables. During the heyday of behaviorism, Clark Hull [1943] found it necessary to introduce intervening variables (e.g., drive, habit strength) into the laws he proposed to describe functional relations between reinforcement and patterns of behavior. But what do these variables stand for? Given the positivistic objectives of behaviorism, this was not a pressing question for Hull; it was sufficient to show how these intervening variables related to the selection, initiation, performance, and reinforcement of behavior.

Hull and other behaviorists bequeathed psychology with a concept of MOTIVATION — characterized as a dispositional variable inferred from reinforced behavior (a 'reinforcer' being defined as anything that will change the probability that immediately prior behaviors will recur) — which was well-suited for some types of experimentation. The advent of cognitivism initially only added to the clutter of motivation constructs, as post-positivist research focused on long-term planning, the influence of emotional factors, construction of prosocial skills, etc. John W. Atkinson's [1957] work, for instance, focused on cognizers' rational calculation of expectations of goal attainment and goal value; Bernard Weiner's [1986] attributional theory construed motivation as the outcome of causal ascriptions to achievement-related success or failure in the context of socially structured endeavors.

Consequently, many researchers were left wondering about the very utility of

about innateness.

[16] An important exception was Henry Murray, whose laboratory at Harvard combined an emphasis on personality factors with then-novel work on homeostatic mechanisms. Murray *et al.* [1938] catalogued dozens of viscerogenic and psychogenic "needs" and environmental "presses," and characterized them (achievement, power, affiliation, succorance, etc.) based on views of physiological regulation circulating at the time (e.g., Cannon, 1929) — namely, as electrochemical "energies" in the brain which, when activated by certain situations, create internal states of tension and insatiety that animate resolutory actions.

motivational constructs, the extent of disunification in the set of phenomena they pick out, and whether the concept of MOTIVATION is even necessary for the explanation of molar behavior. Of course, the term 'motivation' itself — derivative from the Latin verb 'movere' and the German 'Motivierung' — generally signifies that which guides movement or causes one to act; yet, due to a number of theoretical refinements and technological advances, the denotatum had clearly fragmented into a variety of related phenomena at the crossroads of affect, cognition, and action. Reflecting on its extremely wide berth, Judson Brown 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" [1961]; see also [Wong, 2000, 1-2].

In a review of the motivation literature since the advent of psychology as a scientific discipline, Kleinginna and Kleinginna [1981] extracted a core group of common factors from over 100 wildly different conceptions, theories, and hypotheses. They found that numerous conceptions focused on the final common pathways of mechanisms producing goal-directed behavior. Based on this review, the authors made the following recommendation:

> 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, 272]

8.2 From motivation to the mechanisms of brain reward function

Confining motivation research to better-delimited characterizations of the target phenomenon with more tractable constructs was one way in which psychologists were able to get a handle on motivation. This coincided with the move toward models of mechanistic activity in lieu of nebulous instinct-need-desire taxonomies and descriptions of lawful stimulus-response patterns — a move fostered by the serendipitous discovery of brain reward circuitry in the mid-1950s. James Olds and Peter Milner [1954] found that animals with electrodes inserted in certain areas of their brains will work extremely hard when their actions are paired with certain electric pulses. They hypothesized that the structure being stimulated — roughly, the medial forebrain bundle (MFB) — is responsible for directly mediating both

the hedonic effects of, and complex behavioral responses to, all pleasures and rewarding stimuli [Olds, 1965]. This represented an attempt to directly localize a psychological phenomenon in a brain region. Of course, since the initial research involved no decomposition of either component parts or their operations, it did not actually offer an account of a mechanism. It did, however, serve to restrict the focus of much research to the more tractable topic of reward. Moreover, localizing reward in the MFB provided a target neural structure that subsequent research could investigate. As we will see, this led to differentiation of component parts and operations involved in reward and an understanding of how they related to one another, thereby initiating a path of research culminating in a mechanistic explanation of reward.

Olds and Milner's original technique was developed into a reliable experimental test paradigm: intracranial self-stimulation [Bielajew and Harris, 1991]. As with other such test paradigms and animal models, the development of intracranial self-stimulation provided a quantitative measure for accessing a phenomenon previously thought to be impenetrably subjective; its value as a "good" animal model, however, lies not so much in being directly and comprehensively probative of pleasure and reward as in adding a new dimension for moving mechanistic research forward along with other neuropharmacological screening tests, behavioral bioassays, and simulations [Wright, 2002]. As lesion and neuroimaging techniques were eventually added to the experimental portfolio, researchers began to differentiate parts of the system, identify different operations performed by these components, and develop a schematic understanding of the workings of hedonic information processing.

A key result of this research was the identification of a cluster of highly-interconnected anatomical structures broadly constitutive of three interrelated systems: mesocortical, mesolimbic, and mesostriatal. A major component of these systems is a large number of dopamine neurons whose cell bodies are located in two proximal, evolutionarily primitive structures — the substantia nigra pars compacta (SN) and the ventral tegmentum (VTA). From these two nuclear groups, dopamine axons project through the MFB and basal ganglia to innervate a number of structures — including the ventral striatum (VS), ventral pallidum (VP), 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.

Together these two neuronal groups involve an interface between tightly coupled motor and reward mechanisms. Dopamine cell bodies in the SN, for instance, project throughout the mesostriatal system to the caudate (CN), putamen (P), and motor cortex (MC), and are well-known to produce an array of signals for sensorimotor control, locomotion, and the initiation of movement. Their degradation or destruction typically results in Parkinson's-like symptoms. VTA projections provide the basic structure for transmitting reward-related associative learning signals in the mesolimbic system.

With the identification of a host of interconnected brain areas apparently involved in reward, the next challenge was to figure out what each does — that is, to localize operations in the various areas. We will focus on one area that has been the locus of intense scrutiny and additional localization claims — nucleus accumbens (NAc), an area within the VS receiving major afferent dopaminergic projections from the VTA. This area has been further decomposed into a shell and a core, subcomponents which exhibit different neuronal organization and perform different operations. In particular, opioid receptors are numerously expressed at synapses in the NAc shell. These — in conjunction with mesolimbic dopamine activity originating from the VTA — provide key reward-related signals that regulate glutamatergic and GABAergic output to various cortical and striatal structures, respectively. These signals are often marked by the shift from tonic firing rates to phasic bursting.

Naturally rewarding stimuli are known to potentiate NAc shell transmission [Di Chiara, 1999; Berridge, 2003a]. But a wealth of convergent evidence about the contribution of these neurons to reward has also come from self-administration and self-stimulation studies with certain drugs of abuse whose reinforcement profile closely corresponds to those of many natural reinforcers. Generally, chronic drug use compromises the integrity of brain reward function and organization, resulting in long-term structural, functional, and organizational neuroadaptive changes in

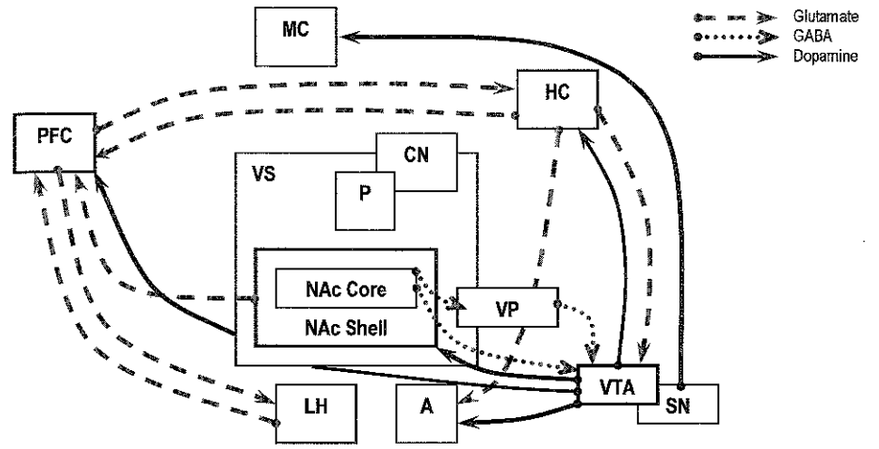

Figure 2. A schematic diagram of glutamatergic, GABAergic, and dopaminergic pathways involved in mesocorticolimbic circuitry, adapted from Berridge (2004, p. 204). Prefrontal cortex (PFC); Motor Cortex (MC); Hippocampus (HC); Ventral Striatum (VS); Putamen (P); Caudate Nucleus (CN); Nucleus Accumbens (NAc); Lateral Hypothalamus (LH); Amygdala (A); Ventral Pallidum (VP); Ventral Tegmental Area (VTA); Substantia Nigra (SN)

the mesolimbic system [Franken, 2003; Koob and Le Moal, 1997; 2001]. For instance, cocaine addiction reduces the expression of endogenous κ-opioid receptors, which provide a natural inhibitory system for keeping tonic dopamine levels in the NAc in check [Chefer *et al.*, 2005]. The mesolimbic system adapts its activities to counteract the physiological changes introduced by drugs by shifting from homeostatic to allostatic mechanisms: 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 and 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.

Recent studies have decomposed NAc shell structure and function yet further. For instance, Ikemoto *et al.* [2005] demonstrated that the NAc shell supports heterogenous operations, as mice differentially self-administer δ-amphetamine in its ventral and medial parts; and Taha and Fields [2005] identified two distinct NAc shell populations with different encoding properties — inhibitory firing immediately before and during the initiation and maintenance of reinforcer consumption, and excitatory firing to encode palatability preferences.

The prominent role of dopamine in the mesolimbic system inspired a series of now-infamous dopamine hypotheses of reward [Wise, 2004]. These hypotheses suggest that dopaminergic transmission is the primary mediator of reward and reinforcement, and is a crucial operation in many reward-related homeostatic and allostatic mechanisms. In tandem, these hypotheses have also been used to advance claims 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" [Berridge, 2003a, 32-3]. However, from increasingly sophisticated localizations and decompositions, it has become clear that this general picture of dopamine stands in need of revision. Although dopamine plays a crucial role, it does not seem to operate in isolation. Berridge and Robinson [1998] report that dopamine-selective neurotoxins such as 6-hydroxydopamine hydrobromine (OHDA-6) knock out virtually all NAc and LH dopamine operations, but not the capacity of rats to spontaneously engage in pleasure-induced behaviors. The result of these lesion studies is to functionally decompose reward-related mesolimbic dopamine operations into those governing certain 'liking' properties of hedonic affect, and those governing 'wanting' properties of incentive salience which can be dissociated from one another. This dissociation is also consistent with evidence distinguishing dopamine's role in appetitive versus consummatory behavior [Robbins and Everitt, 1996, 233].

8.3 From reward back to the complexities of motivation

Research into the production of reward-related phenomena by brain reward circuitry exemplifies how researchers were able to get a handle on a seemingly unwieldly, more complex phenomena such as motivation. By narrowing the context of research to modulation of NAc shell processes by VTA dopamine neurons, deemphasizing the utility of laws in explanation, using decomposition and localization strategies to identify parts and operations at successively lower levels, etc., researchers were able to make inroads into this perplexed area of research. Much of this progress was also born out of better tools and techniques for investigating the structure and function of hierarchical systems (mesolimbic, mesostriatal, mesocortical, etc.).

Yet the modulatory role of dopamine is much more complicated than simply providing a catch-all reward signal. Increasingly finer-grained mechanistic explanations will eventually specify just how complicated the picture actually is. But even then, dopamine is only a small piece of the puzzle: a comprehensive and complete mechanistic story of dopamine would not thereby yield a complete mechanistic explanation of reward, much less motivation. The main reason is simply that dopamine transmission is only causally efficacious in the context of larger subsystems, systems, mechanisms, and circuits working together; for example, the mesostriatal, mesolimbic, and mesocortical systems are organized in particular ways such that they comprise an extremely complex mesocorticolimbic system, in which many other factors besides dopamine turn out to be extremely important (e.g., neuropeptides, VP AMPA-Kainate expression, frontal lobe integrity). (In that sense, motivation mechanisms might share some similarities to Hooker's 'virtual governor', insofar as different mechanisms come together to produce and regulate certain phenomena that would be otherwise uncontrollable.)

Only in the interaction of these mechanistic systems does one begin to "see the forest for the trees" — a point continually emphasized across the various reflections of many of the main scientific players. For instance, in their review, Robbins and Everitt [1996] wrote, "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" [1996, 228]. Berridge and Robinson concur: "[F]urther advances will require equal sophistication in parsing reward into its specific psychological components" [2003, 507]. And in his review of concepts of MOTIVATION, Berridge concludes that higher-level motivation research is necessary to make sense of how the systemic interactions of neuroanatomical structures and neurochemical signals, mechanisms of protein folding, monoamine production, gene transcription, etc. realize psychological phenomena and produce behavior [2004, 205]. Others are explicit about the immediate need for systems-level functional neuroimaging results, which can place the specific components into the broader context of overall brain function [Robbins and Everitt, 1996, 233]. 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, ascending across levels is equally important as descent.

Hence, when Berridge [2003a; 2003b] and others explain motivational states in terms of attributions of the 'wanting' component of reward-related processes that transform perceptual representations into desired incentives for action, and in a way that is independent of hedonic valence, their explanations implicitly invoke models of mechanisms that exhibit much greater organizational complexity. Models of phasic bursting mechanisms of mesolimbic dopamine would fail to be pitched at the *appropriate* psychological level (i.e., in terms of transformation of representations, personal versus subpersonal systems). And indeed, a full explanation of motivation itself — especially beyond that of immediate attributions of incentive salience — must eventually involve models of mechanistic systems governing the production of planning and decision-making, the regulation of emotion and long-term memory, creativity, social role formation, and so forth [Franken, 2003; Ikemoto and Panksepp, 1999]. In sum, explaining motivation mechanistically requires illuminating the organizational collusion and interaction of these various composite systems that engage their environment at increasingly higher levels.

9 SUMMARY

The foregoing sketch of research on motivation and reward exemplifies what is now common explanatory practice in psychology — decomposing a composite, hierarchically organized system into its component parts and operations and then constructing models that abet scientific understanding of how they might be organized so as to comprise the mechanism's activity. Rather than its subsumption under sets of laws, such models feature in narratives about how a mechanism might be directly responsible for the phenomenon. As we noted at the beginning of the chapter, this explanatory practice entered science with the contributions of investigators such as Galileo, Descartes, and Boyle, and gradually became more common. Its broad acceptance in psychology was ushered in by the development of the information-processing perspective, which suggested that new understanding of how complex mechanisms work could help explain important features of psychological phenomena.

After noting this confluence between mechanistic approaches and information-processing perspectives in psychology, we turned to the task of explicating what mechanisms and mechanistic explanations are, drawing upon research from philosophers primarily focused on biology. We argued that explanation is inherently an epistemic or cognitive activity; so rather than misconstruing mechanistic explanation as ontic and nomological explanation as epistemic, what is needed is the development of the appropriate epistemic account of mechanistic explanation — one that focuses on how investigators reason with the models and representations of mechanisms. Such representation and reasoning often involves graphical representations or simulations of operations, not just linguistic representations of laws and initial conditions and deductive inferences.

Mechanistic approaches also reconfigure a number of issues in the philosophy of psychology beyond that of explanation. We have considered two: the question of reductionism, and the question of scientific discoveries. Mechanistic explanation is partially reductionistic, in the sense that it appeals to lower-level parts and their operations in explaining why a mechanism behaves as it does; but mechanistic explanation is not reductionistic in the sense of deriving higher-level theories from lower-level ones, nor in the sense of supplanting explanations of causal processes at higher levels, where the mechanism as a whole engages other entities in its environment. Causal processes at each level are different, and the ultimate result of a mechanistic account is an interfield theory that bridges levels. As to the question of scientific discoveries, mechanistic approaches are particularly apt for analyzing them, despite a tradition in philosophy of science that limits philosophy to characterizing justification of already discovered laws and disavows any prospect of contributing to the understanding of discovery. In particular, philosophers are engaged in articulating heuristics such as decomposition and localization, identifying what different investigatory techniques contribute to discovering components and operations, and understanding how scientists have discovered different modes of organization found in mechanisms, characterized their significance, and articulated relations between phenomena at different levels of organization.

BIBLIOGRAPHY

[Abrahamsen, 1987] A. A. Abrahamsen. Bridging boundaries versus breaking boundaries: Psycholinguistics in perspective. *Synthese, 72*, 355-388, 1987.

[Atkinson, 1957] J. W. Atkinson. Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359-372, 1957.

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

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

[Bechtel, 1995] W. Bechtel. Biological and social constraints on cognitive processes: The need for dynamical interactions between levels of inquiry. *Canadian Journal of Philosophy, 20*, S133-S164, 1995.

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

[Bechtel, 2001b] W. Bechtel. Decomposing and localizing vision: An exemplar for cognitive neuroscience. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 225-249). Oxford: Basil Blackwell, 2001.
[Bechtel, 2002] W. Bechtel. Decomposing the mind-brain: A long-term pursuit. *Brain and Mind, 3*, 229-242, 2002.
[Bechtel, 2006] W. Bechtel. *Discovering cell mechanisms: The creation of modern cell biology.* Cambridge: Cambridge University Press, 2006.
[Bechtel, in press] W. Bechtel. Biological mechanisms: Organized to maintain autonomy. In F. Boogerd *et al.*, eds *Systems Biology: Philosophical Foundations*, New York: Elsevier, in press.
[Bechtel and Abrahamsen, 2005] W. Bechtel and A. Abrahamsen. Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421-441m, 2005.
[Bechtel *et al.*, 1998] W. Bechtel, A. Abrahamsen, and G. Graham. The life of cognitive science. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 1-104). Oxford: Basil Blackwell, 1998.
[Bechtel and Hamilton, in press] W. Bechtel and A. Hamilton. Reduction, integration, and the unity of science: Natural, behavioral, and social sciences and the humanities. In T. Kuipers (Ed.), *Philosophy of science: Focal issues.* New York: Elsevier, in press.
[Bechtel and Richardson, 1993] W. Bechtel and R. C. Richardson. *Discovering complexity: Decomposition and localization as strategies in scientific research.* Princeton, NJ: Princeton University Press, 1993.
[Bernard, 1865] C. Bernard. *An introduction to the study of experimental medicine.* New York: Dover, 1865.
[Berridge, 2003a] K. C. Berridge. Comparing the emotional brain of humans and other animals. In R. J. Davidson, K. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 25-51). New York: Oxford University Press, 2003.
[Berridge, 2003b] K. C. Berridge. Pleasures of the brain. *Brain and Cognition, 52*, 106-128, 2003.
[Berridge, 2004] K. C. Berridge. Motivation concepts in behavioral neuroscience. *Physiology and Behavior, 81*, 179-209, 2004.
[Berridge and Robinson, 1998] K. C. Berridge and T. E. Robinson. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*, 309-369, 1998.
[Bickle, 1998] J. Bickle. *Psychoneural reduction: The new wave.* Cambridge, MA: MIT Press, 1998.
[Bickle, 2003] J. Bickle. *Philosophy and neuroscience: A ruthlessly reductive account.* Dordrecht: Kluwer, 2003.
[Bielajew and Harris, 1991] C. H. Bielajew and T. Harris. Self-stimulation: A rewarding decade. *Journal of Psychiatry and Neuroscience, 16*, 109-114, 1991.
[Broca, 1861] P. Broca. 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, 1861.
[Brodmann, 1909/1994] K. Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde* (L. J. Garvey, Trans.). Leipzig: J. A. Barth, 1909/1994.
[Brown, 1961] J. S. Brown. *The motivation of behavior.* New York: McGraw Hill, 1961.
[Cannon, 1929] W. B. Cannon. Organization of physiological homeostasis. *Physiological Reviews, 9*, 399-431, 1919.
[Chefer *et al.*, 2005] V. I. Chefer, T. Czyzyk, E. A. Bolan, J. Moron, J. E. Pintar and T. S. Shippenberg. Endogenous kappa-opioid receptor systems regulate mesoaccumbal dopamine dynamics and vulnerability to cocaine. *Journal of Neuroscience, 25*, 5029-5037, 2005.
[Chomsky, 1965] N. Chomsky. *Aspects of a theory of syntax.* Cambridge, MA: MIT Press, 1965.
[Churchland, 1986] P. S. Churchland. *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* Cambridge, MA: MIT Press, 1986.
[Churchland and Sejnowski, 1992] P. S. Churchland and T. J. Sejnowski. *The computational brain.* Cambridge, MA: MIT Press, 1992.
[Clement, 2003] J. J. Clement. Imagistic simulation in scientific model construction, *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, p. 25. Mahwah, NJ: Erlbaum, 2003.

[Craver, 2001] C. Craver. Role, mechanisms, and hierarchy. *Philosophy of Science, 68*, 53-74, 2001.
[Craver, in press] C. Craver. *Explaining the Brain: What a Science of the Mind-brain Could Be.* New York: Oxford University Press, in press.
[Craver and Bechtel, in press] C. Craver and W. Bechtel. Explaining top-down causation (away) *Biology and Philosophy*, in press.
[Cummins, 2000] R. Cummins. "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117-144). Cambridge, MA: MIT Press, 2000.
[Darden, 1990] L. Darden. Diagnosing and fixing faults in theories. In J. Shrager & P. Langley (Eds.), *Computational Models of Scientific Discovery and Theory Formation* (pp. 319-353). San Mateo, CA: Morgan Kaufmann, 1990.
[Darden, 1991] L. Darden. *Theory change in science: Strategies from Mendelian genetics.* New York: Oxford University Press, 1991.
[Darden and Maull, 1977] L. Darden and N. Maull. Interfield theories. *Philosophy of Science, 43*, 44-64, 1977.
[Descartes, 1637] R. Descartes. *Discours de la méthode pour bien conduire sa raison & chercher la varité dans les sciences.* Leyden: I. Maire, 1637.
[Descartes, 1644] R. Descartes. *Principia philosophiae.* Amsterdam: Apud Ludovicum Elzevirium, 1644.
[Descartes, 1658] R. Descartes. *Meditationes de prima philosophia.* Amsterdam: J. Janssonium Juniorum, 1658.
[Descartes, 1664] R. Descartes. *Traite de l'Homme.* Paris: Angot, 1664.
[Di Chiara, 1999] G. Di Chiara. Drug addiction as dopamine-dependent associative learning disorder. *European Journal of Pharmacology, 375*, 13-30, 1999.
[Donders, 1868] F. C. Donders. Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool: 1868-1869. *Tweede Reeks, 2*, 92-120, 1868.
[Endicott, 1998] R. P. Endicott. Collapse of the new wave. *Journal of Philosophy, 95*, 53-72, 1998.
[Endicott, 2001] R. P. Endicott. Post-structuralist angst: A critical notice of Bickle's *Psychoneural reduction: The new wave. Philosophy of Science, 68*, 377-393, 2001.
[Feinberg and Farah, 2000] T. E. Feinberg and M. J. Farah. A historical perspective on cognitive neuroscience. In M. J. Farah & T. E. Feinberg (Eds.), *Patient-based approaches to cognitive neuroscience* (pp. 3-20). Cambridge, MA: MIT Press, 2000.
[Flourens, 1846] M. J. P. Flourens. *Phrenology examined* (C. D. L. Meigs, Trans.). Philadelphia: Hogan and Thompson, 1846.
[Franken, 2003] I. H. A. Franken. Drug craving and addiction: Integrating psychological and neuropsychopharmacological approaches. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 27*, 563-579, 2003.
[Freud, 1922] S. Freud. *Beyond the pleasure principle.* London: The International Psycho-Analytical Press, 1922.
[Garber, 2002] D. Garber. Descartes, mechanics, and the mechanical philosophy. *Midwest Studies in Philosophy, 26*, 185-204, 2002.
[Giere, 1988] R. G. Giere. *Explaining science: A cognitive approach.* Chicago: University of Chicago Press, 1988.
[Giere, 1999] R. G. Giere. *Science without laws.* Chicago: University of Chicago Press, 1999.
[Glennan, 1996] S. Glennan. Mechanisms and the nature of causation. *Erkenntnis, 44*, 50-71, 1996.
[Glennan, 2002] S. Glennan. Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342-S353, 2002.
[Hardcastle, 1996] V. G. Hardcastle. *How to build a theory in cognitive science.* Albany, NY: SUNY Press, 1996.
[Harley, 1879] R. Harley. The Stanhope demonstrator. *Mind, 4*, 192-210, 1879.
[Harré, 1960] R. Harré. Metaphor, models, and mechanism. *Proceedings of the Aristotelian Society, 60*, 101-122, 1960.
[Hegarty, 2002] M. Hegarty. Mental visualization and external visualizations. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, (p. 40). Mahwah, NJ: Erlbaum, 2002.

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

[Hempel, 1958] C. G. Hempel. 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, 1958.

[Hempel, 1948] C. G. Hempel and P. Oppenheim. Studies in the logic of explanation. *Philosophy of Science, 15*, 137-175, 1948.

[Holland *et al.*, 1987] J. H. Holland, K. J. Holyoak, R. E. Nisbett and P. R. Thagard. *Induction: Processes of Inference, Learning and Discovery.* Cambridge, MA: MIT Press, 1987.

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

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

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

[Ikemoto *et al.*, 2005] S. Ikemoto, M. Qin, Z.-H. Liu. The functional divide for primary reinforcement of D-amphetamine lies between the medial and lateral ventral striatum: Is the division of the accumbens core, shell, and olfactory tubercle valid? *Journal of Neuroscience, 25*, 5061-5065, 2005.

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

[Jackson, 1931] J. H. Jackson. *Selected writings of John Hughlings Jackson, vol. 1.* London: Hodder and Stoughton, 1931.

[James, 1890/1950] W. James. *Principles of psychology.* New York: Dover, 1890/1950.

[Jevons, 1982] W. S. Jevons. On the mechanical performance of logical inference. *Philosophical Transactions, 160*, 497-518, 1982.

[Kahneman *et al.*, 1982] D. Kahneman, P. Slovic and A. Tversky, eds. *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press, 1982.

[Kauffman, 1971] S. A. Kauffman. Articulation of parts explanations in biology and the rational serach for them. In R. C. Bluck & R. S. Cohen (Eds.), *PSA 1970* (pp. 257-272). Dordrecht: Reidel, 1971.

[Kauffman, 1993] S. A. Kauffman. *The origins of order: Self-organization and selection in evolution.* Oxford: Oxford University Press, 1993.

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

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

[Koob and Le Moal, 1997] G. F. Koob and M. Le Moal. Drug abuse: Hedonic homeostatic dysregulation. *Science, 278*, 52-58, 1997.

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

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

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

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

[Machamer *et al.*, 2000] P. Machamer, L. Darden, and C. Craver. Thinking about mechanisms. *Philosophy of Science, 67*, 1-25, 2000.

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

[Marquand, 1885] H. G. Marquand. A new logic machine. *Proceedings of the American Academy of Arts and Sciences, 21*, 303, 1885.

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

[McCauley, 1987] R. N. McCauley. The not so happy story of the marriage of linguistics and psychology: or why linguistics has discouraged psychology's recent advances. *Synthese, 72*, 341-353, 1987.

[McCauley, 1996] R. N. McCauley. Explanatory pluralism and the coevolution of theories in science. In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 17-47). Oxford: Blackwell, 1996.

[McCauley, 2001] R. N. McCauley and W. Bechtel. Explanatory pluralism and the heuristic identity theory. *Theory and Psychology, 11*, 736-760, 2001.

[McDougall, 1908] W. McDougall. *An introduction to social psychology.* London: Methuen, 1908.

[Miller, 1962] G. A. Miller. Some psychological studies of grammar. *American Psychologist, 17*, 748-762, 1962.

[Miller *et al.*, 1960] G. A. Miller, E. Galanter, and K. Pribram. *Plans and the structure of behavior.* New York: Holt, 1960.

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

[Morrison and Morrison, 1961] P. Morrison and E. Morrison, eds. *Charles Babbage and his calculating engines: Selected writings by Charles Babbage and others.* New York: Dover, 1961.

[Murray *et al.*, 1938] H. A. Murray, W. G. Barrett, E. Homburger, and others. *Explorations in personality: A clinical and experimental study of fifty men of college age, by the workers at the Harvard psychological clinic.* New York: Oxford, 1938.

[Nagel, 1961] E. Nagel. *The structure of science.* New York: Harcourt, Brace, 1961.

[Nersessian, 1999] N. Nersessian. 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, 1999.

[Nersessian, 2002] N. Nersessian. 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, 2002.

[Newell and Simon, 1972] A. Newell and H. A. Simon. *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall, 1972.

[Nickles, 1980a] T. Nickles, ed. *Scientific discovery: Case studies.* Dordrecht: Reidel, 1980.

[Nickles, 1980b] T. Nickles, ed. *Scientific discovery: Logic and rationality.* Dordrecht: D. Reidel Publishing Company, 1980.

[Olds and Milner, 1954] J. Olds and P. Milner. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47*, 419-429, 1954.

[Oppenheim and Putnam, 1958] P. Oppenheim and H. Putnam. 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, 1958.

[Peirce, 1887] C. S. Peirce. Logical machines. *American Journal of Psychology, 1*, 165, 1887.

[Petersen *et al.*, 1989] S. E. Petersen, P. T. Fox, M. I. Posner, M. Mintun, and M. E. Raichle. Positron emission tomographic studies of the processing single words. *Journal of Cognitive Neuroscience, 1*, 153-170, 1989.

[Petersen *et al.*, 1988] S. E. Petersen, P. T. Fox, M. I. Posner, M. Mintun, and M. E. Raichle. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*, 585-588, 1988.

[Posner and Raichle, 1994] M. I. Posner and M. E. Raichle. *Images of Mind.* San Francisco: Freeman, 1994.

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

[Railton, 1978] P. Railton. A deductive-nomological model of probabilistic explanation. *Philosophy of Science, 45*, 206-226, 1978. Reprinted in J. A. Cover (Ed.), *Philosophy of science: The central issues* (pp. 746-765). New York: W. W. Norton and Company, 1998.

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

[Reichenbach, 1938] H. Reichenbach. *Experience and prediction.* Chicago: University of Chicago Press, 1938.

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

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

[Ruiz-Mirazo and Moreno, 2004] K. Ruiz-Mirazo and A. Moreno. Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*, 235-259, 2004.
[Salmon, 1984] W. C. Salmon. *Scientific explanation and the causal structure of the world.* Princeton: Princeton University Press, 1984.
[Schaffner, 1967] K. Schaffner. Approaches to reduction. *Philosophy of Science, 34*, 137-147, 1967.
[Schouten and de Jong, 1999] M. Schouten and H. Looren de Jong. Reduction, elimination, and levels: The case of the LTP-learning link. *Philosophical Psychology, 12*, 237-262, 1999.
[Schroeder, 2004] T. Schroeder. *The three faces of desire.* Oxford: Oxford University Press, 2004.
[Shannon, 1948] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-656, 1948.
[Shannon and Weaver, 1949] C. Shannon and W. Weaver. *The mathematical theory of communication.* Urbana, IL: University of Illinois Press, 1949.
[Simon, 1969] H. A. Simon. *The sciences of the artificial.* Cambridge, MA: MIT Press, 1969.
[Sternberg, 1966] S. Sternberg. High-speed scanning in human memory. *Science, 153*, 652-654, 1966.
[Taha and Fields, 2005] S. A. Taha and H. L. Fields. Encoding of palatability and appetitive behaviors by distinct neuronal populations in the nucleus accumbens. *Journal of Neuroscience, 25*, 1193-1202, 2005.
[Talmy, 2000] L. Talmy. The semantics of causation. In his *Toward a Cognitive Semantics, vol. 1*, pp. 471-549. Cambridge, MA: MIT Press, 2000.
[Thagard, 2003] P. Thagard. Pathways to biomedical discovery. *Philosophy of Science, 70*, 235-254, 2003.
[Titchner, 1907] E. Titchner. *An outline of psychology* (Revised and enlarged ed.). New York: MacMillan, 1907.
[Turing, 1936] A. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 42*, 230-265, 1936.
[van Essen and Gallant, 1994] D. C. van Essen and J. L. Gallant. Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*, 1-10, 1994.
[van Fraassen, 1989] B. van Fraassen. *Laws and symmetries.* Oxford: Oxford University Press, 1989.
[von Eckardt and Poland, 2004] B. von Eckardt and J. S. Poland. Mechanistic and explanation in cognitive neuroscience. *Philosophy of Science, 71*, 972-984, 2004.
[Waskan, forthcoming] J. Waskan. *Models and cognition*, forthcoming.
[Watson, 1913] J. B. Watson. Psychology as the behaviorist views it. *Psychological Review, 20*, 158-177, 1913.
[Weiner, 1986] B. Weiner. *An attributional theory of motivation and emotion.* New York: Springer Verlag, 1986.
[Wernicke, 1874] C. Wernicke. *Der aphasische Symptomenkomplex: eine psychologische Studie auf anatomischer Basis.* Breslau: Cohn and Weigert, 1874.
[Wiener, 1948] N. Wiener. *Cybernetics: Or, control and communication in the animal machine.* New York: Wiley, 1948.
[Wilson, 2003] R. A. Wilson. Pluralism, entwinement, and the levels of selection. *Philosophy of Science, 70*, 531-552, 2003.
[Wimsatt, 1974] W. C. Wimsatt. Complexity and organization. In K. F. Schaffner & R. S. Cohen (Eds.), *PSA 1972* (pp. 67-86). Dordrecth: Reidel, 1974.
[Wimsatt, 1976] W. C. Wimsatt. Reductionism, levels of organization, and the mind-body problem. In G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry* (pp. 202-267). New York: Plenum Press, 1976.
[Wimsatt, 1986] W. C. Wimsatt. Forms of aggregativity. In A. Donagan, N. Perovich, & M. Wedin (Eds.), *Human nature and natural knowledge* (pp. 259-293). Dordrecht: Reidel, 1986.
[Wimsatt, 1994] W. C. Wimsatt. The ontology of complex systems: Levels, perspectives, and causal thickets. *Canadian Journal of Philosophy, 20*, S207-S274, 1994.
[Wimsatt, 1997] W. C. Wimsatt. Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science, 64*, S372-S384, 1997.
[Wise, 2004] R. A. Wise. Dopamine, learning and motivation. *Nature Reviews Neuroscience, 5*, 483-494, 2004.

[Wong, 2000] R. Wong. *Motivation: A biobehavioral approach.* Cambridge: Cambridge University Press, 2000.
[Woodward, 2002] J. Woodward. What is a mechanism? A counterfactual account. *Philosophy of Science, 69*, S366-S377, 2002.
[Wright, 2000] C. D. Wright. Eliminativist undercurrents in the new wave model of psychoneural reduction. *Journal of Mind and Behavior, 21*, 413-436, 2000.
[Wright, 2002] C. D. Wright. Animal models of depression in neuropsychopharmacology qua Feyerabendian philosophy of science. In S. P. Shohov (Ed.), *Advances in Psychology Research, vol. 13*, pp. 129-148). New York: NovaScience Publishers, 2002.
[Wright, forthcoming] C. D. Wright. Is psychological explanation going extinct? In M. Schouten and H. Loren de Jong (Eds.), *The Matter of the Mind: Philsophical Essays on Psychology, Neuroscience, and Reduction.* Oxford: Blackwell, forthcoming.
[Young, 1970] 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.