

---

# MECHANISM

---

Interest in mechanisms has experienced a recent upsurge in the philosophy of science generally (e.g., Salmon 1984; Glennan 1996) and in the philosophy of biology and neuroscience in particular (see e.g., Bechtel and Richardson 1993; Craver and Darden 2001; Machamer, Darden, and Craver 2000). Scientific explanation often involves identifying the mechanism responsible for a phenomenon of interest. This entry provides a generic account of what mechanisms are and how they are appealed to in explanations and then turns to the question of how scientists discover them. What are mechanisms?

## Four Aspects of Mechanisms

The notion of mechanism has four aspects: (i) a phenomenal aspect, (ii) a componential aspect, (iii) a causal aspect, and (iv) an organizational aspect. Mechanisms can differ from one another in each of these aspects. Consider them in turn, first using a common mousetrap as an example and then considering the more complicated mechanism of action potential generation in neurons:

### *The Phenomenal Aspect*

Mechanisms do things; they are the mechanisms *of* the things that they *do*. A mousetrap traps mice, and the mechanism for generating action potentials generates action potentials. These tasks performed by the mechanism as a whole are the phenomena explained by the working of the mechanism. There are no mechanisms simpliciter—only mechanisms *for* phenomena. A mechanism's phenomenon partially determines the mechanism's boundaries (i.e., what is "in" the mechanism and what is not). As Kauffman (1971) clearly emphasized, an item is considered "part of" the mechanism only if it is relevant to a mechanism's phenomenon.

### *The Componential Aspect*

Mechanisms have components, or working parts. Mechanisms all have at least two components. The old-fashioned mousetrap has six: a platform, a trigger, a latch, a catch, a spring, and an impact bar (see Figure 1). Trivially, the components are proper parts of the mechanism as a

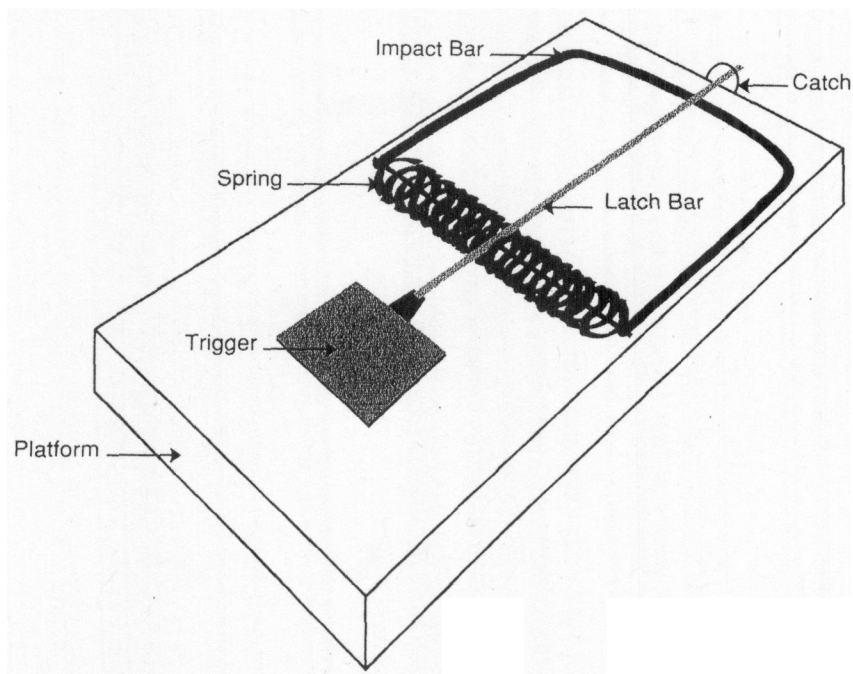


Fig. 1. The Mousetrap.

whole. More restrictively, as just noted, the parts of a mechanism are those that are relevant to the phenomenon explained by the mechanism. The parts are relevant to the phenomenon by virtue of certain of their properties (and not others). But for the rigidity of the bar and the tension on the spring, a mousetrap would catch no mice. The buoyancy of the platform, in contrast, is not properly included in the mechanism for catching mice.

#### *The Causal Aspect*

The components of mechanisms act and interact with one another. If they did not, they would not do anything. *Pressing* the trigger *releases* the catch, *allowing* the spring to *launch* the impact bar. The verbs in this description of the mousetrap refer to the relevant causal relations among the component parts. Talk of causal relations is a schematic placeholder to be filled in with one or more appropriate accounts of the kinds of causing exhibited in a given case. Philosophical attempts to develop univocal analyses of such causal relationships have yet to garner widespread acceptance. Yet the intransigence of the causal relation to a single uniform philosophical analysis should not distract attention from the central role that causal relations play in mechanistic explanations.

#### *The Organizational Aspect*

The components of mechanisms and their causal relations are organized spatially and temporally in the production of the phenomenon. The *spatial organization* of a mechanism includes the relative locations, shapes, sizes, orientations, connections, and boundaries of the mechanism's components. In the mousetrap, the trigger and the catch have to be so located with respect to one another that a small amount of pressure on the trigger moves the trigger bar enough to dislodge from the catch. The catch is circular and accommodates the size of the trigger bar. When the mechanism is loaded, the parts are connected to one another: The trigger bar restrains the blunt bar because it is stuck in the catch. As the mousetrap "fires," temporal organization takes center stage. The temporal organization of a mechanism includes the order, rates, durations, and frequencies of the activities in the mechanism. If a mousetrap is to work, it should work quickly, it should not discharge until there is pressure on the trigger, and there should not be significant delays between the steps of its working. Spatial and temporal organization are two important varieties of mechanistic organization. There are familiar patterns of mechanistic organization that can be found in different mechanisms for different phenomena. Some mechanisms are feed-forward, with

each step following upon its predecessor without forks, joins, or cycles (like the common mousetrap); others may work in parallel or have significant feedback connections.

### A Neurobiological Example: The Mechanism of the Action Potential

Mousetraps fire, and so do neurons. The firing of a neuron is known as an action potential. Action potentials are changes in the electrical potential difference across the cell membrane that propagate along the length of the neuron. This difference, known as the membrane potential ( $V_m$ ), consists of the separation of charged ions on either side of the membrane. In the neuron's resting state, positive ions line up against the membrane's extracellular surface, and negative ions line up on the intracellular side, producing a polarized resting potential ( $V_{rest}$ ) of roughly  $-60$  mV. The action potential (as indicated in Figure 2) consists of (I) a rapid rise in  $V_m$  (reaching a maximum value of roughly  $+20$  mV), followed by (II) an equally rapid decline in  $V_m$  to values below  $V_{rest}$ , and then (III) an extended hyperpolarized afterpotential during which the neuron is less excitable. These three features characterize the phenomenon to be explained by the action potential mechanism.

The components of this mechanism include the cell membrane, positively charged sodium ( $\text{Na}^+$ ) ions, positively charged potassium ( $\text{K}^+$ ) ions, and two types of voltage-sensitive ion channels that selectively allow, respectively,  $\text{Na}^+$  or  $\text{K}^+$  ions to diffuse through the membrane. It is the temporally organized activities of these channels that produce the action potential phenomenon.

The mechanism of the action potential starts with a cumulative depolarization of the cell body (i.e.,  $V_m$  becomes greater than  $V_{rest}$ ), typically through the effect of neurotransmitters on ion channels in the cell's dendrites (the "receiving" ends of the neuron). Action potentials are generated in the axon hillock, an ion-channel-dense region of membrane at the interface of the cell body and the axon (the "sending" end of the neuron). Depolarization of the cell body opens voltage-sensitive  $\text{Na}^+$  channels (increasing membrane conductance to  $\text{Na}^+$ ), allowing  $\text{Na}^+$  to diffuse down its concentration gradient from the  $\text{Na}^+$ -rich extracellular fluid into the relatively  $\text{Na}^+$ -poor intracellular fluid (illustrated by the membrane conductance curve for  $\text{Na}^+$  in Figure 2). The resulting flood of  $\text{Na}^+$  drives the voltage of the cell toward the  $\text{Na}^+$  equilibrium potential ( $E_{\text{Na}}$ ; roughly  $+55$  mV), accounting for the rapid rising phase of the action potential (I).

This rapid depolarization of the membrane has two consequences that account for the declining phase of the action potential (II). The first is the inactivation of the  $\text{Na}^+$  channel, which slows and eventually stops the ascent of  $V_m$  toward  $E_{\text{Na}}$ . The second is the delayed activation of voltage-sensitive  $\text{K}^+$  channels, increasing the  $\text{K}^+$  conductance of the membrane and allowing  $\text{K}^+$  to diffuse down its concentration gradient from the  $\text{K}^+$ -rich intracellular fluid into the  $\text{K}^+$ -poor extracellular fluid. This diffusion of  $\text{K}^+$  drives the membrane potential back down toward the  $\text{K}^+$  equilibrium potential ( $E_{\text{K}}$ ; roughly  $-75$  mV) and even below the resting potential of the membrane.

Thus begins the final, afterpotential phase of the action potential (III), which is characterized by both the hyperpolarization of the membrane (i.e.,  $V_m$  is

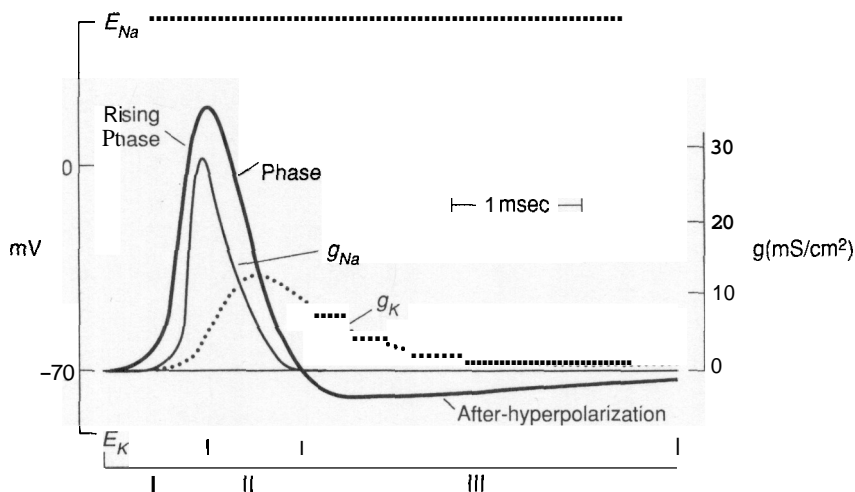


Fig. 2. The Action Potential.

lower than  $V_{m,m}$ ) and a period of reduced excitability. The membrane hyperpolarizes after the action potential because  $K^+$  channels are slow to return to their resting closed state. The residual  $K^+$  conductance tugs  $V_m$  away from  $V_{rest}$  and toward  $E_K$ .

The parts in the mechanism for generating action potentials are the membrane, the ions, and the ion channels. These parts are causally connected; they act and interact in regular ways to produce the action potential. These activities depend crucially upon the spatial organization of the components; ion channels *span* the membrane, allowing ion *movement* between the intracellular and extracellular fluids. Spatial organization is also fundamental to understanding the molecular mechanisms of channel activation and inactivation and for understanding the propagation of action potentials along axons. Yet, it is temporal organization that is most evident in the mechanism of the action potential; it is the relative orders and durations of the activation and inactivation of  $Na^+$  and  $K^+$  channels that explain the characteristic waveform (I, II, and III) of the action potential.

### Levels of Mechanisms

Often mechanisms are nested within mechanisms. In such cases, some phenomenon ( $\psi$ ) of a mechanism ( $M$ ) is explained by the organized activities ( $\phi$ ) of lower-level components ( $X$ ) that can themselves be taken as phenomena to be explained by the activities ( $\rho$ ) of still lower level Components ( $Z$ ). Thinking about mechanisms provides a straightforward way to think about levels (see Craver 2001b). In this case, the relationship between lower and higher mechanistic levels is a compositional relationship with the additional restriction that the lower-level parts are components of (and hence organized within) the mechanism at the higher level. The requirement that lower-level parts be organized (at least spatially and temporally) within the higher-level mechanism distinguishes mechanistic levels from mere aggregates, such as piles of sand (Wimsatt 1986); from mere collections of improper parts, such as the cubes into which a television might be arbitrarily sliced (Haugeland 1998); and from mere inclusive sets, such as the collected songs of the Ramones. Lower mechanistic levels are entities and activities organized to exhibit the behavior of the mechanism as a whole.

Mechanistic levels should not be confused with intuitive ontic levels (e.g., Oppenheim and Putnam 1958), which map out a monolithic stratigraphy of levels across theories, sciences, and types of entities. Just as there are no mechanisms *simpliciter*,

there are no mechanistic levels *simpliciter*. Mechanistic levels, instead, are defined only with respect to some highest-level mechanism  $M$  and its phenomenon  $\psi$  (pronounced “psi”). This, however, does not mean that the investigator cannot move upward, treating  $M$  as part of a yet higher level mechanism that generates its own phenomenon. Different levels of a mechanism involve different entities and activities. Accordingly, different vocabularies are typically used to describe mechanisms at different levels (Bechtel 1995). Exactly how many levels there are and how they are to be individuated are empirical questions that are answered differently for different phenomena.

### Representing Mechanisms

There are many conventions for describing and representing mechanisms. Verbal accounts are generally insufficient to convey an understanding of a mechanism, especially if there are any nonlinearities in its behavior. Accordingly, verbal descriptions are often accompanied by diagrams representing the components, their activities (often depicted with arrows), and the relevant features of their organization (see Figure 1 above). Temporal relations are often represented spatially, either with labeled events conjoined by arrows or in separate frames. Diagrams afford the viewer the opportunity to follow through the parallel sequences of activities within the mechanism in one glance. With increasing frequency, the working of a mechanism may be represented in animated shorts. Extremely complicated mechanisms, however, frequently require the viewing time afforded by static two-dimensional representations so that aspects of the mechanism can be taken in piecemeal.

Descriptions of mechanisms, whether verbal or pictorial, may be more or less gappy, with holes or question marks to be filled in as details of the mechanism are discovered. Sometimes these are appreciated by the person portraying the mechanism, but many times the gaps are not even recognized until, for example, another component is discovered and researchers try to figure out what it contributes. Descriptions of mechanisms may also be more or less abstracted from the details of the operation of any particular mechanism, highlighting broad patterns of organization (e.g., with equations) or exhibiting precisely the spatial, temporal, and hierarchical organization of the components and activities of the mechanism.

Often the activities within a mechanism are characterized mathematically. For example, in describing the action potential mechanism, equations

are advanced describing the changes in magnitude of  $\text{Na}^+$  concentrations over time. Once such equations are developed, mathematical models of the overall operation of the mechanism can be advanced.

### Mechanistic Explanations

Since mechanisms are often responsible for generating phenomena for which explanations are sought, it is not surprising that scientists frequently advance accounts of mechanisms as explanations. That is, to explain an action potential, they proceed much as in the example above—identifying the components of the responsible mechanism, describing the activities performed by the components, and showing how these components and activities are organized. They frequently present this information in diagrams, and often the account offered is gappy. Presenting a mechanism as an explanation, however, does not fit the standard deductive-nomological account of explanation, according to which explanation involves deriving a statement of the phenomenon to be explained from laws and relevant initial conditions. It is not laws that do the explanatory work but the account of the operation of the mechanism.

One might try to reconcile the two accounts of explanations by insisting that there is a law characterizing each mechanism. Typically, however, there is too much variability in a given mechanism (e.g., in the generation of action potentials in different neurons) for this to be plausible. It is better to recognize mechanistic explanation as an alternative model of explanation. Its prevalence in a variety of sciences such as physiology and neuroscience may account for the fact that these sciences have not been the primary source of examples of deductive-nomological explanation and have been relatively neglected by philosophers of science. Once mechanisms are recognized for their explanatory role in these sciences, though, we can also identify a number of philosophical issues to be pursued. One of these concerns their discovery.

### How Are Mechanisms Discovered?

#### *Characterizing the Phenomenon*

One of the first tasks in discovering mechanisms is identifying the phenomenon—determining what it is that the mechanism does. The world does not come obviously prepackaged in terms of phenomena. How one characterizes the phenomenon critically affects how one goes about trying to discover the responsible mechanism and whether

that quest will prove successful. Accordingly, characterizations of phenomena prove controversial and frequently are revised in the course of inquiry as one discovers that the mechanism does something different than one thought.

Phenomena are often subdivided, consolidated, or reconceptualized entirely as the discovery process proceeds. Researchers may recognize the need to subdivide a phenomenon into many distinct phenomena, as when learning and memory researchers were forced to recognize that there were many different kinds of memory requiring more or less distinct mechanisms to explain them. Alternatively, researchers may be forced to consolidate many different phenomena into a single phenomenon, as when it became understood that burning, respiring, and rusting were all due to a common mechanism and thus are examples of one phenomenon, oxidation. Finally, investigators may need to reconceptualize the phenomenon to be explained entirely. For example, early physiologists focused on the fact that animals burn foodstuffs and release heat. But after further investigation, researchers recharacterized this phenomenon as transforming energy into usable forms (e.g., ATP bonds).

#### *Identifying Components*

The discovery of mechanisms also involves identifying the components of the mechanism and their activities. Bechtel and Richardson (1993) used the term *decomposition* to describe analysis of a phenomenon into activities that, when properly organized, exhibit the phenomenon. In one of their main examples, they describe how the biological process of fermentation, over three decades of research, was decomposed into a set of more basic chemical reactions (oxidations, reductions, phosphorylations, etc.). This is *functional decomposition*. But frequently the process of decomposition begins by breaking the mechanism apart into component entities and only then investigating what the components do. This is *structural decomposition*. Ultimately, one measure of the adequacy of either form of decomposition is that it maps onto the other so that specific components are related to particular activities. Bechtel and Richardson call this identification of activities with components *localization*.

Often the search for the components of a mechanism is guided by an accepted store of components and activities that are reasonably well understood by a science at a particular time and that are available for use in thinking about how a mechanism works (Craver and Darden 2001). In the early stages of mechanism discovery, there may be no

such store; there is either no idea of, or considerable controversy over, what the components and activities might be. The brain provides a useful example (Mundale 1998). There has been considerable controversy, for example, about what counts as a brain region, with different investigators using different criteria to divide the brain into parts at different times. Early attempts to map brain areas focused on the sulci and gyri resulting from the folding of cortex. Although prominent features of the brain, these tend not to be closely linked to component activities. With the identification of different types of neurons and the existence of cortical layers of varying thicknesses, numerous early-twentieth-century scientists, including Korbinian Brodmann, used these cytoarchitectural features to demarcate brain areas. Brodmann explicitly thought that different areas were likely to perform different operations, but he lacked any means for linking the regions he differentiated with function. More recent brain mappers have invoked yet additional criteria such as connectivity to other regions to identify brain areas. A major reason for controversy over these is that researchers are interested in components that perform the activities that generate the relevant phenomena. As the sulci and gyri of the brain illustrate, it is possible to differentiate structures within a mechanism that are not working components, that is, parts that carry out the relevant activities. In the relevant sense, these are not components of the mechanism. Similar challenges arise in functional decomposition— one may propose a decomposition into activities but not ones performed by any of the mechanism's components. Moreover, as hinted above, the search for components and for activities is interdeterminate— conceptions of the activities thought to be performed guide the identification of components, and vice versa.

How do scientists arrive at satisfactory decompositions that describe mechanisms in terms of their organized parts and activities? Often scientists begin the discovery process by proposing that there is a single component in the mechanism that alone is responsible for the phenomenon (e.g., attributing pleasure to activation of the brain's pleasure center). Sometimes this claim is correct, but even when it is, the task of identifying the mechanism that generates the phenomenon awaits decomposition of that component itself.

True decomposition is frequently guided either by the available store of components or by the available tools for investigating these components. Often scientists, functioning much like engineers, attempt to organize known components and activities in such a way that they might possibly produce the

phenomenon. This process may involve reasoning analogically from other mechanisms (discovered in nature or human artifacts) and the activities performed in them. Such "how possibly" reasoning is, of course, fallible, since even two phenomena that are very similar may be generated by two very different mechanisms. In fact, sometimes the discovery process is slowed dramatically by pursuit of false leads generated by this engineering heuristic. On the other hand, even an erroneous proposal often advances the inquiry, since now experimental evidence can be generated that points to a more adequate decomposition. Experimental strategies for decomposing a mechanism are discussed further below.

### Discovering the Organization of a Mechanism

Beyond delineating the phenomenon and revealing the components, a third major goal in the discovery of a mechanism is to determine how these components and activities are organized in the mechanism. Typically there are both spatial and temporal aspects to the organization of a mechanism. For example, the rate and duration of the phenomenon places time constraints on the activities of the components, and uncovering the order, rate, and duration of the steps in a mechanism often provides important clues into how the mechanism works. Likewise, discovering aspects of the spatial organization of a mechanism (the size, shape, position, orientation, etc., of the components) is often crucial for suggesting possible mechanisms and for ruling out others (see Craver and Darden 2001).

The relative importance of spatial and temporal organization varies from mechanism to mechanism. Spatial organization is of fundamental importance in, for example, the mechanisms of enzyme degradation because enzymes that can break down cellular substances need to be kept separate from other cellular substances that are not to be broken down. Spatial organization also helps provide efficiency in production mechanisms in which intermediate products are literally passed from one activity to the next (as in the Krebs's cycle).

If a phenomenon involves a change from one state or set of conditions to another (e.g., from glucose to alcohol, from sensory stimulus to recognition), it is common to think of that change as being executed by a linear sequence of steps. In part this is common because human conscious cognitive activities are serial— humans proceed from thinking of one thing to thinking of another. But, for very good reasons, such as ensuring proper regulation of a process, many natural mechanisms

are not organized linearly. As a result, they are difficult for humans to conceptualize, at least without the aid of external representations such as diagrams in which one can represent backward as well as forward linkages.

The naturalness of linear organization means that in trying to fit multiple parts and activities together into a coherent description of a mechanism, researchers often begin by trying to organize them linearly. Often researchers begin to appreciate more complex modes of organization only when these attempts fail to account for the phenomenon. In modeling a chemical process, for example, one may find that there is no way to link together known basic reactions to get from the initial input to the product. This often leads to the exploration of more complex modes of organization such as a cycle. Thus, one common pattern in the process of discovering mechanisms is to begin with linear organization and then add complexity as required.

### Experiments in Mechanism Discovery

Typically, the components, activities, and organization of a mechanism cannot be understood without the aid of well-designed experiments. Experimentation figures not just in the testing of models of mechanisms that have been hypothesized independently, but in the very process of discovering the mechanism.

Experimentation requires some means of intervening in the operation of the mechanism as well as a means of recording the effects of those interventions. Sometimes interventions into a mechanism are performed "by nature," through accidental damage, disease, or genetic mutation or variation.

Other times the interventions are intentional and designed by the researcher to perturb some isolated aspect of the phenomenon or some component or activity in the mechanism.

A taxonomy of experimental approaches to developing and testing descriptions of mechanisms can be developed by focusing on where the intervention and recording techniques are applied (Bechtel and Richardson 1993; Craver 2001a). In the sense discussed above, a phenomenon and the mechanism that produces it are at two mechanistic levels, the phenomenal level ( $L_P$ ) and the level of the mechanism ( $L_M$ ) (see Figure 3). As illustrated in Figure 3, experiments may intervene and record entirely at the phenomenal level, bridge phenomenal and mechanistic levels, or intervene and record entirely within the mechanistic level.

First, both the intervention and the recording may be conducted at  $L_P$  without going down to  $L_M$  (see Figure 4). For example, one can intervene to vary the inputs to a mechanism or the conditions under which it operates (e.g., temperature) and record variations in the phenomenon. Much experimentation in cognitive psychology (e.g., requiring subjects to perform a task under varying conditions, such as cognitive load, and using reaction time as the measure of the effect) is of this sort and, when done well, can provide abundant information about the internal design of the mechanism. For example, evidence that two tasks interfere with each other provides further evidence that some component or components may be involved in both tasks. A great deal can also be learned about a mechanism by determining the range of input conditions under which it works properly and under which it fails or malfunctions.

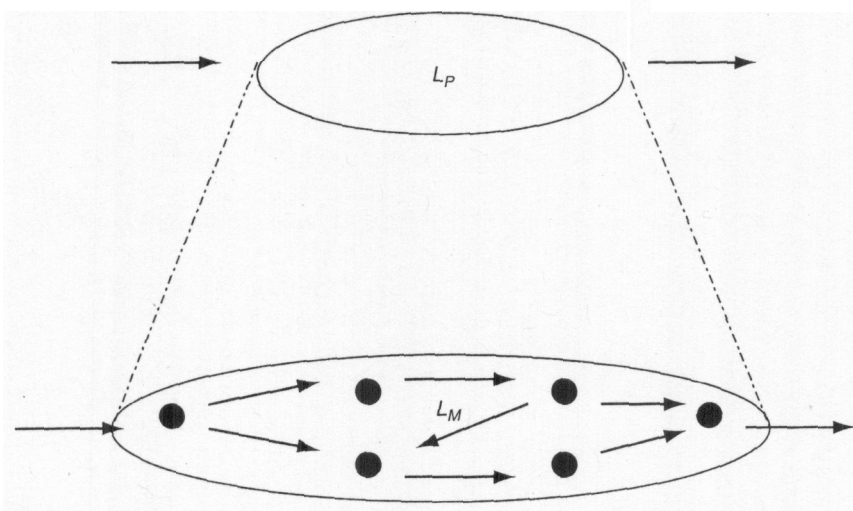


Fig. 3. Phenomenal Level (top) and Mechanism Level (bottom).

## MECHANISM

Second, experiments may bridge  $L_P$  and  $L_M$ . (Many experiments bridge several such levels at once.) Such experiments may be top-down (intervening at  $L_P$  and recording at  $L_M$ ) or bottom-up (intervening at  $L_M$  and recording at  $L_P$ ), and the experimental intervention may be either excitatory (somehow stimulating the target of the intervention) or inhibitory (somehow removing or impairing the target of the intervention). Top-down excitatory experiments are prevalent in cognitive neuroscience, where researchers intervene to engage an organism in some cognitive task while recording the activities of component brain regions, neurons, or molecules. Bottom-up excitatory experiments are also common. Neural stimulation studies, for example, use electrodes to excite individual neurons, and the effects are recorded for the cognitive phenomenon in which those neurons are involved. Additionally, bottom-up inhibitory experiments are a staple of most sciences that search for mechanisms. In neuroscience, for example, one may intervene to remove a brain region, a receptor molecule, or a neurotransmitter and record the effects on the phenomena in which those components are putatively involved. It is not uncommon for researchers to find a way to impair the activity before they figure out what the relevant components are or which are being affected. For example, one can discover a chemical poison that impairs a metabolic process but not know what component of the mechanism the poison is acting upon.

Third, inhibitory and excitatory techniques can also be applied within  $L_M$ . In this case, one intervenes to excite or inhibit some component or activity in the mechanism and then records the results of that intervention elsewhere in the mechanism. This form of experiment is especially important for determining how the components of the mechanisms are organized together in the production of the phenomenon.

There are significant epistemological challenges in interpreting the results of excitatory and inhibitory interventions into the working of the mechanism. Bottom-up inhibitory experiments may be foiled by redundancy, reorganization, and failures of specificity in the intervention. Intervention to remove or inhibit a component or activity may result in little or no change to the phenomenon if the removed or inhibited component is redundant (like the human kidney). Likewise, the mechanism may reorganize in the face of a loss of its component, leaving the phenomenon intact or only mildly transformed. In general, in removing a part of a mechanism and observing the behavior of the mechanism as a whole, researchers learn not what the removed part does but rather what the rest of the mechanism can do in its absence. Finally, the intervention may have nonspecific effects on other components in the mechanism, thereby indirectly altering the phenomenon and foiling the inference from the recorded changes to the function of the inhibited part. This problem is often exacerbated in “natural

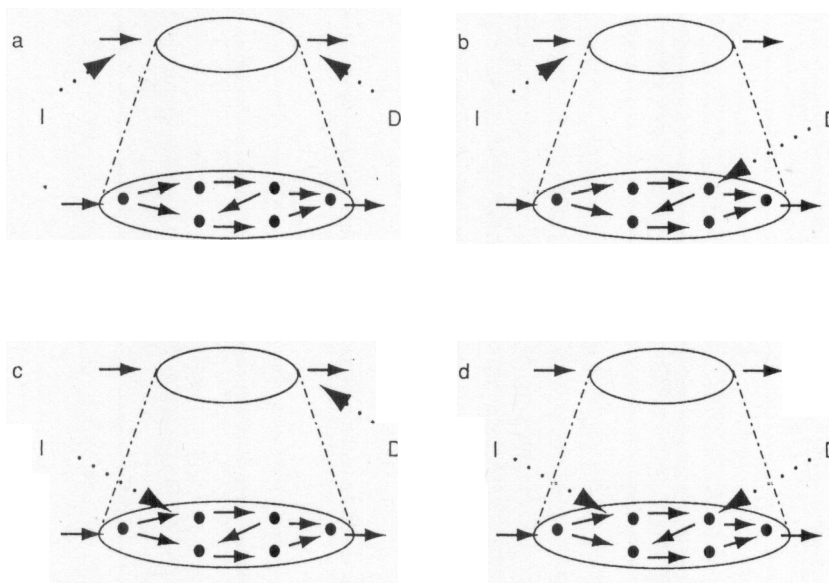


Fig. 4. Points of intervention and recording in experiments. Experiments may (a) both intervene and record at the phenomenal level, (b) intervene at the phenomenal level and record at the mechanistic level, (c) intervene at the mechanistic level and record at the phenomenal level, or (d) both intervene and record at the mechanistic level.



experiments” in which the intervention has not been tightly controlled by an investigator and so may have had a rather nonlocal impact on the components of the mechanism.

Similar epistemological difficulties attend the use of top-down excitatory experimental strategies. One example of such an experiment is to provide a stimulus to an organism and record from individual neurons in its brain or to use neuroimaging to record where there is increased blood flow in the brain. The epistemic challenges here are no less than when the intervention is within the mechanism. Activity in a part of a mechanism when the whole mechanism has been stimulated shows only that the component in question does respond to the stimulation. It does not yet show what activity it performs. Many neurons in the brain, for example, will respond when the organism is presented with a visual stimulus. One can gain more of a clue as to what a component is contributing by varying the intervention and determining the range of interventions to which the component is responsive (e.g., that it is only responsive to visual stimuli moving to the left). Even so, a given active neuron may perform an activity that is largely incidental to the phenomenon one is investigating (e.g., how objects are identified).

One way investigators begin to acquire confidence in their physical and functional decompositions is by drawing upon multiple modes of investigation, especially by invoking both inhibition and excitatory interventions. If lesioning a component eliminates the phenomenon of interest and exciting it produces the phenomenon, compelling evidence is provided that the component figures in generating that phenomenon. But just what does it contribute? Often answering that question depends on formulating a hypothesis about what many different components are contributing, and developing an account of how the components together produce the phenomenon. For example, researchers working on how the brain recognizes objects identified different brain regions in which individual cells would respond to different aspects of a stimulus—some responded whenever a given color was present, another when a given shape was present, and yet others when a particular object was present. By also knowing how these various brain regions were connected to each other, researchers began to piece together an account of the overall mechanism (Bechtel 2001).

As researchers reach the stage of reasonably worked out hypotheses about what different components contribute, additional tools can be invoked to help figure out the mechanism. For example, researchers often begin to build models, including

computational ones, that characterize what each component is thought to contribute and to simulate their interaction. To the degree that the model predicts the phenomenon, one acquires confidence that one’s account is at least close to correct. (The fit between a model and the phenomenon is often a matter of degree, and the degree of fit deemed sufficient often changes as research on the mechanism proceeds.) But failures are equally informative, since they often lead researchers to posit yet unidentified components and activities and begin to seek evidence for them.

Not surprisingly, there is no foolproof procedure for discovering mechanisms. But there are a range of strategies that can be identified by careful examination of actual science.

## Conclusion

Four aspects of mechanisms have been identified: (i) the phenomenal, (ii) the componential, (iii) the causal, and (iv) the organizational. The generation of a phenomenon is often the product of a mechanism, and describing the mechanism provides an explanation of the phenomenon. The sciences concerned with identifying mechanisms have developed a variety of conceptual and experimental tools for this purpose. The philosophical analysis of mechanisms and their discovery is still in a relatively early stage but has advanced far enough that it is safe to predict that careful attention to mechanisms and mechanistic explanation is likely to yield significant advance in the philosophical understanding of science.

CARL CRAVER  
WILLIAM BECHTEL

## References

- Bechtel, William (1995), “Biological and Social Constraints on Cognitive Processes: The Need for Dynamical Interactions Between Levels of Organization,” *Canadian Journal of Philosophy* 20 (supplement): 133–164.
- (2001), “Decomposing and Localizing Vision: An Exemplar for Cognitive Neuroscience,” in Bechtel, Pete Mandik, Jennifer Mundale, and Robert S. Stufflebeam (eds.), *Philosophy and the Neurosciences: A Reader*. Oxford: Basil Blackwell, 225–249.
- Bechtel, William, and Robert C. Richardson (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.
- Craver, Carl (2001a), “Interlevel Experiments and Multi-level Mechanisms in the Neuroscience of Memory,” *Philosophy of Science* 68 (supplement): S83–S97.
- (2001b), “Role Functions, Mechanisms, and Hierarchy,” *Philosophy of Science* 68: 53–74.
- Craver, Carl, and Lindley Darden (2001), “Discovering Mechanisms in Neuroscience: The Case of Spatial Memory,” in Peter K. Machamer, Rick Gush, and Peter

## MECHANISM

- McLaughlin (eds.), *Theory and Method in Neuroscience*. Pittsburgh: University of Pittsburgh Press.
- Glennan, Stuart (1996), "Mechanisms and the Nature of Causation," *Erkenntnis* 44: 49–71.
- Haugeland, John (1998), *Having Thought*. Cambridge, MA: Harvard University Press, chap. 10.
- Kauffman, Stuart A. (1971), "Articulation of Parts Explanation in Biology and the Rational Search for Them," in R. C. Bluck and R. S. Cohen (eds.), *PSA 1970*. Dordrecht, Netherlands: Reidel.
- Machamer, Peter, Lindley Darden, and Carl Craver (2000), "Thinking about Mechanisms," *Philosophy of Science* 67: 1–25.
- Mundale, Jennifer (1998), "Brain Mapping," in William Bechtel and George Graham (eds.), *A Companion to Cognitive Science*. Oxford: Basil Blackwell, 129–139.
- Oppenheim, Paul, and Hilary Putnam (1958), "Unity of Science as a Working Hypothesis," in Herbert Feigl, Grover Maxwell, and Michael Scriven (eds.), *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 3–36.
- Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- Wimsatt, William (1986), "Forms of Aggregativity," in A. Donagan, A. Perovich, and M. Wedin (eds.), *Human Nature and Natural Knowledge*. Dordrecht, Netherlands: Reidel. 259–291.

**See also** **Explanation; Cognitive Science; Neurobiology; Reductionism**