Explanation: The Mechanist Alternative

William Bechtel and Adele Abrahamsen
University of California, San Diego

Abstract

Explanations in the life sciences frequently involve presenting a model of the mechanism taken to be responsible for a given phenomenon. Such explanations depart in numerous ways from nomological explanations commonly presented in philosophy of science. This paper focuses on three sorts of differences. First, investigators are not limited to linguistic representations and logical inference in presenting explanations but frequently employ diagrams and reason about mechanisms by simulating them. Thus, the epistemic resources for presenting mechanistic explanations are considerably richer than those available for nomological explanations. Secondly, the fact that mechanisms involve organized component parts performing coordinated component operations provides direction to the processes of both discovery and testing of mechanistic explanations. The strategies scientists employ in investigating mechanisms offers a rich area for philosophical inquiry. Finally, models of mechanisms are developed for specific exemplars and are not represented in terms of universally quantified statements. Generalization involves investigating both the similarities of new instances to the already studied exemplars and the variations that arise in those cases and thus involves a long-term endeavor.

Key terms: mechanistic explanation, diagrams, simulation, discovery, generalization

Recent years have been marked by increasing discontent within philosophy of science regarding the longstanding approach to scientific explanation in which in laws play an essential, central role (e.g., (Cartwright, 1983, 1999; Giere, 1999). Laws may even be impossible in principle, at least in the biological sciences (Beatty, 1995; Rosenberg, 1994). My goal in this paper is not to rehearse or advance the philosophical objections to laws, either generally or specifically in the context of biology. I rather start with the observation that most *actual* explanations in the life sciences do not appeal to laws in the manner specified in the traditional deductive nomological (D-N) model. For example, biologists who want to explain phenomena involving ATP—the molecule that stores the energy harvested by metabolic processes in the cell—would not be satisfied by such putative laws as one stating that under specified conditions, the ratio of oxygen molecules consumed to ATP molecules produced does not exceed 1:3. Even if accorded the status of a law, this statement would merely provide a characterization of one metabolic phenomenon. It would do nothing to explain the phenomenon.[1]

How are biological phenomena explained, if not by means of laws? Perusing the biological literature, it quickly becomes clear that the term biologists most frequently invoke in explanatory contexts is *mechanism*.[2] In cell biology, for example, one finds frequent references to, and accounts of, mechanisms of fermentation, protein synthesis, secretion, action potential generation, and so forth. Although mostly neglected in the literature of 20th century philosophy of science,[3] mechanistic explanation has recently attracted the attention of a number of philosophers of biology. One way they have begun to address the absence of an appropriate framework is to offer a variety of characterizations of mechanisms, their discovery, and their role in explanation (see, for example, Bechtel & Richardson, 1993; Glennan, 1996; Machamer, Darden, & Craver, 2000). These characterizations show substantial overlap. Our own version is that a mechanism is a system whose behavior produces a phenomenon in virtue of organized component parts performing coordinated component operations.[4]

[1] Cummins (2000) notes that what elsewhere are called *laws* are labeled *effects* in psychology and notes examples such as the Garcia effect (avoidance of a food whose intake on a previous occasion was followed by nausea or other gastrointestinal distress). Such effects, he maintains, are not invoked to explain things, but rather are themselves in need of explanation. Moreover, they are less likely to be direct targets of explanatory activity than to serve as constraints on explanations of the more basic psychological phenomena that underlie them (e.g., classical conditioning).

[2] In this paper we are targeting explanations in the life sciences. We are not claiming all scientific explanations involve appeals to mechanisms or even that in biology all explanation take the form of identifying the responsible mechanism. The claim is far more modest—that in many cases in the life sciences, and potentially in other sciences, the quest for explanation is the quest for a specification of the appropriate mechanism.

[3] Mechanism figured prominently in the early modern philosophy of Descartes and Boyle. Even Newton, whose appeal to forces seems to reject the quest for mechanism, in some other contexts did advance mechanistic explanations. Newton's appeal to forces, however, provided the prototype for the deductive-nomological model of explanation promoted by the Logical Positivists. In the late 20th century Wesley Salmon (1984) set out to revive the mechanical philosophy. He focused more on causal explanations, however, than on mechanisms as characterized here.

[4] There are some salient differences between the various accounts of mechanism. Bechtel and Richardson (1993) focus on the "functions" (operations) that parts perform, whereas Glennan focuses on the properties of parts in stating what he originally (1996) called laws and now (2002)calls "invariant change-relating generalizations." These are instantiated in "interactions" in which "a change in a property of one part brings about a change in a property of another part" (2002, p. S344). Machamer, Darden, and Craver (2000) pursue the metaphysical status of "entities" (parts) and "activities" (operations). Tabery (2004) has proposed a partial synthesis in which activity and property changes are seen as complementary. We use the term *operation* rather than *activity* because we want to draw attention to the involvement of parts; for example, enzymes operate on substrates so as to catalyze changes in the substrates. For a more complete account of the multiple roles played by parts, see (Bechtel, in press-a) } Finally, Machamer et al. (p. 3) include a characterization of mechanisms as "productive of regular changes from start or set-up to finish or termination conditions." I am concerned that such an emphasis 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 on the conditions in which an operation is elicited and on its contribution to the behavior of the overall mechanism. However we often have to counter this analytical perspective to appreciate the dynamics at work in the system.

For example, Figure 1 provides a sketchy illustration of a familiar mechanism: the heart (a *system*) pumping blood (the system's *behavior*). The major *parts* of the heart are familiar enough that the abbreviated labels should suffice. The *operations* performed by the parts include contraction and relaxation (by all four chambers) and blockage of reverse movement of blood (by all four valves). The heart is itself part of a larger mechanism, the circulatory system, that includes such parts as veins, arteries, and the blood itself (which is operated upon by the heart chambers, valves, veins, arteries, etc.). These various component parts must be *organized* (positioned and connected) such that blood can flow on each side from atrium to valve to ventricle to valve to aorta or pulmonary artery into the rest of the circulatory system, as suggested by the arrows. At least as important, the operations must be timed to achieve a *coordinated* effect. (How this is achieved is a complex story involving additional parts, such as the heart's pacemaker.)

For the purposes of this paper, we will employ this conception of mechanism[5] and focus on how explanations that appeal to mechanisms differ from those that appeal to laws. Before turning to this endeavor, however, a few characteristics of mechanistic explanation should be clarified.

First, mechanisms are in the world, whereas providing explanations, including mechanistic explanations, is a cognitive activity. There has been a tendency, originating with Salmon (1984), to treat the mechanism operative in the world as itself providing explanation. Thus, Salmon identifies his approach to explanation as *ontic* insofar as it appeals to the actual mechanism in nature, and contrasts it with an *epistemic* conception of explanation that appeals to derivations from laws, which are clearly products of mental activities. Salmon's insight is important, but the ontic/epistemic distinction fails to capture it. The important insight is that mechanisms are real systems in nature, and hence one does not have to face questions comparable to those faced by nomological accounts of explanation about the ontological status of laws. But it is crucial to note that offering an explanation is still an epistemic activity and that it is not the mechanism in nature that directly performs the explanatory work.[6] This is particularly obvious when one considers incorrect mechanistic explanations—in such a case one has still appealed to a mechanism, but not one operative in nature. Another way to appreciate the point is to note that in many instances the mechanism in question was operative long before scientists discovered the mechanism and thereby explained the phenomenon.

Thus, since explanation is itself an epistemic activity, what figures in it are not the mechanisms in the world, but representations of them. These representations may be internal mental representations, but they may also take the form of representations

[5] This conception ultimately is not sufficient for characterizing biological mechanisms. An important recent way of viewing biological mechanisms is as autonomous systems in a condition far from equilibrium. On this view, they face a variety of additional demands, including being self-organizing systems that reside between a source and sync for energy and have the capacity to capture and utilize energy to develop and maintain themselves. While these additional considerations do not invalidate the features of mechanism that I will focus on in this paper, they impose additional constraints on the sorts of mechanisms that occur in biological systems, as opposed to humanly engineered systems which employ both external designers and repair persons. (See Bechtel, forthcoming; Ruiz-Mirazo, Peretó, & Moreno, in press)

[6] I thank Cory Wright for impressing on me that a mechanism in nature does not itself explain anything.

external to the cognitive agent—diagrams, linguistic descriptions, mathematical equations, physical models, etc. Generically, one can refer to these internal and external representations as *models* of the mechanism. A model of a mechanism describes or portrays what are taken to be its relevant component parts and operations,[7] the organization of the parts into a system, and the means by which operations are coordinated so as to simulate the system's behavior. When they are correct, models of mechanisms accurately describe the mechanism operative in the world that is responsible for the phenomenon.

Second, identifying the component parts and operations of a mechanism and their organization is only part of the overall endeavor of mechanistic explanation of a phenomenon. Looking outwards, the mechanism generating a phenomenon typically does so only in appropriate external circumstances. A relatively simple example from cell biology is that yeast cells carry out fermentation only when glucose and ADP are available and oxygen is not. For numerous other phenomena—such as those of gene expression in cell biology and speciation in evolutionary biology—the relevant external circumstances are more complex. Nonetheless, it is crucial to identify them and to explore how variations affect the behavior of the mechanism. Often, such explorations reveal that the external circumstances are best understood as involving a larger mechanism in which the target mechanism is embedded, rather than as features of an environment. Moreover, many of the components of a mechanism are themselves mechanisms which could be targeted in another round of explanation. Thus, whether we look outwards or inwards from the targeted mechanism, the same important point becomes evident: mechanistic explanation can be recursive.

Traditional reduction (Causey, 1977; Nagel, 1961) is also recursive, but there is an important difference. In traditional reduction, the most primitive level must offer a comprehensive account of all phenomena. In mechanistic explanation, successively lower level mechanisms account for *different* phenomena. What is achieved is a cascade of explanations, each appropriate to its level and not replaced by those below (see Bechtel, 1994, 1995, 2001). From any one level, going down a level offers a kind of reduction (to component parts and operations). And going up a level provides a different perspective: that the mechanism's behavior may be modulated by that of a larger mechanism in which it is embedded (see Craver & Bechtel, submitted).

Mechanistic explanations do not merely trade laws for models of mechanisms, leaving the rest of the standard account of explanation unchanged. Rather, thinking in terms of mechanisms as the vehicles of explanation transforms how one thinks about a host of other issues in philosophy of science. My goal in this paper is to draw out some of the important consequences of mechanistic explanation for some of these traditional issues. Although the D-N model has long since lost its status as the generally accepted model of explanation in philosophy of science, I will frequently invoke it by way of contrast in

[7] When we are emphasizing the thing performing the operation, we use the term *part* or *component part* and when we are emphasizing what the part does, we speak of *operation* or *component operation*. We use the term *component* alone where it is not important to be specific or where we wish to jointly designate the part and the operation it performs.

what follows because it still provides the backdrop for much philosophical thinking about explanation.

1. Representing and reasoning about mechanisms

Two of the major legacies of the D-N model are the assumptions that (1) the primary mode of representing explanations is propositional and (2) logic provides the tools for reasoning about these representations. In particular, explanation is viewed as a process of logical inference from laws and theories to statements describing the phenomenon to be explained (Hempel, 1965; Nagel, 1961). Although scientific papers do not typically include formalized arguments, such arguments are assumed to be implicit and philosophers often take it to be part of their job to reconstruct scientific explanations as proper derivations from statements of laws and initial conditions. But are propositional representations and logical inference the most appropriate devices for representing and reasoning about mechanisms? A prominent feature of almost any paper in biology (in print, or especially when presented to an audience) is the reliance on visuospatial representations. The figures in a paper might include a photograph showing an apparatus, a micrograph revealing a subcellular structure, a graph summarizing the results of an experiment, and so forth. Of particular interest here, figures—especially diagrams—can play a key role in presenting an explanation. From a perspective in which linguistic representation is treated as primary, the use of figures and diagrams seems to be comparable to their invocation in formal geometry, where they are often construed as crutches to help one understand the relations specified in the laws. As such, they may be discounted as not themselves central to the explanatory endeavor.

When one considers the actual practice of scientists in reading papers, however, the tables seem to be turned. It is not uncommon for a reader to scan the abstract and then jump to key figures. To the extent that crutches are involved, it is the labels and figure captions providing commentary on the figures that play this role. Consider a paper in which a mechanistic explanation is proposed. The diagram provides a vehicle for keeping in mind the complex interactions among operations, while the commentary can only characterize these one at a time. The text of the paper then provides yet further commentary: about how the mechanism is expected to operate (introduction), how evidence as to its operation was procured (methods), what evidence was advanced (results), and the interpretation of how these results bear on the proposed mechanism (discussion). The detailed commentary is important, but it is the diagram that fixes the mechanism in the reader's mind. As just one example, de Duve (de Duve, 1969, p. 5) recollects that his discovery of the lysosome was sparked by an unexpected failure in his biochemical investigation of a liver enzyme. "By some fortunate coincidence, my recent readings had included" two 1946 papers by Claude and "I immediately recalled Claude's diagrams showing the agglutination at *p*H 5 of both large and small granules, and concluded that our enzyme was likely to be firmly attached to some kind of subcellular structure."

The importance scientists place on diagrams should lead us to question whether they are in fact superfluous. Are there reasons a scientist might prefer to represent information diagrammatically rather than propositionally? More importantly, are there different processes of reasoning with diagrams than with propositions such that an account of science that focused only on logical inference would fail to capture an important aspect of explanatory reasoning?

The motivation for using diagrams to represent mechanisms is obvious. Unlike linguistic representations (except those found in signed languages), diagrams make use of space to convey information.[8] As we have already seen in the diagram of a heart, spatial layout and organization is often critical to the operation of a mechanism. As in a factory, different operations occur at different locations. Sometimes this serves to keep operations separate from one another and sometimes it serves to place operations in association with one another. These spatial relationships can be readily shown in a diagram. Even when information about the specific spatial layout is lacking or not significant, moreover, one can use space in the diagram to relate or separate operations conceptually.

Time is at least as important as space to the operation of a mechanism—one operation proceeds, follows, or is simultaneous with another operation. This can be captured by using one of the spatial dimensions in a diagram to convey temporal order. This of course presents a problem: with most diagrams laid out in two dimensions, that leaves just one dimension for everything other than time. One solution—as exemplified in the heart diagram—is to make strategic use of arrows to represent temporal relations, leaving both dimensions free to represent the mechanism's spatial or similarity relations. Another solution is to use techniques for projecting three dimensions onto a two-dimensional plane.

Whether the temporal order of operations is represented by means of a spatial dimension or by arrows, a diagram has clear advantages over linguistic description. The most obvious advantage—that all parts and operations are available for inspection simultaneously—probably is the weakest one. Due to processing limitations, people can only take in pieces of the diagram at a time. Nonetheless, more so than when reading text, they have the freedom to move around it in any number of ways. A stronger advantage is that diagrams offer relatively direct, iconic resources for representation that can be invaluable. For example, it is immediately apparent in the heart diagram that blood is being pumped simultaneously from the two atrial chambers to the two ventricles and that these two parallel operations are in a sequential relationship to two other parallel operations (pumping from the two ventricular chambers). The value of consulting a diagram in this way is even more apparent in mechanisms with feedback loops, through which an operation that is conceptually downstream (closer to producing what is taken to be the product of the mechanism) has effects that alter the execution of operations earlier in the steam at subsequent time steps. Multiple examples can be found within the oxidative metabolism mechanism that functions within our cells to harvest and store energy from food. It is composed of three connected submechanisms, which when further unpacked are seen to involve coordinated biochemical operations, including feedback

[8] In addition to spatial features, diagrams also take advantage of other features that visual processing can access, including color and shape.

operations. Figure 2 shows one of these. It involves the last few components of the first system, glycolysis, which converts carbohydrates into Acetyl-CoA and is self-regulating. When the need for energy is high, most of the Acetyl-CoA feeds into the second major metabolic system, the Krebs cycle (solid line). When the need is modest, however, Acetyl-CoA accumulates and feeds back (dotted line) to inhibit the operation of the enzyme pyruvate kinase, which catalyzes the reaction converting phosphoenolpyruvate to pyruvate. The diagram aids understanding by offering a spatial layout of the parts of the system (compounds such as pyruvate) and by using the vertical dimension as well as solid vs. dotted arrows to indicate the sequence of operations (chemical reactions).

Although a mechanism can be represented with a diagram, it can also be described linguistically. Is there any fundamental difference between linguistic and diagrammatic representations? Larkin and Simon (1987) considered diagrams and linguistic representations that are informationally equivalent and analyzed how they can nonetheless differ with respect to ease of search, pattern recognition, and the inference procedures that can be applied to them. In part these differences stem from the fact that information that may be only implicit in a linguistic representation may be made explicit, and hence easier to invoke in reasoning, in a diagram.[9]

An important principle of computational modeling of reasoning is that it is essential to coordinate the modes of representation and procedures of inference. If diagrams are an important vehicle for representing mechanisms, then it is necessary to consider how one reasons about diagrams. In particular, if diagrams represent information that is not represented (or easily represented) in linguistic representations, then deductive inference, the traditional glue that relates laws and statements of the phenomena to be explained in D-N models, will not capture the reasoning involved in understanding how a given mechanism produces the phenomenon.[10] What, then, provides the glue of mechanistic explanations invoking diagrams? Here it is important to keep in focus the fact that mechanisms generate the phenomenon in virtue of their components performing their own operations in a coordinated manner. The kind of reasoning that is needed is reasoning that captures the actual operation of the mechanism, including both the operations the components are performing and the way these operations relate to one another.

One limitation of diagrams when it comes to understanding mechanisms is that they are static.

[9] Larkin and Simon comment: "In the representations we call diagrammatic, information is organized by location, and often much of the information needed to make an inference is present and explicit at a single location. In addition, cues to the next logical step in the problem may be present at an adjacent location. Therefore problem solving can proceed through a smooth traversal of the diagram, and may require very little search or computation of elements that had been implicit (Larkin & Simon, 1987, p. 65)

[10] It is ironical to observe that one of the leading fields in which investigators explore the power of diagrams is logical inference itself. Interest in diagrammatic representation of logical inference dates back to Euler and Venn. For a discussion by logicians of diagrams in logical reasoning itself, see Barwise and Etchemendy (1995). For an attempt to characterize logical reasoning in terms of mental diagrams, see Johnson-Laird (1983).

Even if they incorporate arrows to characterize the dynamics of the mechanism, the diagram itself doesn't do anything. Thus, it cannot capture the relation of the operation of the parts to the behavior of the whole mechanism. Accordingly, the glue holding these together must be provided by the cognitive agent. The cognizer must imagine the different operations being performed, thereby turning a static representation into something dynamic.[11] Mary Hegarty (1992) calls the activity of inferring "the state of one component of the system given information about the states of the other system components, and the relations between the components" *mental animation* and emphasizes its importance to the activities of designing, troubleshooting, and operating mechanical devices. Obtaining reaction time and eye movement data while people solved problems about relatively simple pulley systems, she investigated the extent to which inference processes are isomorphic to the operation of the physical systems. One way they were not isomorphic is that the participants made inferences about different components of the system (i.e., individual pulleys) separately and sequentially even though in the physical system the components operated simultaneously. The participants found it considerably harder, however, to make inferences that required them to reason backwards through the system rather than forwards, suggesting that they animated the system sequentially from what they represented as the first operation, in this respect preserving isomorphism with the actual system.

Accepting the claim that people, including scientists, understand diagrams of mechanisms by animating them, a natural follow-up question concerns how they do this. A plausible initial proposal is that they create and transform an image of the mechanism so as to represent the different components each carrying out their operations. In perception we have experience of parts of the system changing over time, and so the proposal is that in imagination we animate these components by invoking the same processes that would arise if we were to watch an animated diagram. This proposal needs to be construed carefully, as a potential misunderstanding looms. Reference to a mental image should not be construed as reference to a mental object such as a picture in the head. Recent cognitive neuroscience research indicates that when people form images they utilize many of the same neural resources that they do in perception (Kosslyn, 1994).[12] Thus, what occurs in the head in forming an image is activity comparable to that which would occur when seeing an actual image. Barsalou (1999) speaks of this neural activity as a *perceptual symbol*. Thinking with perceptual symbols then involves the brain initiating sequences of operations that correspond to what it would undergo if confronted with

[11] Animated diagrams relieve people of this difficult task and are often far more instructive to novices. Thomas M. Terry of University of Connecticut has produced some excellent ones that make clear how the many operations in cellular metabolism are related. He has them posted at http://www.sp.ucon.edu/~terry/images/anim/ETS.html. Another good site for such diagrams, which also provides links to Terry's diagrams, is http://www.people.virginia.edu/~rjh9u/atpyield.html.

[12] Within cognitive science there has been a heated controversy over whether the representations formed in the cognitive system are really image-like (Kosslyn, 1981, 1994; Pylyshyn, 1981, 2003). This discussion can remain neutral on this issue since the fundamental issue is not how the cognitive system encodes its representations but what it represents something as. The visual system represents objects as extended in space and changing through time. What is important here is that a scientist can represent mechanisms in much the way that she represents diagrams that she encounters (albeit with less detail than when she actually is looking at the diagram).

actual input from visual objects behaving in a particular manner. On this account, the mental image is not a real object but the intentional object of the mental activity.

Although humans are relatively good at forming and manipulating images of rather simple systems, if the mechanism is complex and involves multiple components interacting with and changing each other, we often go astray. We fail to keep track of all the changes that would occur in other components of the system in response to the changes we do imagine. Thus, the usefulness of mental animation for understanding a system does reach a limit. Ordinary people may simply stop trying at this point, but scientists and engineers often find it important to do better and hence have created tools that supplement human abilities to imagine a system in action. One tool involves building a scale model (or otherwise simplified version) of a system and operating on the scale model to determine how the actual system would behave. The behavior of the scale model *simulates* that of the actual system. For example, the behavior of objects in wind tunnels can be used to simulate phenomena involving turbulence in natural environments. If instead an investigator can devise equations that accurately characterize the changes in a system over time, the investigator can often determine how the system will behave by solving the equations without actually building a simulator. In this case the simulation is done with a mathematical model rather then a physical model. The advent of the computer provided both a means of solving the equations of a mathematical model and an additional means of simulating systems. Higher level computer languages are designed to represent complex structures and their interactions, and by using these resources, one can often create a computer simulation of the interactions in a complex system (Jonker, Treur, & Wijngaards, 2002).

These different modes of simulating a system all provide an important advantage when a system is complex with multiple operations occurring simultaneously—they do not lose track of some of the interactions as a human imagining the operation of the system often does. But even when it is a human that is doing the imagining, what he or she is doing can also be characterized as simulating the system.

Representation and inference in mechanistic explanation is thus quite different from representation and inference in nomological explanation. While it is possible to give a linguistic description of a mechanism, the linguistic account is not privileged. Frequently diagrams provide a preferred representation of a mechanism. Inference involves a determination of how a mechanism behaves, and this is typically not achieved via logical inference but by simulating the activity of a mechanism, either by animating a diagram or by creating mental, computational, or scale model simulations.

2. Discovering and testing models of mechanisms

Advocates of the D-N account of explanation have been able to say very little about the discovery of explanations.[13] On the nomological account, discovery involves

[13] Accordingly, Reichenbach (1966) made a principled distinction between the context of discovery and the context of justification. Justification was the proper focus of philosophy, while the context of discovery might be explored by psychologists.

constructing a law that fits a range of cases. When the law involves only observation terms, the challenge is to identify the appropriate terms and determine how they are related. A variety of mathematical techniques are available for doing this, and some AI researchers have marshaled such techniques into computer programs designed to discover new laws (Holland, Holyoak, Nisbett, & Thagard, 1987; Langley, Simon, Bradshaw, & Zytkow, 1987). But when laws require positing theoretical terms that do not refer directly to observables, it is far more difficult to provide insight into how discovery proceeds. A diagnosis of this problem is that within nomological accounts, theoretical terms are not grounded in observations. Rather, scientists are taken to posit them for use in theories from which laws using only observation terms can be derived. This leaves the process by which a scientist develops such hypotheses rather obscure.

Mechanistic explanations, in contrast, seek to identify component parts and operations of a mechanism. Even an investigator who does not observe the components, but instead infers them, is construing them as the parts and operations of an actually existing mechanism. Accordingly, there is a great deal that can be said about the discovery process. The very conception of a mechanism lays out the tasks involved—one must identify the working parts of the mechanism, determine what operations they perform, and figure out how they are organized so as to generate the phenomenon. This requires taking the mechanism apart, either physically or conceptually, a process that Bechtel and Richardson (1993) called *decomposition*. There are two ways researchers decompose mechanisms—structurally or functionally—depending on whether they focus on component parts or component operations. Interestingly, it is generally different researchers in different fields who achieve these two kinds of decomposition for a given mechanism. Combining these into a complete account generally is a later achievement.

To begin with functional decomposition, here the strategy is to start with the overall behavior of the system (its task or function) and figure out what lower-level operations contribute to achieving it. These operations are characterized differently in different sorts of mechanisms, but often involve transformations to some substrate. The biochemical system that performs glycolysis in cells, for example, catabolizes glucose to carbon dioxide and water. The component operations are then characterized in terms of individual chemical reactions on a series of substrates (e.g., oxidizing or reducing them, adding or removing H_2O, etc.). In information processing systems, informational structures (representations) are the substrate, and the operations are information processing activities (e.g., moving or altering representations). Often it is possible to determine, at least to a first approximation, what the internal operations of a mechanism are without knowing what parts perform these operations. In the case of fermentation, biochemists were able to determine the chemical reactions involved in the overall activity and designate the responsible agents (enzymes) for each without knowing the chemical structure of the agents (Bechtel, 1986).[14] Likewise, cognitive psychologists are often

[14] This does not mean that first guesses about the operations in the mechanism are necessarily correct. In the case of fermentation, researchers first tried to link the beginning and end products (glucose and alcohol or lactic acid) through various 3-carbon sugars that were plausible intermediates, not aware that the intermediate compounds were phosphorylated. But in this case the corrections to the initial proposal were developed through further experimentation on the overall process and determining intermediate substrates

able to specify the internal information processing operations involved in performing a given task without access to the brain components that execute these operations. Instead of direct access to the structures operative within the system, such decompositions rely on cleverly designed perturbations of the functioning of the system that provide clues to the operations being performed within the system. In such cases, then, functional decomposition precedes structural decomposition.

Turning now to structural decomposition, it is important to emphasize at the outset that the structural components into which researchers seek to decompose a system are ones which perform the operations that figure in the functional decomposition. The majority of ways of structurally decomposing a system will not result in components that perform operations. As Craver (forthcoming) notes, one might dice any system into cubes, but these cubes do not individually perform operations in terms of which one can explain the phenomenon. To reflect this fact, I will, following Craver, refer to the components as *working parts*. Although the goal is to find working parts, it is possible to decompose a system structurally independently of actually being able to determine the operations the various components perform. This involves, for example, appraising that component structures are likely to be distinct working parts on other grounds.

Cytological research on cell organelles provides an example of structural decomposition. Figure 3 shows the outcome of several rounds of decomposition of the cell (which required decades of research to achieve). I focus here on one important organelle, the mitochondrion, which can be observed in the cell's cytoplasm with an ordinary light microscope. The finer features of its structure were discovered through electron microscopy several years before their functional significance was recognized. Palade (1952) discovered that mitochondria not only have an outer membrane but also an inner membrane, which folds into the substance (matrix) of the mitochondrion. The infoldings are called *cristae*. Later Fernández-Morán (Fernández-Morán, Oda, Blair, & Green, 1964) discovered small knobs on these cristae which were determined to be comprised of an enzyme, ATPase.

Although researchers might frequently differentiate component operations before linking them with a part, or identify component parts without yet knowing what operations they perform, the ultimate goal is to link operations with parts. I refer to proposals of this kind as *localizations*. For example, as early as the 19th century it was recognized that the cell is the physiological unit that carries out the overall operation of metabolism. The next level of mechanistic explanation of metabolism (an achievement of mid-20th century cell biology) differentiates three major metabolic operations and localizes each in a different part of the cell, as illustrated in Figure 4. Anaerobic oxidation is an operation carried out in the cytoplasm (in animal cells this is glycolysis, a portion of which was illustrated in Figure 2). Aerobic oxidation (via the Krebs cycle) is an operation of the inner part (matrix) of the mitochondrion. The coupled operations of electron transport and oxidative phosphorylation are carried out in the cristae (infoldings of the inner membrane of the mitochondrion).

without any direct access to the enzymes that were hypothesized to catalyze the various intermediate reactions.

The ability to link parts with operations provides a means of corroborating each decomposition. Thus, linking a component operation with an independently identified component part provides evidence that both really figure in the mechanism. Failure to link operations with parts, on the other hand, can be grounds for doubting the existence of either the part or the operation. For example, from the time of its discovery in the 19th century until substantiated information about its operation was obtained in the 1960s, the Golgi apparatus was frequently suspected to be an artifact. In 1949 two future Noble Laureates, Albert Claude and George Palade, were among the last to argue that the Golgi apparatus was an artifact and did not really exist in living cells (Palade & Claude, 1949a, 1949b). Their reasoning was exemplary and included demonstrations of the ability to create structures with the appearance of the Golgi apparatus from materials much like those used to stain cell preparations. However, Palade later demonstrated that newly synthesized proteins migrate from the ribosome to the Golgi area, where they are concentrated into secretory vesicles. Serious investigators (including Palade) no longer questioned the existence of Golgi apparatus: association with a clearly delineated operation had vindicated the existence of the part itself.

The tasks of actually decomposing mechanisms into component parts and operations and of localizing operations in parts require invoking a variety of experimental procedures such as inhibiting a component to observe its affect on the overall operation of the mechanism or imaging changes internal to the mechanism when it is operative under various conditions. The investigation into how these and other research strategies can provide clues to the mechanisms responsible for particular phenomena is a relatively new undertaking in philosophy (Bechtel, in press-b; Bechtel & Richardson, 1993; Craver, 2002; Darden, 1991; Darden & Craver, 2002). But it is already clear that when the goal of discovery is the articulation of mechanisms, there is much more to be said about the discovery process than when the goal is simply the articulation of laws. In the sketch above I have focused on the discovery of the component parts and component operations in mechanisms, but a major part of the discovery process involves their organization (Bechtel & Richardson, 1993). Although we often think of component operations as linear (in the sense of occurring in sequence), living systems often make use of various forms of non-linear organization. Non-linear interactions can yield rather surprising forms of behavior, including self organizing behavior, which got little attention until the last decades of the 20th century (Barabási & Albert, 1999; Kaneko & Tsuda, 2001). These investigations often rely on computational modeling, and the techniques for developing and applying such models in understanding the organization of mechanisms are yet to be targeted in philosophical inquiry.

So far in this section I have focused on the discovery of mechanisms. However, science involves not just the advancement of hypotheses but also testing whether they are true in a given situation. While advocates of nomological explanations had little to say about discovery, they did attempt to articulate criteria for evaluation of proposed laws. Essentially, this involved making predictions based on the laws and evaluating the truth of these predictions. The challenge was to articulate a logic that would relate the truth or falsity of a prediction to the confirmation or falsification of the law. This is not the occasion to review the problems and solutions that have been proposed but simply to note the genesis of the problems for confirmation or falsification. Confirmation is challenging because there are always alternative possible laws from which one might make the same prediction (underdetermination). Falsification is challenging since a false prediction might be due to an error either in the proposed law or in one of the auxiliary hypotheses that figured in deriving the prediction (credit assignment).

My goal here is to focus on how such testing occurs when mechanisms rather than laws are the vehicle of explanation. In Section 1, I argued that inferences about mechanisms often involve simulation rather than formal logical deduction, but modulo that difference, the challenge for testing hypotheses is much the same for mechanisms as for laws. A researcher tests hypothesized mechanisms by inferring how the mechanism or its components will behave under specified conditions and uses the results of actually subjecting the mechanism to these conditions to evaluate the proposed mechanism. In principle the same challenges confront tests of mechanisms as tests of laws—different mechanisms might produce the same predictions (underdetermination) and when a prediction fails, the problem might lie either with the model of the mechanism or with auxiliary hypotheses invoked in making the prediction (credit assignment).

Although the problems of underdetermination and credit assignment are not eliminated just by focusing on mechanisms, there is much more to be said about testing mechanistic hypotheses than about testing laws. When a researcher sets out to test a model of a mechanism, the focus is typically not on the mechanism as a whole, but on specific components of the mechanism. Evidence is sought that a given component actually figures in the generation of the phenomenon in the way proposed. Thus, predictions are diagnostic—specifically targeted to the effects such a component would have (e.g., that if one intervened in a way known to incapacitate that component, there would be a specific change in the way the mechanism as a whole would behave). Consequently, the results of tests of mechanisms are frequently much more informative than those of laws.

An important aspect of discovering and testing mechanisms is that inquiry does not simply consist of postulating and testing a mechanism. Rather, research typically begins with an oversimplified account in which only a few components and aspects of their organization are specified. Over time, it is repeatedly revised and filled in (Bechtel & Richardson, 1993). Machamer, Darden and Craver (2000) refer to the simplified account as a *sketch* of a mechanism. Much of the discovery and testing involved in mechanistic explanation focuses on proposing components or forms of organization that are to be added to or used to revise parts of a sketch, and (often late in the process) localizing the worked-out component operations in the appropriate component part or parts. The entire process is typically a long-term endeavor (Bechtel, 2002). The result looks much more like a research program (Lakatos, 1970) than like the classical account of theory-testing.

3. Generalizing without laws

An important desideratum for scientific explanations is that they generalize to cases beyond those for which they were initially proposed. This is a seeming virtue of invoking general laws in explanations. Laws are commonly represented in universal conditional statements, and hence apply to any situation in which the antecedent of the conditional is satisfied. So generalization is automatic.[15] A well-known example is Newton's first law of motion: If there is no force on a body, its momentum will remain constant. This applies to any body and its conclusion can be applied to any body with no force operating on it. In contrast, models of mechanisms can be highly specific, taking account of the particular factors at work in a specific case in which a phenomenon is studied. As research proceeds, scientists find variants of what initially might seem to be the same mechanism, for example, the mechanisms responsible for oxidative phosphorylation in liver versus heart cells in cows. The mechanistic approach seems to make explanation—discovery of the mechanism responsible for a phenomenon—highly context bound. What sort of generality is then possible?

To address this issue, we can consider a related problem from another domain, that of concepts and categorization. Most philosophers and psychologists prior to the early 1970s construed concepts as having definitions that provided necessary and sufficient conditions for satisfaction of the concept. An exception was Wittgenstein (1953), who objected that it did not seem possible to provide definitions even for such ordinary concepts as *game*. He suggested that instances of a concept might not share any distinctive common properties, but rather merely resemble each other in the way members of a family do. Psychologist Eleanor Rosch provided an empirical foundation for this idea by showing that people could rate the typicality of instances with respect to a category and that these ratings predicted performance measures such as response time to verify category membership (Rosch, 1975, 1978). With respect to *bird*, for example, robins are highly typical, chickens atypical, and penguins highly atypical exemplars. These findings posed a challenge to the traditional view, since if defining features marked the set of items satisfying a concept, all instances that possessed the features should be equally good exemplars of it. Alternatives emerged, especially exemplar theories, in which prototypes play a key role and are themselves based on the best exemplars (Rosch, 1975), properties of exemplars, or both (Smith & Medin, 1981). Membership in a category is a matter of degree, based on similarity to its prototype.

Prototype and exemplar theories suggest a way to approach the issue of generalizing mechanistic explanations. Different mechanisms may exhibit similarity relations to each other without being exactly the same. For example, mechanisms of protein synthesis may be similar in different organisms or different cell types in the same organism without being identical. Certain memory encoding mechanisms, to take another example, may be similar across some delimited range of species.

Appeals to similarity have been notoriously suspect in philosophy in the wake of the objections raised by Goodman (1955). There are an infinite number of respects in which one can judge similarity and any two objects in the universe are similar to others on some of these dimensions. To make the appeal to similarity precise, the dimensions and metrics need to be specified. Yet, without making the dimensions and metrics precise,

[15] It was in fact this generalizability of laws that led David Hull (1978) to deny that there were laws about biological species once he argued that species were in fact individuals, not generalized classes.

scientists do make judgments about similarity. They seem to have an intuitive sense of which dimensions are pertinent and which are not.

I started this section by noting that nomological explanation provided generalization in a straightforward manner. The need to invoke similarity relations to generalize mechanistic explanations seems to be a limitation of the mechanistic account. But in fact it may be the mechanistic account that provides a better characterization of how explanations are generalized in many sciences. Laws are generalized by being universally quantified and their domain of applicability is specified by the conditions in their antecedents. On this account, no instance better exemplifies the law than any other. But in actual investigations of mechanisms, scientists often focus on a specific exemplar when first developing their accounts. Such an exemplar is often referred to as a *model system* and may be chosen for a variety of reasons. For example, much of the research on neural transmission was conducted on the giant squid axon because its size rendered it easy to study. Many investigations of oxidative metabolism focused on cow heart mitochondria since cow hearts were readily available from slaughter houses and mitochondria are plentiful in heart tissue. Choices of model systems sometimes are rooted in, and maintain, differences of tradition or orientation between closely related disciplines. For example, cell biologists who had developed techniques for cell fractionation with mammalian cells used liver and pancreatic cells from rats to study microsomes and their relation to protein synthesis. Molecular biologists, in contrast, generally preferred to work with bacteria and bacteriophages and did so in developing their own models of protein synthesis.

Having used a particular model system to initiate investigation into the mechanism responsible for a given phenomenon, researchers eventually need to determine experimentally how well their accounts will generalize. Examining the counterpart mechanism in other organs and species, any differences can be identified and their importance can be assessed. Unlike research in the D-N framework, in which the conditions of application of a proposed law are incorporated by refining its antecedent, variations are articulated in the description of the mechanism itself. For example, it may be that two variations on the mechanism exist in which a minor part is or is not included, and that this has a small but systematic impact on several component operations and their coordination. Or minor variations on what is essentially the same operation may be found. Findings of this kind are a regular part of the scientific literature in biology. Such papers do not serve only the traditional role of independent confirmation of a theoretical idea—they also identify variations in the mechanism and their significance.

4. Conclusions

Explanation in the life sciences often takes the form of identifying the mechanism responsible for a given activity. Producing a mechanistic explanation is a cognitive activity involving representing and reasoning about nature. But mechanistic explanations are different in many respects from nomological explanations. First, linguistic representations are not privileged and often diagrams provide a better vehicle for representing mechanisms. Making inferences about mechanisms often involves simulating the operation of the mechanism, not generating logical deductions. Second, the fact that mechanisms consist of organized systems of components entails that discovering mechanistic explanations involves procedures for decomposing and modeling mechanisms. Thus, it is possible to articulate procedures for discovering mechanistic explanations. Moreover, the inferences invoked in testing of mechanistic explanation are often much more constrained than those used to test nomological explanations. Investigators focus on tests that are likely to be diagnostic of the operation and organization of components. Third, generalization involves not just applying the same law to different conditions, but identifying the similarities and differences between mechanisms operative in different circumstances.

References

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512.

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

Barwise, J., & Etchemendy, J. (1995). Heterogeneous logic. In B. Chandrasekaran & J. Glasgow & N. H. Narayanan (Eds.), *Diagrammatic reasoning: Cognitive and computational perspectives*. Menlo Park, CA: AAAI Press.

Beatty, J. (1995). The evolutionary contingency thesis. In G. Wolters & J. Lennox (Eds.), *Theories and rationality in the biological sciences, The second annual Pittsburgh/Konstanz colloquium in the philosophy of science* (pp. 45-81). Pittsburgh: University of Pittsburgh Press.

Bechtel, W. (1986). Biochemistry: A cross-disciplinary endeavor that discovered a distinctive domain. In W. Bechtel (Ed.), *Integrating scientific disciplines* (pp. 77-100). Dordrecht: Martinus Nijhoff.

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

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

Bechtel, W. (2001). The compatibility of complex systems and reduction: A case analysis of memory research. *Minds and Machines, 11*(483-502).

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

Bechtel, W. (forthcoming). Mechanism and biological explanation.

Bechtel, W. (in press-a). *Discovering cell mechanisms*. Cambridge: Cambridge University Press.

Bechtel, W. (in press-b). The epistemology of evidence in cognitive neuroscience.

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

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

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

Causey, R. L. (1977). *Unity of science.* Dordrecht: Reidel.

Craver, C. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science, 69*, S83-S97.

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

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

Cummins, R. (2000). "How Does It Work?" versus "What Are the Laws?": Two Conceptions of Psychological. In F. Keil & R. Wilson (Eds.), *Explanation and cognition*. Cambridge, MA: MIT Press.

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

Darden, L., & Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in the History and Philosophy of the Biological and Biomedical Sciences, 33*, 1-28.

de Duve, C. (1969). The lysosome in retrospect. In J. T. Dingle & H. B. Fell (Eds.), *Lysosomes in biology and pathology* (pp. 3-40). Amsterdam: North Holland.

Fernández-Morán, H., Oda, T., Blair, P. V., & Green, D. E. (1964). A macromolecular repeating unit of mitochondrial structure and function. Correlated electron microscopic and biochemical studies of isolated mitochondria and submitochondrial particles of beef heart muscle. *The Journal of Cell Biology, 22*, 63-100.

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

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

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

Goodman, N. (1955). *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084-1102.

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Macmillan.

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

Hull, D. L. (1978). A matter of individuality. *Philosophy of Science, 45*, 335-360.

Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, England: Cambridge University Press.

Jonker, C., Treur, J., & Wijngaards, W. C. A. (2002). Reductionist and anti-reductionist perspectives on dynamics. *Philosophical Psychology, 15*, 381-409.

Kaneko, K., & Tsuda, I. (2001). *Complex systems: Chaos and beyond*. Berlin: Springer.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review, 88*, 46-66.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). Cambridge: Cambridge University Press.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative process.* Cambridge: MIT Press.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

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

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

Palade, G. E. (1952). The fine structure of mitochondria. *Anatomical Record, 114*, 427-451.

Palade, G. E., & Claude, A. (1949a). The nature of the Golgi apparatus. II. Identification of the Golgi apparatus with a complex of myelin figures. *Journal of Morphology, 85*, 71-111.

Palade, G. E., & Claude, A. (1949b). The nature of the Golgi apparatus. I. Parallelism between intercellular myelin figures and Golgi apparatus in somatic cells. *Journal of Morphology, 85*, 35-69.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 111-133.

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.

Reichenbach, H. (1966). *The rise of scientific philosophy*. Berkeley: University of California Press.

Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General, 104*, 192-233.

Rosch, E. (1978). Principles of categorization. In E. Rosch & C. Mervis (Eds.), *Cognition and categorization* (pp. 24-48). Hillsdale, NJ: Erlbaum.

Rosenberg, A. (1994). *Instrumental biology and the disunity of science*. Chicago: University of Chicago Press.

Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (in press). A universal definition of life: Autonomy and open-ended evolution. *Origins of Life*.

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

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Tabery. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science, 71*, 1-15.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: MacMillan.

Figure 1: The heart pumping blood. RA: right aorta; LA: left aorta; RV: right ventricle; LV, left ventricle; T: tricuspid valve; M: mitral valve; P: pulmonary valve; A: aortic valve.

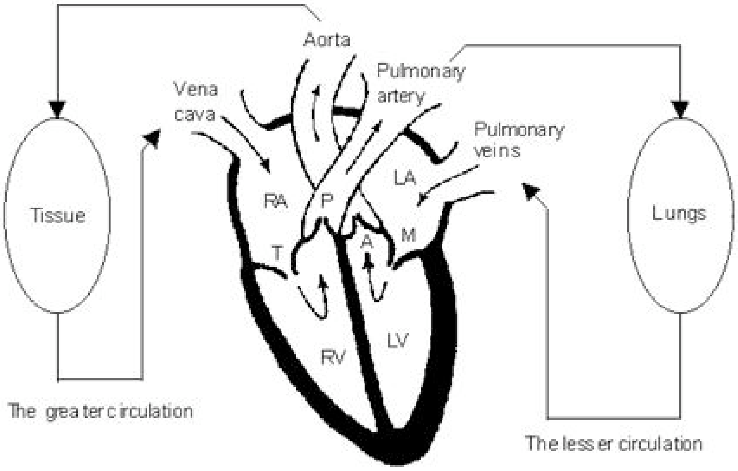


Figure 2: Feedback loop in the linkage between the end of the glycolysis pathway and the Krebs cycle. Phosphoenolpyruvate is usually metabolized to Pyruvate, and that in turn to Acetyl-CoA (dark arrows), but if Acetyl-CoA accumulates, it feeds back (dotted arrow) to inhibit Pyruvate kinase, the enzyme responsible for the first step in the reaction, thereby halting the process.

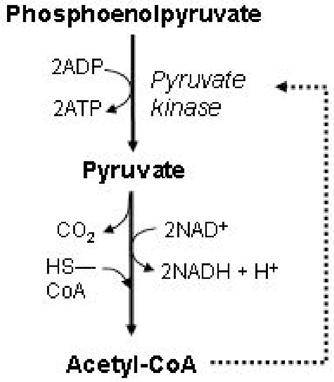


Figure 3: A partial structural decomposition of the cell. The mitochondrion is an organelle located in the cell cytoplasm. The inner membrane of the mitochondrion folds into the inner matrix of the mitochondrion, creating cristae, on which are located small knobs that contain the enzyme ATPase.

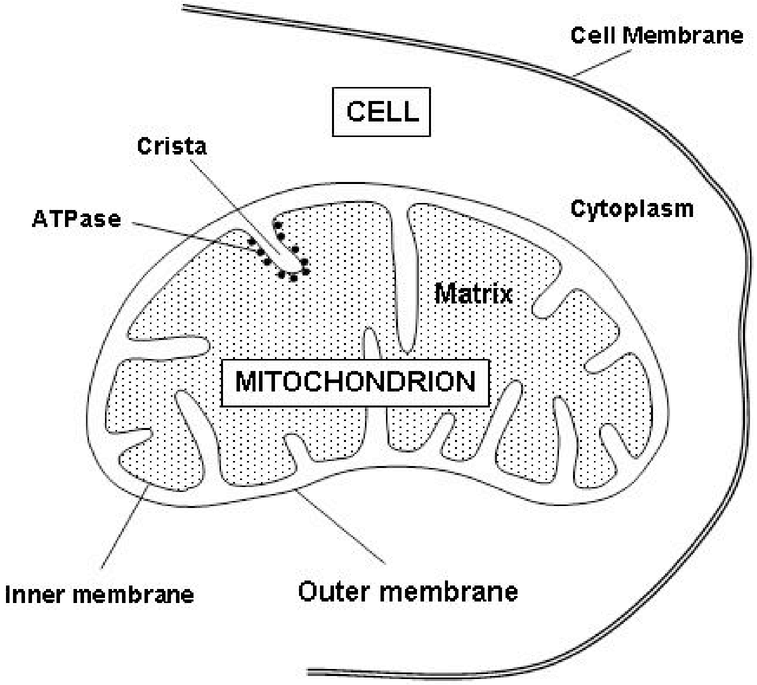
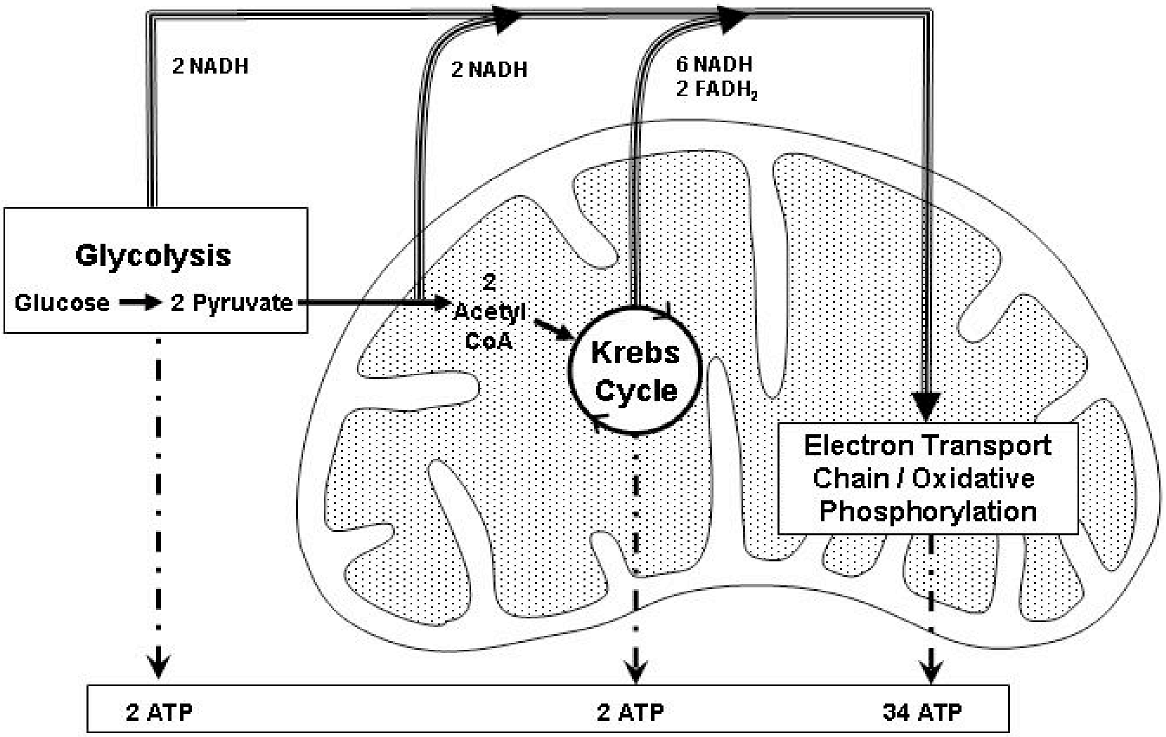


Figure 4: The three major operations of energy metabolism localized in parts of the cell: glycolysis in the cytoplasm, the Krebs cycle in the mitochondrial matrix, and the electron transport chain and oxidative phosphorylation in the cristae of the mitochondrion. Also shown is the energy harvest from each operation.


2 NADH
2 NADH
6 NADH
2 FADH$_2$
Glycolysis
Glucose → 2 Pyruvate
2 Acetyl CoA
Krebs Cycle
Electron Transport Chain / Oxidative Phosphorylation
2 ATP
2 ATP
34 ATP