This is a draft manuscript. Comments are most welcome,
but please do not quote without permission.

Lecture 1
Introducing a Philosophical Stance: Naturalism and Mechanism

The brain has long been an object of interest and fascination for philosophers. Some philosophers, such as Descartes, even offered proposals as to how it worked. But philosophers, using the tools of their profession, do not have the ability themselves to investigate the operation of the brain any more than they have tools to investigate the cell, the atom, or the solar system. With respect to these other domains of phenomena, philosophers have made their contribution by engaging the practitioners of the relevant sciences who do have the tools to conduct the appropriate investigations of the phenomena. Thus, in the last part of the 20th century philosophy of science has largely transformed into the philosophy of particular sciences. This explains the term "engages" in the title of these lectures, which are intended both to characterize and exemplify the manner in which philosophy is currently engaging the cognitive neurosciences.

The focus of this first lecture is on the stance from which I maintain philosophy ought to pursue the engagement, one I characterize in terms of naturalism and mechanism. Before turning to explaining this stance, though, I need to pinpoint a bit more the sciences to be engaged. Let's start with neuroscience. Although the sciences of the brain have been developing over the past several centuries, and major advances were made in the understanding of the brain in the 19th century and the first half of the 20th century, the research efforts were distributed over a number of disciplines, including anatomy, physiology, biochemistry, genetics, psychology, and psychiatry. Energized by the efforts of Francis Schmitt[1] and inspired by the success of the Neurosciences Research Program which he launched in 1962, *neuroscience* came to be recognized as a distinct and integrated field of inquiry. (As a indication of the scope of the endeavor, the Society for Neuroscience was founded in 1970 with 500 members and now has more than 30,000 members and is recognized as the principal professional organization in the neurosciences..)

During the last part of the 20th century neuroscientific inquiry made major advances in understanding the brain, especially at the cellular and molecular level. But neuroscience has not just focused on these nuts and bolts of the cognitive machinery, but on how the neural system as a whole supports animals in their engagement with their environment. Within neuroscience, this effort is often referred to as *behavioral neuroscience*. During the same period in which neuroscience was developing, another interdisciplinary activity was taking shape, which came to

---

[1] For Schmitt, this effort was a reprise of his previous endeavor of discipline creation on behalf of biophysics for which, in 1958 he had organized the four week long summer program entitled Intensive Study Program in Biophysics at the University of Colorado, Boulder. This brought together 200 biologists, chemists, physicists, psychologists, and engineers and culminated in the publication of *Biophysical Science: A Study Program* that helped to delineate the scope and chart the research agenda for the new enterprise.

be referred to as *cognitive science* (Bechtel, Abrahamsen, & Graham, 1998). The central disciplines in this endeavor were psychology, computer science (artificial intelligence), and linguistics, with philosophy and anthropology playing less central roles. Although the study of the brain was often regarded as pertinent to cognitive functioning, the major research programs in cognitive science proceeded independently of investigations into the brain. Sometimes this independence was encouraged by theoretical arguments, usually supplied by philosophers (Fodor, 1974; Putnam, 1967). However, the primary motivation for pursuing cognitive inquiry independently of neuroscience was that the tools for relating the processes identified in cognitive science accounts with brain processes were not available. Rather, drawing inspiration from formal information theory, and relying on behavioral measures such as error patterns or reaction times, these disciplines developed information processing models of cognitive processes. As I noted, philosophers have been a part of the cognitive science endeavor, and their involvement represents the sort of engagement exemplified in the engagement of philosophy with other disciplines, both seeking to understand the activity of cognitive science and sometimes contributing to it at the theoretical level.

One of the major factors that separated cognitive science from behavioral neuroscience is that the research tools available in neuroscience typically could not be applied to humans (one of the main exceptions was the study of naturally occurring lesions). These tools, most notably recording electrical impulses from individual neurons, were invasive in ways judged to be ethical objectionable with humans. But in the 1980s a new approach to studying neural activity, neuroimaging, provided not only the opportunity to study brain processes with little or no invasion but also the opportunity to study brain processes relevant to cognitive activity. Inspired in significant part by the availability of functional neuroimaging, a new interdisciplinary endeavor, cognitive neuroscience, began to develop. Increasingly, philosophers with sufficient familiarity with the brain and cognitive research are beginning to engage this new interdisciplinary pursuit.

With this exposition of the title for these lectures, let me turn more directly to philosophy, and the challenge for this lecture of articulating the philosophical stance I will be adopting in these lectures. Methodologically, philosophers can be divided roughly into two communities. Some practice philosophy as an *a priori* endeavor which tries to access its own body of truths. In philosophy of science this approach is exemplified by the attempt to offer accounts of the structure of scientific theories, of scientific explanation, and of the evidentiary relations in science in terms of logic. To the degree those who practice this approach engage the sciences, they do so to show applications of philosophical accounts and, potentially, to offer normative advice to scientists (your theories ought to be characterized as axiomatizeable systems or your search for evidence ought to take the form of attempting to falsify your hypotheses). Naturalism is an alternative to this *a priori* approach to philosophy. Instead of restricting himself or herself to *a priori* methods, the naturalist adopts an *a posteriori* stance. But lacking his or her own unique methods for discovering *a posteriori* truths, the naturalist in philosophy happily draws upon those discovered by others (or those he or she can discover using the same methods). Roughly, the naturalistic method in philosophy of science is to treat the relevant empirical sciences as the subject matter of inquiry and to construe the task of philosophy as an attempt principally to understand how these sciences function: what is offered as a theory in these sciences, what counts as explanation and what is accepted as adequate evidence. If the naturalist is to make normative judgments about science, they will be based on the understanding of the enterprise itself and consist of showing how, for example, certain strategies have or have not succeeded in other scientific endeavors. In the next section I will develop the naturalist's perspective in greater detail.

One consequence of adopting the naturalist perspective is the recognition that not all sciences are the same. Many of the traditional philosophical ideas about science, whether developed by philosophers pursuing the *a priori* approach or the naturalist approach, were most applicable to domains of classical physics. But, starting in the 1970s and 1980s, a number of philosophers who turned their attention to the biological sciences found that these frameworks did not apply all that well to different biological domains. Recognizing this, the naturalist is committed to developing accounts that work for specific sciences, postponing the question of determining what is in common between all sciences. In this sense, the naturalist is led to be a pluralist about sciences.[2]

At the center of the naturalistic account of the cognitive neurosciences that I will be offering in these lectures is an understanding of the scientist as engaged in the quest to understand the mechanism responsible for a particular phenomenon. The term *mechanism* is widely used, not just in the cognitive neuroscience, but across the life sciences. Although the *mechanical philosophy* has deep roots in 17th and 18th century science, the particular version of mechanism it espoused is rightly recognized as inadequate. Until very recently, the notion of mechanism as it figures in the contemporary life sciences has not been much analyzed. Accordingly, the second part of this lecture will develop the necessary account of mechanism.

1. Naturalism

Although a naturalistic approach to philosophy is identifiable in many figures in the history of philosophy, its modern incarnation can be traced to W. V. O. Quine, who contrasted naturalized epistemology to what he termed *first philosophy*. According to *first philosophy*, the goal of epistemology is to answer the skeptic by showing how, from indubitable foundations, we can derive our knowledge of the world via valid arguments. Quine maintained that no derivations could take us from stimulation of sensory receptors to knowledge of things in the world. But nonetheless, with nothing else to go on but sensory stimulation, we somehow arrive at what we take to be knowledge about the world. Since the process is not derivation, Quine seems to be

---

[2] This does not entail abandoning the unity of science project if that means revealing how different sciences relate to one another. Even if the different sciences pursue their investigations differently, there may well be ways in which they relate to each other. In fact, given the emergence of broad integrating research programs in science that involve practitioners from domains of the physics, chemical, biological, and psychological sciences, one would expect that the naturalist should be able to discover means of integration amongst the sciences.

repudiating the project of epistemology. But in fact he advances an alternative project—determine how humans do construct their understanding of the world from their sensory stimulations. “Why not settle for psychology?”, he asks (Quine, 1969, p. 75). If we take this as the project, then the epistemological inquiry itself becomes a scientific inquiry into the processes by which subjects acquire knowledge, an inquiry within the scope of psychology. Quine maintains, therefore, that “Epistemology in its new setting . . . is contained in natural science, as a chapter of psychology” (p. 83).[3]

While situating epistemology within psychology, Quine also tries to maintain that psychology and natural science fall within the scope of epistemology:

> But the old containment [of natural science by epistemology] remains valid too, in its own way. We are studying how the human subject of our study posits bodies and projects his physics from his data, and we appreciate that our position in the world is just like his. Our very epistemological enterprise, therefore, and the psychology wherein it is a component chapter, and the whole of natural science wherein psychology is a component book—all of this is our own construction or projection from stimulations like those we are meting out to our epistemological subject (p. 83).

But Quine’s naturalism does not make much of this later containment. One of the main objectives of traditional epistemology, which took the containment of natural science by epistemology as central, was to advance a normative claim about what ought to count as knowledge. An important question for the naturalist is whether the containment Quine contemplates allows one to advance any normative claims. To this I will return.

Because they want to address these normative questions, many who have pursued the naturalization of epistemology have pursued a program much weaker than Quine’s. While not subjugating epistemology to natural science, they advocate invoking the results of natural science in the pursuit of answering epistemology’s normative questions. Since the fundamental activity in question is how to *reason* from evidence to knowledge, the relevant empirical information is to come from empirical inquiries in psychology about how we reason. One of the noteworthy features of psychological inquiries of this kind is that they have tended to focus on

---

[3] In a widely quoted passage, he explains: “Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input – certain patterns of irradiation in assorted frequencies, for instance –a and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology: namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence...But a conspicuous difference between old epistemology and the epistemological enterprise in this new psychological setting is that we can now make free use of empirical psychology” (pp. 82-3).

how we reason badly, for example, how we ignore base rates and thereby attribute too much weight to individual pieces of evidence or how we are taken in by visual illusions. Although results of this type in psychology are evidence about the challenges confronting human reason, the very fact that they are taken to indicate *failures* of human reason raises questions as to their value in addressing epistemic judgments about how we ought to reason so as to acquire knowledge.

Quine's focus was broadly epistemological and although he made claims about scientific practice (e.g., about the role of non-epistemic factors such as simplicity in the evaluation of scientific theories) and about the limitations on scientific knowledge (e.g., the underdetermination of theories by all possible evidence), he generally did not engage science as a naturalist. For example, he did not study the investigatory strategies or reasoning employed by particular scientists or the specific explanations they advanced. But in the thirty years since Quine advanced the project of naturalized epistemology, numerous philosophers of science have taken up just this endeavor (for early developments of naturalism in philosophy of biology, see Callebaut, 1993). Many of these efforts involve examining the reasoning and inferential practices of particular scientists, although often with the goal of elaborating more general strategies. For example, Nancy Nersessian has examined in detail the reasoning of physicists Faraday and Maxwell, especially their use of diagrams and analogy in developing explanatory models (Nersessian, 2002). Ronald Giere has not only examined how scientists construct models for particular reasoning projects but also how the epistemic efforts are distributed across both multiple scientists and the instruments they employ (Giere, 1988, 2002). Lindley Darden has focused on how scientists revise theories by localizing problems and employing specific repair strategies (Darden, 1991).

In addition to the theoretical reasoning of scientists, philosophers pursuing a naturalistic approach have also focused on the reasoning involved in the design and evaluation of new instruments and research techniques. One of the pioneering efforts of this sort was Ian Hacking's examination of the microscope and the manner in which scientists come to have confidence that they are confronting real phenomena, not just an artifact produced by their instrument (Hacking, 1983). Hacking emphasized the importance of concurrence of results from new modes of intervention (e.g., microscopes working on different principles or different modes of staining specimens) with existing techniques. In my own work on what I term *the epistemology of evidence* such concurrence actually plays a relatively minor role, as older techniques serve more to guide the calibration of new investigative strategies than to evaluate them (Bechtel, 2002). It appears that often in actual science one of the chief criteria for the acceptability of evidence from a new technique is whether it produces information that can be fit into an acceptable theoretical framework (Bechtel, 2000). This, of course, reverses the usual foundationalist perspective of focusing on how theories are evaluated by evidence, but is an example of the revisions in our account of science that are prompted by a naturalistic inquiry into science.

Naturalized philosophers are not alone in examining the actual practices of science. Historians, sociologists, and psychologists often pursue similar objectives. This broader, interdisciplinary inquiry into the activity of science is often referred to as *science studies*. In many cases it is difficult to draw any strong distinctions between the pursuits of philosopher and other practitioners of science studies, although philosophers tend to be characterized by a focus on epistemology and the question of how the activities of scientists can be understood as productive of knowledge. Moreover, many retain the normative aspirations of earlier epistemology. That is, they hope that an understanding of how scientific knowledge is actually procured, it will be possible to make recommendations for future practice. Of course insofar as any such recommendations are based on understanding what has and has not worked well in previous science, the force of these recommendations is limited. We cannot guarantee that a strategy of reasoning that worked in the past will work for new problems. But in noting this I am merely noting that a naturalized pursuit cannot rise above its own naturalism.

One of the most jarring results of adopting a naturalistic perspective and focusing on the life sciences is that in many parts of biology one seems to look in vain for what philosophy has commonly taken to be the principal explanatory tool of science, laws. The few things that have been called laws in biology and psychology, such as Mendel's laws, are often recognized to be incorrect, but they are generally not replaced by more adequate statements of the laws. Some philosophers have offered explanations of the failure to find laws—laws are supposed to be universal in scope, and biological discoveries are restricted to the specific life forms that have evolved on this planet (Sober, Beatty). But that does not mean that biologists and psychologists are not developing explanations. If one looks at what biologists and psychologists seek and find sufficient to provide explanation, it often turns out to be a mechanism. Since this notion has been insufficiently developed in the philosophical literature, I turn now to it.

2. Mechanism

The idea that explanation involves the quest to understand the mechanism responsible for the phenomenon in question is in fact an old one whose roots go back at least as far as the development of the ancient atomic theory by Leucippus and Democritus in the 5th century BCE and to the writings of Pseudo Aristotle. The central explanatory strategy of the ancient atomic theory was to explain the features of observable objects in terms of the shape and motion of their constituents. This strategy was revived in the 17th and 18th century in the works of, among others, Galileo, Descartes, Boyle, and Newton. Descartes's explanation of magnetism, in which he proposed the existence of screw shaped particles and threads in iron into which the screws were drawn, provides an exemplar of the program. The strategy was to show how natural phenomena could be accounted for in terms of the shape, size, and motion of elementary particles (Boas, 1952; Westfall, 1971).

As with many other mechanists, Descartes's models for his mechanistic explanations were the machines humans built. He says:

> 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 (Principia IV 203).

But the difference is important. Whereas the builder of human artifacts must know the nature of the building blocks before undertaking the construction of a mechanism, the mechanical scientist of the 17th century hypothesized about the features of the atoms or corpuscles which are to explain macroscopic phenomena. Moreover, often these hypotheses were guided by the effects to be explained. Moreover, in general the features of the atoms were identified differently for each phenomenon to be explained—they did not constitute a common set of components which could explain a wide range of phenomena and thereby gain credibility from the systematicity they provided in the understanding of macroscopic phenomena.

For mechanism to be more than a metaphysical commitment and to become a compelling mode of explanation, knowledge of the building blocks or better tools for determining their capacities became crucial. The mechanistic program, especially as it applied to living phenomena, took on a new form in the 19th century as advances in physiology provided building blocks which had status independent of the particular phenomena that were explained in terms of them.[4] A key step for biochemical explanations of the phenomena within cells was the development of a set of chemical reactions at the appropriate level out of which to propose mechanisms of metabolic processes.[5] With these more or less in place, the program for the mechanist was to explain effects and activities of the compounds built out of these. Moreover, as researchers pursued this project, another feature of mechanistic explanation acquired prominence. It was not sufficient just to know the elementary composition of higher level structures, but to know how they were *organized*. Understanding organization and how different modes of organization can have different effects is extremely difficult. Humans tend to conceptualize processes linearly and only when that fails recognize more complex modes of organization.

In a telling example, when Justus Leibig sat down (literally) to figure out the processes involved in animal nutrition, he identified starting points (the foodstuffs animals consume) and end points (the generation of heat or the building up of muscles in the animal body, which would then be

---

[4] The build blocks did not always have to be chemical or even microscopic. Xaiver Bichat, for example, catalogued twenty-one different types of tissues out of which he proposed that all organs of the body were constructed and articulated a program in terms of which one could explain the properties of organs in terms of the properties of the tissues out of which they were composed.

[5] Holmes summarizes this critical foundation for biochemistry: “By 1900, organic compounds relevant to metabolic processes could be characterized structurally. Characteristic groups such as hydroxyl, carboxyl, and amino groups were linked at specific sites to the carbon backbones of organic molecules. The general classes of reaction mechanisms in which each reactive group participated, and some of the influences of their relative placements on the carbon skeleton upon their reactivities, were known. The information provided strong foundations for interpreting the chemical changes linking any compounds that could be shown to take part in metabolic processes” (Holmes, 1992, p. 56).

broken down in the course of animal work) and proposed a linear series of transformations that would break down the complex compounds provided by plants through a series of reactions to generate heat and waste products (in the case of fats and carbohydrates) or to produce muscular activity and waste products (in the case of proteins). When Claude Bernard tried to establish empirically just where these reactions were performed in the animal body, he discovered that animals not only broke down sugars but also synthesized them. This discovery figured heavily in his proposal that animals were organized in terms of an internal environment and that different constituents of the organism were designed so as to carry out their activities (e.g., glycogenesis) when appropriate conditions arose in this internal environment. The introduction of an internal environment as an organizing feature of animal bodies also enabled Bernard to offer a response to mechanism's critics who charged that mechanism could never explain the apparent ability of organisms to resist the forces of the inanimate world. They did so, he argued, as a result of the various organs being both buffered from the external environment as a result of living in the internal environment and each carrying out activities that ensured the stability of this internal environment.

Discovering both the components out of which it is built and the organization that is imposed on those components are two of the major tasks in understanding a mechanism. They also are key elements in the characterization of a mechanism. I will adopt as a working characterization of a mechanism that *a mechanism is an enduring system that regularly performs some activity. It is made of component parts, each of which performs its own operation, which are then coordinated so as to accomplish the activity of the overall mechanism*.[6] There are some features to note about this characterization. First, a mechanism is identified in terms of the activity it performs (Glennan, 1996). Example activities for which mechanisms are sought are synthesizing proteins, circulating blood, visually identifying objects, and encoding episodic memories.[7] Different

---

[6] In *Discovering complexity* Richardson and I characterized mechanistic explanations as accounting "for the behavior of a system in terms of the functions performed by the parts and the interactions of these parts. . . . A mechanistic explanation identifies these parts and their organization, showing how the behavior of the machine [system] is a consequence of the parts and their organization." (Bechtel & Richardson, 1993, p. 17). More recently, Machamer, Darden, and Craver have characterized mechanisms as "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions." (Machamer, Darden, & Craver, 2000, p. 3). My main reluctance to adopting this characterization of a mechanism is that the emphasis on set-up and finish conditions helps to retain a focus on linear processes whereas mechanisms, when they are embedded in larger mechanisms, are continuously responsive to conditions in the larger mechanism. For tractability we tend to focus, like Bernard, on conditions in which their activity is triggered and their contribution to the operation of the overall mechanism, but often we have to counter this analytical perspective to appreciate the dynamics at work in the system.

[7] If a mechanism breaks and is not able to perform the activity, one may still identify it as a mechanism in terms of the activity it was designed or evolved to perform.

mechanisms may reside in the same physical entity such as the brain or the liver if that entity performs different activities. The characterization of the activity of a mechanism is critical not only for identifying it but also for investigating out how it works—the researcher is interested in those things within it that could explain that activity (and not others). This does not mean that investigators must have a fixed idea of what a mechanism does prior to their investigation. Sometimes they will arrive at the characterization of the activity of the mechanism only as their investigation proceeds or they will revise their assessment of what task it performs as they discover the responsible mechanism (Bechtel and Richardson, 1993, refer to this as *reconstituting the phenomena*). A reason why identifying the task or activity is important as it enables us to do is settle what is part of the mechanism and what is not.

Second, the characterization of a mechanism identifies three of the central tasks investigators face in coming to understand a mechanism. First, they must determine the parts of the mechanism and the operations each performs. This involves taking the system apart or *decomposing* it. Second, they must determine how these components are organized so that their activities are coordinated. Finally, they must determine the context in which the mechanism carries out this activity. Often a given mechanism in contained within a larger mechanism and there are various procedures operative in that larger mechanism that determine when the particular mechanism carries out its activity.

Since identifying the relevant components is an essential part of the process of discovering the mechanism responsible for a particular phenomena, we need to say more about what this involves. In general, there are two ways investigators can identify the components of a mechanism—they can specify their structure or they can specify what they do. For a given mechanism at a given time one of these modes of decomposition may be easier to accomplish than the other. At times, researchers may have ways of differentiating the components of a system structurally without yet knowing what they do. A clear example is found in the research at the beginning of the 20th century on mapping the brain, of which Brodmann is today the best known practitioner. Brodmann pursued his project by identifying features on which different parts of the cortex differed, such as the type of neurons present and the thickness of the six layers of which it was comprised. He identified an area as a region of cortex which was reasonably homogenous in terms of these features but which varied from the regions surrounding it (Brodmann, 1909/1994). Likewise, through development of microscopes and stains, researchers of the period were able to identify organelles in the cell such as the mitochondria and the Golgi apparatus. Although the goal of researchers identify component structures was to identify parts that would turn out to perform functions in the system, the techniques that permitting differentiation of structural components are usually not themselves able to provide information about the operations the parts perform.

At other times researchers may be able to identify the operations performed in the system without being able to identify the component structures that perform the operations. Very often these operations can be identified in terms of the transformations made in a substrate. In biochemical processes, an initial metabolite undergoes a number of operations until the final product is produced. Sometimes evidence as to the intermediate metabolites can be discovered by such techniques as blocking a particular operation in the process. In this manner the set of operations can be identified. For example, in the history of biochemistry, researchers identified the major operations involved in the process of fermentation before they were able to isolate and crystalize the responsible enzymes, the key step in specifying the structural components involved. Research in the second half of the 20th century in cognitive psychology and cognitive science likewise attempted to differentiate steps of information processing without being able to identify the brain structures responsible for these operations. This approach of decomposing an overall activity into component operations is known as *functional* analysis (Cummins, 1975, 1983) or *functional* decomposition.

Although typically either functional or structural decomposition is developed before the other, the ultimate goal is to identify structural components that perform the operations that explain the activities of the mechanism, and hence it is important to bring functional and structural decompositions into alignment. This I refer to as *localization*. One of the important benefits of localizing component operations in component structures is that one obtains independent evidence for both accounts. This is particularly important in the case of component operations. Often researchers propose operations in a mechanism hypothetically by considering what operations would suffice to do what the mechanism does. Sometimes these hypotheses are grounded in analogy with other mechanisms. But these hypotheses are fallible and the proposed functional decomposition may not identify the operations performed in the real mechanism. But if these operations are linked with structures, which can be independently investigated, then one provides important independent evidence for the proposed decomposition.

To flesh out the account of the mechanism, one needs to not only identify the components and specify the operations they perform, but also to determine how they are organized. Since components interact locally, spatial relations are often critical part of the organization involved. In biochemical processes, there is a greater likelihood that two substances will interact if they are isolated in a component of the system than if they are permitted to disperse through the whole system. Also, it is possible to impose barriers in space to keep components that should not interact apart. But spatial relations are not the only organizing principle. Mutual dependencies between different operations can relate them. Cyclic processes in biochemistry, such as the Krebs (citric acid) cycle, serve to coordinate a set of reactions.

Organization serves in the first place to coordinate the operation of different components so that each of the components contributes to a larger overall activity than any of them can accomplish in themselves. Sometimes, as in catalyzed reactions in chemistry, by coordinating three components together it is possible to produce changes (binding two of the components together) which would not happen in simple pair-wise interactions. But it can also facilitate regulation so that individual operations only occur when they are required in a large process. In a cyclic process, if an earlier stage in a transformation depends upon a product produced at a later stage, then the earlier stage is regulated by the later stage.

Having identified the central components in the account of a mechanism, let me take the biochemical mechanism of fermentation as an exemplar. From one perspective, the mechanism transforms sugar (glucose) into either lactic acid or alcohol. However, from another perspective, the mechanism extracts energy from glucose when it is needed for cellular activities. Initially working only from the first perspective, a number of investigators tried in the first 30 years of the 20th century to formulate a coherent pathway from glucose to alcohol that involved only known chemical substances as intermediates and involved known types of chemical reactions—splitting of molecules, oxidations, reductions, etc. The best of the schemes that were developed involved as an intermediate a substance not known to occur in living cells and which could not stimulate the overall reaction if added to the reaction process. But as well it was recognized that inorganic phosphates needed to be added to the reaction when performed *in vitro*, although they did not seem to have a role in the process. Around 1930, though, two phosphorylated substances, creatine phosophate and adenosine triphosphate (ATP) were discovered and recognized to play an intermediate role in cell energetics by virtue of the high energy stored in the phosphate bonds. This information explained the need for phosphates and suggested the possibility that fermentation first involved the creation of a high-energy phosphate bond on an intermediate. The discovery of other substances which could transport hydrogen from the oxidation of one substrate to a reduction of another substrate further filled in the account of the mechanism, resulting in the articulation of the Embden-Meyerhof pathway in the early 1930s. The overall mechanism can be represented as in Figure X, which exhibits the dependencies between the different operations in the overall process. The various cycles involving NAD/NADH and ADP/ATP serve to relate a number of operations into a system which generate useable energy from glucose, yielding alcohol or lactic acid as a waste product.

This example from biochemistry provides a context to emphasize an important feature of mechanistic explanations. They are often incomplete. Sometimes researchers are unaware of the gaps in their explanation. They think they have provided all the components and operations that are necessary and offer an account of a mechanism that seems to be fully filled in. But then research reveals that another component is required. This often requires revisiting the previous account of what operation given components performed. They may have only performed part of the operation assigned to them, with other components performing the remainder. Sometimes researchers may be aware that there is a significant gap in their account. They have identified some of the components and are pretty confident in their assessment of what they contribute, but they recognize that there are gaps between these operations. They have more of what counts as a sketch of a mechanism that requires being filled in (Machamer et al., 2000). Very often mechanistic explanations advanced in the literature are works-in-progress, not final products. We will confront an example of such in the next lecture.

Thus far I have focused on explanation by identifying mechanisms without discussing the view to which it stands in opposition—the view that explanation consists of subsumption under laws. This view too has a long history, going back at least as far as Aristotle. It too was much in prominence in the 17th and 18th centuries. After Copernicus advanced his alternative conception of the solar system, Kepler offered laws to describe the motion of the planets. Newton generalized these laws into general laws regarding the attraction of any two bodies. From the perspective of the Cartesian mechanist, in allowing action at a distance, Newton was surrendering the mechanistic program and reintroducing occult forces. Unlike the occult forces of Medieval science, however, Newton's force laws had a powerful unifying power—they applied to any bodies the universe and, as developed by subsequent Newtonian scientists, covered a wide range of natural phenomena.

In the 20th century the strategy of explaining phenomena by subsuming them under laws was formalized in the deductive-nomological model of explanation. The framework renders explanation a linguistic activity—one explains a phenomenon by deriving a statement describing that phenomenon from a statement of one or more laws and statements identifying initial conditions (Hempel & Oppenheim, 1948). In this construal, commonly referred to as the deductive-nomological (D-N) model, logic becomes the glue of explanation and laws become the primary explainers.

Over the past twenty years this view of explanation has been challenged within the physical sciences by authors such as Nancy Cartwright (Cartwright, 1983, 1999) and Ron Giere (1999), who have maintained that laws do not truly apply to any real circumstances but only to idealized circumstances, which are more or less approximated in experimental setups. In the life sciences, a different problem is manifest. There simply are not many examples of laws. And yet there are a host of examples of explanation. For a naturalist, this is a primary factor encouraging setting aside the view that explanation requires laws and pursuing a different model of explanation that more closely captures what scientists offer as explanations. (In offering it as an alternative, I am not proposing that are no laws or that there are no explanations that are better understood on the D-N model, but only that in many cases it is mechanisms that do the explanatory work in science).

To further flesh out the mechanist alternative to the covering law account of explanation, I will conclude this discussion by consider briefly two features on which the two accounts divide. First, as I noted above, in the covering law framework, explanation is understood as a linguistic activity involving the derivation of a sentence describing the phenomenon from a statement of laws. The mechanistic framework does not require this linguistic turn. Mechanisms are things existing in the world and it is the ability of a mechanism to causally generate the phenomenon of interest that explains the phenomenon (Wimsatt, 1976). Scientists do, in fact, use language to describe mechanisms, but the language here serves simply to identify the components of the mechanism and how they are related to one another.[8] Moreover, this is often better done not in words but in diagrams since they facilitate representing multiple simultaneous activities, which are often critical for the operation of a mechanism.

---

[8] One might try to make a similar move within the covering law framework by claiming that it is the law that does the explanatory work, not the linguistic expression of it. But that raises difficult metaphysical issues for the covering law model. What, precisely, is a law beyond the statement of it? One might try to characterize it as a relation between universals (Armstrong, 1983), but then both how universals can stand in relations and the connection between the universals and the individual entities in the world requires explanation. In the covering law model, it is deduction that is the glue connecting statements of laws to statements of the phenomena, and it is not clear how to cash out this relation without the linguistic turn.

It is the case that in offering an explanation, one advances an account of the mechanism, whether in language or diagrams. Moreover, to understand the explanation, the theorizer must provide the dynamics to the account. For example, the theorizer must envisage how the components in the account perform their operations in coordination with each other. But these activities are in the service of representing the mechanism that is assumed to be operating in nature.

A second feature of mechanistic explanation is that, insofar as it emphasizes the important contributions parts of a mechanism make to the operation of a mechanism, a mechanistic analysis inherently looks to lower levels of organization. It is thus a reductionistic approach. But it is not only a reductionistic perspective. Insofar as it also recognizes the importance of the organization in which the parts are embedded and the context in which the whole mechanism is contained, it not only sanction looking to lower levels but also upwards to higher levels.

I will address the question of reduction in more detail in Lecture 5, but because the multi-level perspective is so important in understanding the mechanistic framework, let me comment just a bit more on the issue now. What often raises worries about reductionism is the assumption that a reductionist seeks to explain everything at the lower level. This assumes that there is a complete account of phenomena at the lower level. But at the core of the mechanistic perspective is the assumption that different types of causal relations are found at each level of organization and that no level is complete. A lower level analysis might explain how the parts of a mechanism behavior.[9] But critical to a mechanism is the organization imposed on these parts. As a result of this organization, the mechanism as a whole is able to engage in its behavior. What it does is different from what its components do, and is described in a different vocabulary. The account of the mechanism straddles the two levels of organization, showing how the mechanism performs its activity as a result of the processes performed in the organized context by the components. And the mechanism carries out this activity as a result of being itself situated in a particular context, perhaps as a component of a yet higher-level mechanism. The account of the activity therefore must take into account these higher-level interactions. Thus, far from sanctioning only a focus downwards on the components, a mechanistic perspective requires as well an account of engagements with other systems at the same level and potentially of the constrains imposed by being incorporated into a higher-level mechanism. In fact, therefore, the mechanistic perspective is inherently multi-level.

3. Looking Ahead

In this lecture I have developed the stance characterized by naturalism and mechanism that I will

---

[9] Whether the lower level provides even a complete account of the behavior of the parts of a mechanism is itself in question. The behavior an entity exhibits is often dependent upon context and there is no reason to think that the account of an entity offered by any inquiry considers how it will behave under all conditions but only those which are the focus of inquiry. As engineers are well aware, how a component will behave when inserted into a particular kind of system often needs to be investigated empirically.

adopt in considering how philosophy engages the cognitive neurosciences. In the next lecture I will develop an exemplary case in which neuroscience has already developed a mechanistic model of a mental activity—visual perception. This will serve both to fill out further the conception of a mechanistic explanation that characterizes neuroscience research and provide a perspective on how mechanistic models are developed over time. Then in lecture 3 I will consider one of the phenomena that has driven much of the philosophical discussion about neuroscience—can neuroscience account for phenomenal experience. Addressing this topic will require providing a new perspective on what is involved in identifying mental processes with brain process, a perspective I characterize as heuristic identity theory.

In lecture 4 I take up another challenge for cognitive neuroscience—whether it can address not just "low-level" mental processes such as perception, but higher-level processes such as reasoning. What is characteristic of cognitive accounts of these processes is that they involve operations over representations. But even visual neuroscience involves appeal to representations. The question will be whether the type of representations posited in neuroscience can be the basis for building up to the type of representations seemingly needed to account for higher cognitive processes. Then in lecture 5 I will return to a topic just adumbrated above. Insofar as the mechanistic framework seems to be reductive insofar as it appeals to lower-level components of mechanisms to explain what they do, it may seem to remove interest in higher-level phenomena such as cognitive or culture. But I will argue that this is a misconstrual of reduction and that in fact the mechanistic perspective is inherently a multi-level perspective which insists on the importance of higher levels of organization and describing phenomena at these higher levels. Reduction is just one of the bugaboos confronting the mechanistic perspective, and in the final lecture I will turn briefly to others involving human freedom, responsibility, and dignity.

References

Armstrong, D. (1983). *What is a law of nature?* Cambridge, MA: Cambridge University Press.

Bechtel, W. (2000). From imaging to believing: Epistemic issues in generating biological data. In R. Creath & J. Maienschein (Eds.), *Biology and epistemology* (pp. 138-163). Cambridge, England: Cambridge University Press.

Bechtel, W. (2002). Aligning multiple research techniques in cognitive neuroscience: Why is it important? *Philosophy of Science, 69*, S48-S58.

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

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

Boas, M. (1952). The establishment of the mechanical philosophy. *Osiris, 10*, 412-541.

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

Callebaut, W. (1993). *Taking the naturalistic turn, or, How real philosophy of science is done*. Chicago: University of Chicago Press.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Cartwright, N. (1999). *The dapled world: A study of the boundaries of science*. Cambridge: Cambridge University Press.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*, 741-765.

Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA: MIT Press/Bradford Books.

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

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

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

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

Giere, R. G. (2002). Scientific cognition as distributed cognition. In P. Carruthers & S. Stich & M. Siegal (Eds.), *The cognitive bases of science*. Cambridge: Cambridge University Press.

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

Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.

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

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

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

Nersessian, N. (2002). The cognitive basis of model-based reasoning in science. In P. Carruthers

& S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 133-153). Cambridge: Cambridge University Press.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind and Religion*. Pittsburgh: University of Pittsburgh Press.

Quine, W. V. O. (1969). Epistemology naturalized. In W. V. O. Quine (Ed.), *Ontological relativity and other essays*. New York: Columbia University Press.

Westfall, R. S. (1971). *The construction of modern science*. New York: Wiley.

Wimsatt, W. C. (1976). Reductive explanation: A functional account. In R. S. Cohen & C. A. Hooker & A. C. Michalos & J. v. Evra (Eds.), *PSA-1974. Boston Studies in the Philosohpy of Science* (Vol. 32, pp. 671-710). Dordrecht: Reidel.