

Theory and Method in the Neurosciences

EDITED BY Peter K. Machamer,
Rick Grush, and
Peter McLaughlin

University of Pittsburgh Press

Contents

Published by the University of Pittsburgh Press,
Pittsburgh, Pa. 15261
Copyright © 2001 University of Pittsburgh Press
All rights reserved
Manufactured in the United States of America
Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

ISBN 0-8229-4140-6

Introduction	1
Peter K. Machamer, Rick Grush, and Peter McLaughlin	
1 The Origin and Development of the Neurosciences	7
Olaf Breidbach	
2 Theory Structure in the Neurosciences	30
Valerie Gray Hardcastle and C. Matthew Stewart	
3 On the Nature of Explanation in the Neurosciences	45
Antti Revonsuo	
4 Mechanisms, Coherence, and the Place of Psychology: Commentary on Revonsuo	70
Stephan Hartmann	
5 Cognitive Neuroscience: Relating Neural Mechanisms and Cognition	81
William Bechtel	
6 Discovering Mechanisms in Neurobiology: The Case of Spatial Memory	112
Carl F. Craver and Lindley Darden	
7 The Explanatory Power and Limits of Simulation Models in the Neurosciences	138
Holk Cruse	

6

Discovering Mechanisms in Neurobiology

The Case of Spatial Memory

Carl F. Craver

Department of Philosophy, Florida International University, Miami, Florida

Lindley Darden

Committee on the History and Philosophy of Science, University of Maryland, College Park, Maryland

This chapter is about discovery in neurobiology; more specifically, it is about the discovery of mechanisms. The search for mechanisms is widespread in contemporary neurobiology, and, understandably, the character of this product shapes the process by which mechanisms are discovered. Analyzing mechanisms and their characteristic organization reveals constraints on their discovery. These constraints reflect, at least in part, what it is to have a plausible description of a mechanism. These constraints also highlight varieties of evidence that both guide and delimit the construction, evaluation, and revision of such plausible descriptions.

The central example in the following discussion is the continuing discovery of the mechanism of spatial memory. Spatial memory, roughly speaking, is the ability to learn to navigate through a novel environment. The mechanism of spatial memory is multilevel, and recently an integrated sketch of the mechanism at each of these levels has started to emerge. Even though this sketch is far from complete at this time, the example offers a glimpse at the kinds of constraints that are delimiting and guiding this gradual and piecemeal discovery process.

This chapter opens with an analysis of mechanism, discussing their components and their characteristic spatial, temporal, and multilevel organization. Mechanisms are often discovered gradually and piecemeal. The second section introduces conventions for constructing

incomplete and abstract descriptions of mechanisms (namely the mechanism sketch and the mechanism schema) and for describing the constraints under which the gradual and piecemeal discovery of these sketches and schemata proceeds. The third section uses the case study of spatial memory to illustrate how constraints on the organization of mechanisms have guided and delimited their discovery. The final section focuses on hierarchical constraints in particular and discusses the use of multilevel experiments to integrate the different levels in such a description. Throughout, the goal is to show that the products, multilevel descriptions of mechanisms, shape the process by which they are discovered.

Mechanisms and Their Organization

Mechanisms

Neurobiologists often speak of “systems” and “cascades” to describe what we call, also consistently with the field’s language, “mechanisms.” Through our collaboration with Peter Machamer, we have come to think about mechanisms as follows. Mechanisms are collections of entities and activities organized in the production of regular changes from start or setup conditions to finish or termination conditions (Machamer et al. 2000). Entities in neurobiology include such things as pyramidal cells, neurotransmitters, brain regions, and mice. Activities are the various doings in which these entities engage: pyramidal cells *fire*, neurotransmitters *bind*, brain regions *process*, and mice *swim* in water while eagerly *searching* for a means of escape. When neurobiologists speak generally about activities, they use a variety of terms; activities are often called “processes,” “functions,” and “interactions.” Activities are the things that entities do; they are the productive components of mechanisms, and they constitute the stages of mechanisms.

Organization

The entities and activities composing mechanisms are *organized*; they are organized such that they *do* something, *carry out* some process, *exercise* some faculty, *perform* some function, or *produce* some end product. We refer to this activity or behavior of the mechanism as a whole as the *phenomenon* to be explained by the description of the

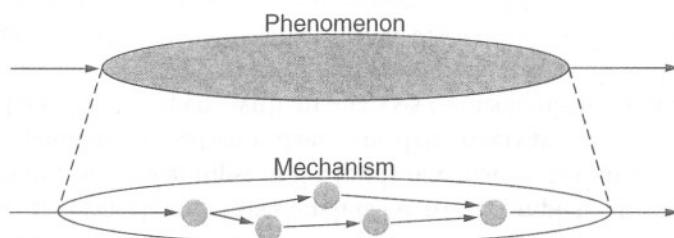


Figure 6.1 Phenomenon and mechanism.

mechanism. This is the activity at the top of figure 6.1. Below it are the entities and activities composing the mechanism for that phenomenon.

The phenomena to be explained by descriptions of mechanisms can be understood in the spirit of Bogen and Woodward (1988, 317). We think of phenomena as relatively stable and repeatable properties or activities that can be produced, manipulated, or detected in a variety of experimental arrangements. Examples of phenomena in neurobiology include the acquisition, storage, and retrieval of spatial memories; the release of neurotransmitters; and the generation of action potentials.

Mechanisms are *organized* in the production of phenomena. One aspect of this organization is *temporal*. The stages of mechanisms have a productive order from beginning to end, with earlier stages giving rise to later stages. The stages of mechanisms also have characteristic rates and durations that can be crucial to their operation. Order, rate, and duration are crucial, for example, for the generation of action potentials in neurons: sodium channels open before potassium channels, and the respective timing and duration of their opening account for the characteristic waveform of the action potential.

A second aspect of the organization of mechanisms is *spatial*. Different stages of the mechanism may be *compartmentalized* within some boundary or otherwise *localized* within some more or less well-defined region. These stages are *connected* with one another by, for example, motion and contact. Often, the connection between stages depends crucially upon the *structures* of the entities in the mechanism and upon those structured entities being *oriented* with respect to one another in particular ways.

These temporal and spatial aspects of the organization of mechanisms trace the productive relationships among the component stages—the relationship of one stage giving rise to, driving, making, or

allowing its successor. Importantly, mechanisms exhibit a productive continuity, without gaps, from setup to termination. Mechanisms require productive continuity to work, and accordingly our understanding of mechanisms turns on our ability to establish a seamless continuity between setup and termination. The discovery of mechanisms is often driven by the goal of eliminating gaps in this productive continuity.

Consider an example from contemporary neurobiology, the mechanism of long-term potentiation (LTP). LTP is a means of strengthening synapses, and many think that LTP, or something like it, is a crucial activity in the mechanism of spatial memory. The idea is essentially Hebb's (1949): when the presynaptic and the postsynaptic neuron are simultaneously active, the synapse is strengthened. LTP is commonly studied in the synapses of the mammalian hippocampus, a medial temporal lobe structure that is thought to be an important entity in the mechanism of spatial memory. A cross section of the hippocampus highlighting some of its major anatomical regions is shown in figure 6.2. Spatial memories are thought to be formed through the changing strengths of synapses between neurons in the hippocampus, and this is how LTP is thought to fit into the context of the mechanism of spatial memory.

The mechanisms of LTP are not yet completely understood. Robert Malinow, an LTP researcher, has worried in print that the LTC (long-term controversy) over LTP is becoming an LTTP (long-term tar pit) for neurobiologists (Malinow 1998, 1226). Nonetheless, one popular sketch of the mechanisms of LTP, visually represented in figure 6.3, includes the following organized collection of entities and activities. When the presynaptic neuron is active, it releases glutamate. This glutamate binds to *N*-methyl-D-aspartate (NMDA) receptors on the postsynaptic cell. The NMDA receptors change their conformation, exposing a pore in the cell membrane. If the postsynaptic cell is inactive, the channel remains blocked by large Mg^{2+} ions. But if the postsynaptic cell is depolarized, these Mg^{2+} ions float out of the channel, allowing Ca^{2+} to diffuse into the cell. The rising intracellular Ca^{2+} concentration sets in motion a long chain of biochemical activities terminating in the question marks of figure 6.3.

A number of gaps arise in the story at this point, but three things are thought to happen. In the short term, it is thought that this cascade leads to an increase in the number or sensitivity of α -amino-3-

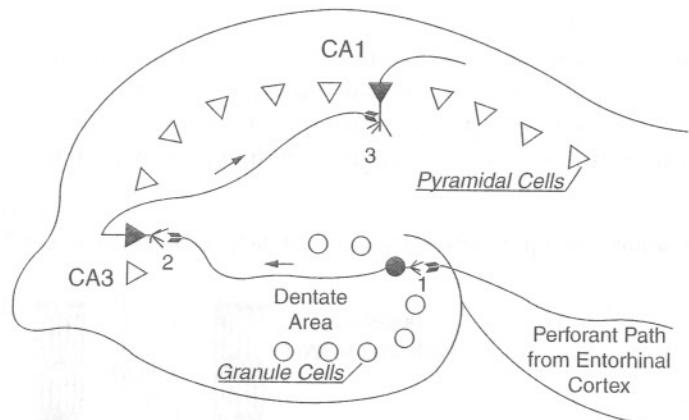


Figure 6.2 Cross section of the hippocampus. The CA1, CA3, and dentate regions are labeled. Numbers 1–3 denote the three synapses in the hippocampal trisynaptic loop. Circles and triangles represent granule cells and pyramidal cells, respectively.

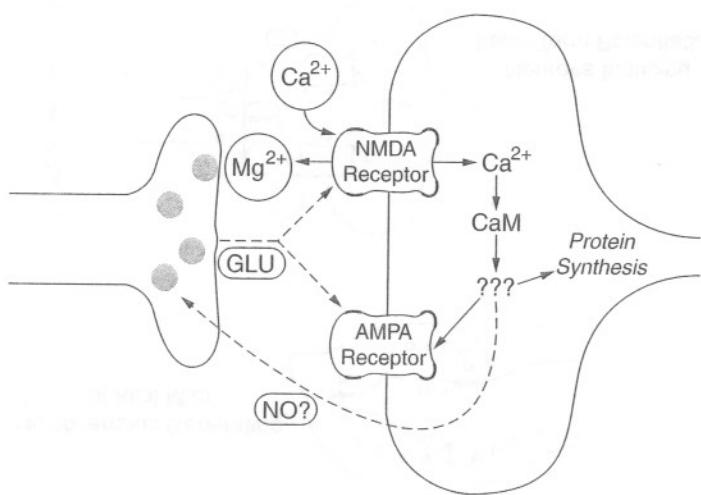


Figure 6.3 Sketch of a possible mechanism for the phenomenon of LTP.

hydroxyl-5-methyl-4-isoxazolepropionic acid (AMPA) receptors (perhaps by phosphorylation). These changes account for the rapid induction of LTP. In the long term, the cascade leads to the production of proteins in the postsynaptic cell body. These proteins are then used to alter the structure of the dendritic spines at that synapse. Some suspect

that there is also a presynaptic component of the LTP mechanism, whereby, for example, the presynaptic cell releases more glutamate. Although incomplete, this description will suffice as a sketch of the mechanism of LTP. (More detail concerning this mechanism and the evidence that supports it can be found in Kuno 1995 and Frey and Morris 1998.)

The entities in this mechanism are glutamate molecules, NMDA receptors, Ca^{2+} ions, and the like. The activities include binding, diffusing, phosphorylating, and changing conformation. These entities and activities are organized in the production of LTP. The components exhibit a temporal organization that begins with the release of glutamate and terminates in structural changes that strengthen the synapse. The rates and durations of the different stages are crucial for the working of the mechanism; for example, there are the short-term modification of the AMPA receptors and the long-term structural changes to the dendritic spine. Stages of the mechanism are compartmentalized or localized in cells, membranes, and pores. Ignoring the question marks, the early stages of the mechanism are connected with one another, mostly through the motion, binding, and breaking of molecules. These molecular activities depend crucially on the structures and orientations of the entities involved; the size of the pore and the complementary shapes of glutamate and the NMDA receptor allow these entities to engage in the activities that produce the later stages of the mechanism.

Levels

There is often a third aspect to the organization of mechanisms in addition to the spatial and temporal; this is a hierarchical aspect.¹ Mechanisms in contemporary neurobiology are organized into multi-level hierarchies (figure 6.4). The mechanism of spatial memory is a good illustration. The description of this mechanism includes mice learning to navigate, hippocampi generating spatial maps, synapses inducing LTP, and macromolecules (like the NMDA receptor) binding and bending.

The levels in this sort of hierarchy stand in part-whole relations to one another with the important additional restriction that the lower-level entities and activities are components of the higher-level mechanism. The binding of glutamate to the NMDA receptor is a lower-level activity in the mechanism of LTP, and LTP is thought to be a lower-

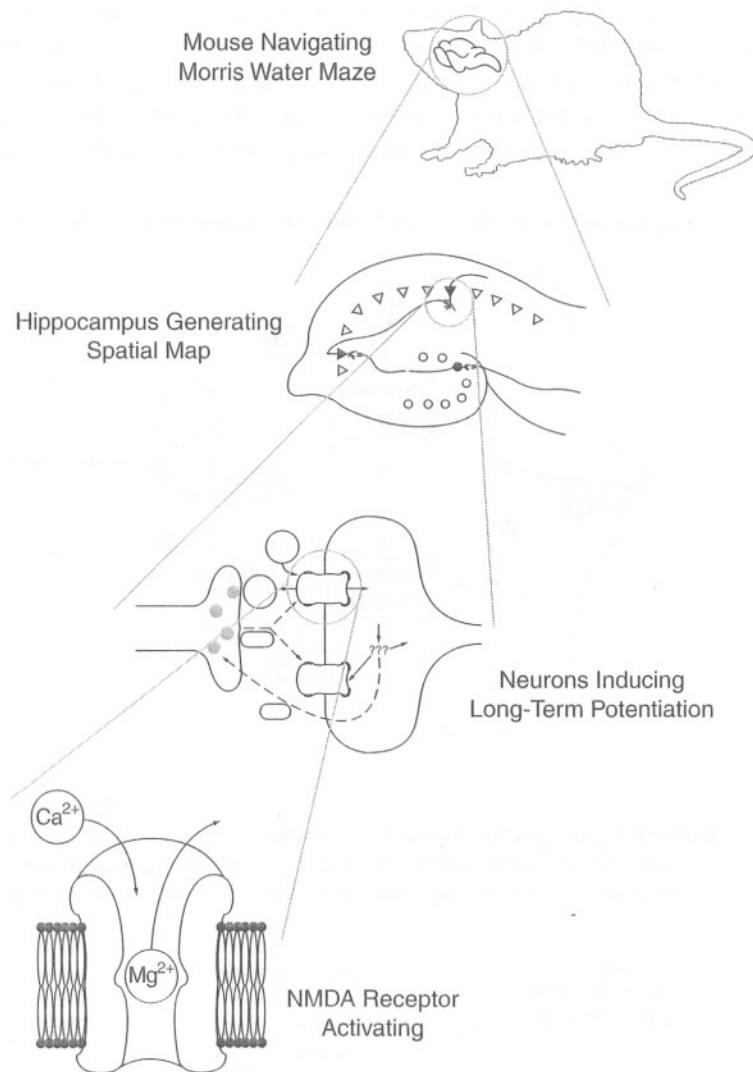


Figure 6.4 Levels in the hierarchical organization of the mechanism of spatial memory.

level activity in spatial map formation, which, of course, is thought to be an activity in the mechanism of spatial memory.

The elaboration and refinement of these hierarchical descriptions typically proceeds piecemeal with the goal of *integrating* the entities and activities at different levels. Integrating a component of a mecha-

nism into such a hierarchy involves, first, contextualizing the item within the mechanism of the phenomenon to be explained. This involves “looking up” a level and finding a functional role for the item in that higher-level mechanism. There is some question, for example, over the role of LTP in the mechanism of spatial memory; not only is there a debate over the correct role for LTP in this mechanism, but there is even some debate as to whether it has a role at all. Contextualizing an item within a relevant mechanism for the phenomenon to be explained is one step in the integration of the different levels in a hierarchy.

The second means of integrating a component into such a hierarchical description is downward looking. This involves showing that the properties or activities of an entity can be explicated in terms of a lower-level mechanism. The persistent failure to find a mechanism for a postulated property or activity signals that there is an irremediable gap in the productive continuity of the mechanism. The activity or property cannot be integrated with those at lower levels. Thus integrating multilevel mechanisms involves both contextualizing an item within higher-level mechanisms and explicating that item in terms of lower-level entities and activities. (For more on the relationship between functions, mechanisms, and levels, see Craver 1998, 2001.)

Describing Mechanisms

The preceding discussion of mechanisms and their organization is motivated by the idea that thinking carefully about the abstract structure of mechanisms can provide insight into how they are discovered. Philosophical discussions of discovery should be sensitive to the fact that there are many different kinds of things to be discovered and that these different kinds of things are not all discovered in the same way. The product shapes the process of discovery.

The mechanism for a given phenomenon, for example, is typically not discovered all at once. Instead, descriptions of mechanisms are typically constructed piecemeal. Often neurobiologists understand some stages of the mechanism quite well and have only the sketchiest understanding of other stages. The question marks in the LTP diagram in figure 6.3 make these gaps in the mechanism explicit. The descriptions of the various components of a mechanism are often evaluated, and so revised, independently of others. In order to capture this feature

of the discovery of mechanisms, it is useful to distinguish mechanism schemata from mechanism sketches.

Mechanism schemata are abstract descriptions of mechanisms that can be instantiated to yield descriptions of particular mechanisms. The term *mechanism schemata* is fitting because their components are placeholders that can be filled in with detailed stages between the setup and termination.² Schemata are thus complete in the sense that they can be filled in without gaps in the productive continuity of the mechanism. Schemata and their component placeholders typically have less than universal scope, and their scope can vary considerably.³

Mechanism sketches, in contrast to mechanism schemata, are abstract descriptions of mechanisms that cannot yet be filled in. Mechanism sketches have black boxes—they leave gaps in the productive continuity of the mechanism, such as the question marks in the LTP diagram. Such black boxes in mechanism sketches are useful in providing guidance about where further elaboration is needed. This role is especially important in the discovery of multilevel mechanisms.

The discovery of mechanisms unfolds gradually and piecemeal through the addition of constraints on plausible mechanism schemata and sketches. These constraints are used to construct plausible descriptions of mechanisms and to revise these plausible descriptions as constraints are added, deleted, or modified. Constraints determine the shape of the space of hypothesized mechanisms. Most simplistically, this space can be understood as a tree with terminal nodes representing possible mechanism schemata for the phenomenon to be explained. The addition of constraints prunes the tree or changes the weights on different branches. The removal of constraints, likewise, can add new branches to the hypothesis space. Understanding the discovery of mechanisms requires an understanding of these different constraints. This is the subject of the remainder of this chapter.

Constraints on the Organization of Mechanisms

Bechtel and Richardson's 1993 book *Discovering Complexity* discusses the use of localization and decomposition as research strategies in the discovery of mechanistic explanations and the conditions under which these strategies are prone to fail. Their book is an important contribution to the relatively sparse discovery literature in the philosophy of biology (see, e.g., Darden 1991; Darden and Cook 1994). Yet

the contribution remains incomplete without a careful look at the products of this discovery process. Thinking carefully about mechanisms and especially their organization highlights a broad variety of constraints on their discovery in addition to those that come from localizing and decomposing. (Bechtel and Richardson's discussion of constraints on "causal and explanatory models" is in some ways complementary to our treatment; see 1993, 235.) For the remainder of the chapter, we focus on five varieties of constraint. These include the character of the phenomenon, compendency constraints, spatial constraints, temporal constraints, and hierarchical constraints. Localization provides one kind of spatial constraint.

Characterizing the Phenomenon

Mechanism schemata and sketches are constrained by the character of the phenomenon for which a mechanism is sought. How one characterizes the phenomenon determines what will count as an adequate description of the mechanism that produces it; the complete description of the mechanism shows how that phenomenon is produced. Spatial memory is the phenomenon to be explained in our working example. But it is not at all obvious in advance of considerable empirical inquiry that there is any such individuable phenomenon—spatial memory—for which there exists an individuable mechanism; and given that there is such a faculty or phenomenon, it is not at all obvious in advance of considerable empirical inquiry how that phenomenon is properly to be characterized. Debates over the taxonomy of memory can be understood as debates about how to characterize and individuate different memory phenomena. The character of the phenomenon, like the description of the mechanism, is open to revision in light of evidence. This is a prevalent feature of the discovery of mechanisms.

Tolman's famous experiments on maze learning are a good example (Tolman and Honzick 1930; Tolman 1948). Tolman's work was instrumental in shaping the way in which contemporary neurobiologists think about spatial memory. Rats trained to navigate a circuitous route through a maze successfully were subsequently placed into the same maze with a new, more direct route from start to reward. If spatial memory were a simple association between stimulus and response, the rats would be expected to take the circuitous route for which they had been reinforced. But they did not; they preferred the more direct route. The rats could also construct efficient detours, shortcuts, and novel

routes to the reward (see, e.g., Olton and Samuelson 1976; Chapuis et al. 1987). Importantly, these experiments and others like them suggest that spatial memory involves the formation of an internal spatial representation—a cognitive map—by which different locations and directions in the environment can be assessed. This characterization of the phenomenon guides the neurobiologist to seek out some entity, property, or activity in the central nervous system that could serve as a representation of space.

The characterization of the phenomenon is also shaped crucially by the accepted experimental protocols for producing, manipulating, and detecting it. As Bogen and Woodward (1988) have argued, phenomena should not be confused with data, which are the evidence for phenomena. Data, among other things, are idiosyncratic to particular experimental arrangements; phenomena, as we think of them, are the stable and repeatable properties or activities that can be detected, produced, and manipulated in a variety of experimental arrangements. For our present purposes, it is important to note that different experimental arrangements reveal different aspects of the phenomenon.

So it was mazes of differing complexity that led Tolman to think of spatial memory in terms of the formation of spatial maps. Spatial memory is also tested in radial arm mazes, three-table problems, and, most importantly for our purposes, the Morris water maze. The Morris water maze is a circular pool filled with an opaque liquid that covers a hidden platform. Mice are trained to escape over repeated trials. They do not like to swim, and so they learn quickly. The aquatic nature of the task also eliminates smell as a sensory cue. So the maze isolates the place of *visual information* in spatial memory. Sherry and Healy (1998, 133) underscore the importance of different experimental protocols for understanding the phenomenon for which a mechanism is sought: “The idea of the cognitive map, first proposed by Tolman (1948), has been an important and influential stimulus to research. But it is really more of a metaphor than a theory. Research on path integration, landmark use, the sun compass and snapshot orientation . . . attempts to specify more concretely exactly what makes up a ‘cognitive map’ of space.” Different experimental assemblies accentuate different features of the phenomenon to be explained. Scientific debates often turn on the appropriateness of a given experimental arrangement for producing, manipulating, or detecting a given phenomenon. Debates over the ecological validity or ethological appropri-

ateness of a task, for example, are debates over the character of the phenomenon. Experimental arrangements are often revised and adjusted over the course of the discovery of a mechanism.

Characterizing the higher-level phenomenon to be explained is a vital step in the discovery of mechanisms. Characterizing the phenomenon prunes the hypothesis space (since the mechanism must produce the phenomenon) and loosely guides its construction (since certain phenomena are suggestive of possible mechanisms). Yet such a top-down approach, as Bechtel and Richardson (1993, 237) agree, cannot itself exhaust the discovery of a mechanism. One also must know the components of the mechanism and how they are organized.

Componency Constraints

Mechanisms, remember, are composed of both entities and activities. For a given field at a given time there is typically a store of established or accepted components out of which mechanisms can be constructed and a set of components that have been excluded from the shelves. The store also contains accepted modules: organized composites of the established entities and activities. In contemporary neurobiology, for example, brain mechanisms will be composed of discrete neurons rather than a “reticulum.” These neurons are connected by synapses, which may be electrical, chemical, or both. If they are chemical, then the mechanism will most likely involve action potentials, quantal release, and allosteric interactions. Modules in neurobiology include different second-messenger cascades, ionophore complexes, and cytoarchitectural structures, such as ocular dominance columns and glomeruli.

The store of entities, activities, and modules out of which mechanisms are constructed expands and contracts with the addition and removal of established entities and activities over time. Contracting the store adds constraints on plausible mechanisms by pruning those branches of the hypothesis space that represent mechanisms with such unestablished or unaccepted components. One commentator recently praised a hypothesized mechanism of LTP by saying, “If nothing else, this model is attractive because it requires only established intracellular signaling mechanisms” (Malinow 1998, 1226). Expanding the store of components loosens constraints by adding branches to the hypothesis space. The addition of nitric oxide (NO) to the store in the 1980s opened the hypothesis space to mechanisms involving retro-

grade transmission from the postsynaptic to the presynaptic neuron. It had previously been assumed that chemical neurotransmission was unidirectional, but this entity, which can diffuse freely through neuronal membranes, expanded the space of possibilities (figure 6.3). Importantly, the store of mechanism components provides guidance in the construction of mechanisms by supplying a set of ingredients out of which mechanisms might be concocted. Introductory neurobiology textbooks acquaint students with this store of entities, activities, and modules.

These textbooks also introduce students to various limitations on the activities in which these entities can engage. These features are important compency constraints on plausible mechanisms. For example, action potentials do not travel at the speed of light. The fastest move at 120 meters per second, and that is in the squid, with its appropriately named giant axons. Human action potentials propagate at roughly 1 meter per second, and our neurons can only fire 500 times in a second. Constraints can also be found in metabolic requirements, computational resources, temperature limits, rates of protein synthesis, and similar facts of carbon-based life on earth. A pair of authors recently dismissed the hypothesis that the proteins for altering the structure of the synapse in LTP were synthesized within each individual synaptic spine. They rejected this hypothesized mechanism as metabolically too demanding (Frey and Morris 1998, 182; they also discuss experimental evidence against the hypothesis). Although these compency constraints are always open to revision in the light of new evidence, they can be decisive in determining the fate of a proposed mechanism schema.

Spatial Constraints

Compency constraints shape the hypothesis space by delimiting the store of entities, activities, and modules that can be included in a mechanism and by limiting the possible activities in which the entities can engage. More specific constraints arise from empirical discoveries concerning the spatial organization of the mechanism. The components of mechanisms are often compartmentalized, localized, connected, structured, and oriented with respect to one another. Evidence concerning these sorts of spatial relationships among the components of a mechanism also constrains and guides the discovery process.

Often the components or stages of mechanisms are *compartmentalized* within reasonably well-defined regions. As the term suggests, these regions are often sectioned off by physical boundaries, like a nuclear membrane, a cell membrane, or skin. Compartmentalization often provides a natural way to individuate the stages of a mechanism. Transcription happens in the nucleus and translation happens in the cytoplasm; there are pre- and postsynaptic components of the mechanism of LTP. Compartmentalization also guides one to seek activities capable of linking the components inside this boundary with those outside—activities such as diffusion, active transport, and second messenger systems.

Closely related to compartmentalization is *localization*. Localizing components, the major focus of Bechtel and Richardson (1993), is often essential for understanding the spatial layout of a mechanism. For reasons to be discussed shortly, researchers are now reasonably confident that some components of the mechanism of spatial memory are to be found in the hippocampus. This finding opens up two new sets of research questions grounded in the spatial organization of the mechanism. One can, for instance, look inside the hippocampus to see what makes it work. Such investigation allows one to restrict the store of components to just those that can be identified within this spatial region. One is constrained to understand the activity of the hippocampus in terms of the cells, synapses, neurotransmitters, and circuits that can be found there.

One can then begin to describe the connections of these hippocampal components. *Connectivity* is yet a third variety of spatial constraint; the productive continuity of mechanisms relies upon the spatial connections among the components. Early research on the hippocampus has operated under the assumption that the anatomical connectivity of the hippocampal regions exhibits a characteristic clockwise “trisynaptic” loop (figure 6.2). Perforant path fibers from the entorhinal cortex make the first synapse onto granule cells in the dentate gyrus. These in turn project their axons to the pyramidal cells of the CA3 region, which in turn project to CA1 pyramidal cells. Revising this simple wiring diagram by adding new connections or new types of cells alters the space of plausible mechanisms by changing the scaffolding upon which the mechanism can be constructed. More recent research is beginning to emphasize the recurrent connections within these different regions.

Knowing that part of the spatial memory mechanism is localized to the hippocampus also constrains the description of the mechanistic context of the hippocampus. The anatomical connections into and out of the hippocampus become further constraints on the mechanism of spatial memory. This connected anatomy is the spatial scaffolding of the components of the mechanism. Localization is thus an important tool for revealing the connectivity of the mechanism.

Details concerning the geometrical *structure* and *orientation* of the entities in a mechanism, on the other hand, are important for understanding the productivity of mechanisms. As we noted earlier, the stages of mechanisms often depend crucially upon entities with appropriate structures having appropriate orientations with respect to one another. Discovering these structures and orientations can place important constraints on the hypothesis space. Indeed, structural aspects of the LTP mechanism are a major focus of recent research on LTP. Recent articles (Engert and Bonhoeffer 1999; Maletic-Savatic et al. 1999) present evidence for the addition of new dendritic spines to recently potentiated synapses. Although far from conclusive, such evidence argues for the existence of the structural basis for one plausible mechanism sketch for LTP.⁴ The idea, yet to be confirmed, is that the addition of new dendritic spines makes the postsynaptic cell more responsive to glutamate. Structure thus provides clues to the activities that sustain the productive continuity of the mechanism. What remains to be shown, if this mechanism sketch is to be viable, is that the new postsynaptic spines are oriented properly with respect to the presynaptic axons.

The general point of this discussion is that mechanisms are organized spatially in the production of a phenomenon. Identifying aspects of that spatial organization guides and constrains the search through the hypothesis space in a number of ways that go beyond the strategy of localizing. Locations, boundaries, connections, positions, shapes, and orientations are especially important characteristics of, or relations among, the entities in mechanisms; these characteristics are especially important because they constrain the activities in which those entities can engage and so constrain the way that the mechanism can work. It is for this reason that characterizing these spatial aspects of the organization of mechanisms contributes to our understanding. But details of the spatial organization alone do not allow one to under-

stand what mechanisms do. This spatial organization must be set in motion.

Temporal Constraints

Our efforts to understand just how a given mechanism moves are delimited and guided by knowledge of the mechanism's temporal organization. Knowledge of the order, rate, duration, and frequency of the activities in which the component entities of the mechanism engage provides clues to how the mechanism works.

Consider the temporal order of the activities of the entities composing a mechanism—that is, their relative position in the series, forks, and cycles that make up the mechanism. Spatial organization, in and of itself, does not reveal the direction of the productivity in the mechanism—the idea that the circuit of neurons in the hippocampus goes clockwise or that NMDA receptors allow Ca^{2+} influx into the postsynaptic cell, thereby initiating protein synthesis. Of course, temporal sequence is not by itself sufficient to establish these productive relationships, but given the temporal asymmetry of causality, temporal relations can place constraints on which entities and activities can be seen as productive of which others.

Temporal constraints on the discovery of mechanisms also include constraints imposed by the rates and durations of both the phenomenon to be explained and the stages of the mechanism. The speed limit for generating and propagating action potentials and for transmission at chemical synapses, for example, places limits on the number of sequential steps involved in a phenomenon of a given duration. Temporal constraints have been important in the discovery of the mechanisms of LTP. Researchers who believe that enduring LTP may be sustained by the addition of dendritic spines to the postsynaptic cell, for example, cannot use this mechanism to explain the initial induction of LTP because it takes around 30 minutes to produce the required proteins, distribute them, and insert them into the membrane. Short-term induction of LTP must rely on some faster mechanism, like the phosphorylation of AMPA receptors. Possible mechanisms are pruned from the hypothesis space on the grounds that the stages or steps take too long or happen too slowly to produce a phenomenon with a given rate or duration. Because mechanisms are active—because mechanisms do things—they take time to work, and the order, rate, and

duration of the stages in a mechanism are therefore important tools for discovering their organization and culling the hypothesis space.

Experiments for Testing Hierarchical Mechanisms

The final set of constraints on the discovery of mechanisms is grounded in the hierarchical organization of mechanisms. Neurobiologists conduct experiments to reveal this hierarchical organization. Although there is a great deal to be said about how experiments are used to test mechanisms, the focus here is upon the role of experiments in the integration of levels in multilevel schemata and sketches.

Experiments can be understood in terms of an abstract experimental protocol, which clearly has some affinities with Hacking's (1988, 1992) discussion of experimentation.⁵ Hacking does not discuss experimentation in the discovery of mechanisms per se, nor does he address their role in the integration of levels. We can achieve a more detailed understanding of experimentation by attending to the mechanistic organization that these experiments are used to reveal.

The abstract protocol in figure 6.5 is most easily described in the case of a unilevel mechanism; the protocol can then be easily extended to a multilevel case. The connected circles and arrows represent a hypothesized mechanism schema putatively instantiated in some experimental preparation. The schema may be instantiated *in vivo*, *in vitro*, or *in silico* (that is, in a mouse, in a petri dish, or in a computer). Having found or created such an experimental preparation, one then typically uses some *intervention technique* to perturb some component. The perturbation produced in the experimental preparation presumably has downstream results that are detected or amplified with the help of a *detection technique*. There is much to be said about this idealized protocol; we introduce it here simply to extend it to the multilevel case. And this is easy to do. Experiments designed to test the hierarchical organization of a mechanism typically involve intervening at one level and detecting at another. Sometimes, a single set of experiments involves intervening and detecting at multiple levels at once; we shall get to such a case shortly.

For simplicity, though, we start with two-level experiments. The left-hand side of figure 6.6 exhibits a case of intervening at the lower level and detecting at the higher level; we call these "bottom-up" experiments. The right-hand side of that figure shows the opposite:

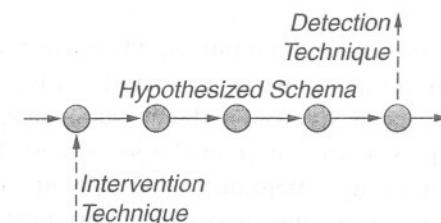


Figure 6.5 Abstract experimental protocol. Horizontal circles and arrows represent a hypothesized mechanism schema putatively instantiated in an experimental preparation. Vertical arrows represent techniques for intervening (left) to perturb the mechanism and for detecting (right) the results of that intervention.

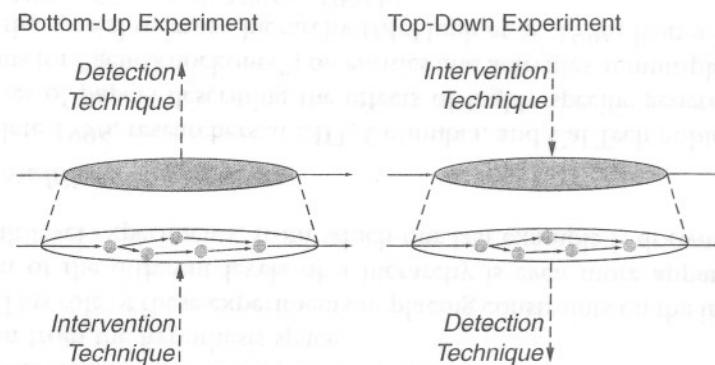


Figure 6.6 Multilevel experiments.

intervention at the higher level and detection at the lower level. These can be thought of as "top-down" experiments. Interventions, in either of these cases, may be either stimulatory or inhibitory. Our first examples are inhibitory, bottom-up experiments. Our second set of examples is top-down and stimulatory. Both sets of examples are drawn from experiments that, taken together, forcefully suggest that the hippocampus is involved in the formation of spatial maps.

Bottom-Up Inhibitory Experiments

The first example is the now-familiar case study H.M., as reported by Scoville and Millner (1957). Because H.M. was plagued by incapacitating epileptic seizures, he consented to an experimental surgical procedure to remove portions of his medial temporal lobes, including the hippocampus. After the surgery, it quickly became apparent that H.M.

had lost the ability to remember recent facts, even though he retained the ability to learn new skills. H.M.'s case famously suggested to researchers that the human hippocampus is a crucial entity in the mechanisms of what has since been called "declarative" memory (Zola-Morgan and Squire 1993).

Subsequent experiments in rats and mice have shown that bilateral removal of the hippocampus leads to profound deficits in spatial memory. For example, although rats with intact hippocampi learn very quickly, over repeated trials, to swim directly to the hidden platform in the Morris water maze, rats with bilateral hippocampal lesions continue over multiple trials to swim randomly through the pool, stopping only when they stumble onto the platform (Morris et al. 1982).

Both the case of H.M. and the subsequent ablation experiments in mice are examples of the two-tiered experimental structure exhibited in the left-hand side of this figure.

Top-Down Excitatory Experiments

The findings of these inhibitory, bottom-up experiments are reinforced by excitatory, top-down experiments like those on the right-hand side of the figure. In the early 1970s, O'Keefe and Dostrovsky (1971) recorded the electrical potentials of individual pyramidal cells in the CA1 region of the rat hippocampus while rats navigated a standard maze. The intervention in this case involves activating the spatial memory system by putting the rat in a maze. The detection technique is the electrical recording. They found that certain of those pyramidal cells generate bursts of action potentials whenever the rat enters a particular location while facing in a particular direction. These cells have come to be called "place cells," and the region of space occupied by the rat when the place cell increases its activity is likewise known as the cell's "place field." The place cells of CA1 have slightly overlapping place fields that cover the animal's immediate spatial environment. The pattern of activity across this subpopulation of CA1 pyramidal cells could therefore play the role of a spatial map.

These findings have recently been confirmed with multiunit electrodes that allow one to record from 70 to 150 CA1 pyramidal cells at once. Astonishingly, it is possible to *predict* the path taken by the rat on the basis of these recordings (Wilson and McNaughton 1993). This is a remarkable top-down stimulatory finding.

Top-down and bottom-up experiments of both the stimulatory and inhibitory variety are quite common in neurobiology. They are common because the findings of such experiments, taken together, reveal aspects of, and thereby place constraints upon, the hierarchical organization of a mechanism. More specifically, when experiments of this sort go well, they place constraints on the possibilities for integrating the different levels. Top-down and bottom-up experiments help to situate an item, like the hippocampus or the NMDA receptor, within the context of a higher-level mechanism. They also identify components in the mechanisms that produce higher-level activities and properties. These experiments tell us what the relevant entities and activities are, how they are nested in component-subcomponent relations, and how the activities of the component entities fit into their mechanistic context. Persistent failure to situate an item within a hierarchical mechanism, or persistent failure to uncover a lower-level mechanism for that item, prunes mechanism schemata involving that item from the hypothesis space.

This role of these experiments in placing constraints on the integration of the different levels of a hierarchy is even more apparent in multilevel experiments, from which our last example is drawn.

Multilevel Experiments

In late 1996, researchers at MIT, Columbia, and Cal Tech published a series of papers describing the effects of highly specific genetic deletions (or "gene knockouts") on entities and activities at multiple levels in the spatial memory hierarchy (McHugh et al. 1996; Rottenberg et al. 1996; Tsien et al. 1996a, 1996b).

The experiments that are our focus are bottom-up and inhibitory. Specifically the researchers invented a molecular scalpel for deleting the *NMDAR1* gene, a gene encoding an essential subunit of the NMDA receptor (Tsien et al. 1996a), and for deleting it only in the pyramidal cells of the CA1 region of the mouse hippocampus. The deletion was also timed to occur only after normal hippocampal development is thought to be complete. The trick was to couple the deletion to a promoter of a gene that is expressed selectively in CA1 pyramidal cells and that is expressed only in the later developmental stages of the hippocampus. This intervention technique gives researchers finer-grained spatial and temporal resolution in their manipulation of the brain's activities than has ever been possible. This, in

turn, provides higher spatio-temporal resolution on the organization of the mechanism of spatial memory.

These experiments have been praised as the first to investigate the mechanism of spatial memory, “at all levels in a single set of experiments, from molecular changes through altered patterns of neuronal firing to impaired learning” (Roush 1997, 32) and for taking an important first step toward the “dream of neurobiology . . . to understand all aspects of interesting and important cognitive phenomena—like memory—from the underlying molecular mechanisms through behavior” (Stevens 1996, 1147). More specifically, we claim that these experiments advance the goal of integrating the different levels in this multilevel mechanism.

Knockout mice, those without functional NMDA receptors, had difficulty escaping the Morris water maze. They performed far worse than controls in learning to escape. When placed in a maze *without* a platform, control mice concentrated their swimming in the platform’s previous location. Knockout mice swam about randomly (Tsien et al. 1996b). Multiunit recordings from CA1 pyramidal cells in the knockout mice revealed significant impairments in spatial map formation. The researchers found, to their surprise, that CA1 cells in the knockout mice *did* exhibit place-related firing. But the place fields were much larger and much less sharply defined. These deficits in spatial map formation can reasonably be attributed to the absence of LTP at synapses lacking functional NMDA receptors. The researchers found that knocking out the NMDA receptor eliminated LTP induction in CA1 and not in any other region of the brain (Tsien et al. 1996b).

This complicated experiment is a bottom-up inhibitory experiment with detection at multiple levels. The intervention technique intervenes to perturb the NMDA receptor by deleting the *NMDAR1* gene. The detection techniques register the results of this intervention on LTP, spatial map formation, and spatial memory. There is a lot to be said about the strength of these experimental findings, but this is not our focus here. Instead we are interested in how these multilevel experiments constrain hypotheses about the integration of multilevel mechanisms.

This set of experiments is designed to test a popular sketch of the multilevel mechanism of spatial memory. It is a sketch because we are not remotely in a position to trace out all of the mechanisms at all of the different levels. Instead, this sketch is a hypothesis of how the components at different levels are integrated with one another. In

particular, it is the hypothesis that the NMDA receptor is a necessary component of the mechanism of LTP, which is a necessary component of the mechanism of spatial map formation, which is a necessary component of the mechanism of spatial memory. If these nesting relationships do hold, then knocking out an essential gene for the NMDA receptor would be expected to eliminate the induction of LTP, to eliminate spatial map formation, and to leave the mice hopelessly lost in the Morris water maze. The rough conformity of the findings to these expectations is heartening in this respect.

It is important to note, however, that the genetic deletion did not eliminate spatial map formation in the CA1 region of the hippocampus. Instead, knocking out this essential gene made the map less precise. This lack of precision still had behavioral implications, and so it is consistent with some role for LTP within the context of the spatial memory mechanism as a whole. However, this anomalous finding forced the researchers to rethink the role of CA1 LTP in the context of this mechanism. The persistence of place-related firing in CA1 suggests that place fields must be established in an earlier stage of the mechanism. So the role of CA1 within the context of the spatial memory mechanism is not the formation of spatial maps. Instead the researchers suggest, with characteristic caution, that CA1 has the role of “learn[ing] associations between entorhinal inputs and place [information projecting from the CA3 region of the hippocampus]” (McHugh et al. 1996, 1347). This finding, in other words, constrains our understanding of the role of CA1 LTP in the spatial memory mechanism; as this case of revision suggests, it also helps to establish the productive organization of the components of the mechanism as a whole. (For a more systematic discussion of anomaly resolution, see Darden 1991, 1992; Darden and Cook 1994.)

It is in this way that multilevel experiments furnish constraints on the integration of components at multiple levels. These hierarchical constraints, in conjunction with the character of the phenomenon and the spatial, temporal, and compency constraints, guide and delimit the discovery of mechanisms.

Conclusion

It is now possible to step back and look at this discovery process as a whole. The continuing discovery of the mechanism of spatial memory has proceeded gradually through the piecemeal elaboration and revi-

TABLE 6.1 Summary of Constraints on the Discovery of Mechanisms Exhibited in the Discovery of the Mechanism of Spatial Memory

1. Character of the phenomenon
2. Componency of constraints
Store of entities and activities
Modules
3. Spatial constraints
Compartmentalization
Localization
Connectivity
Structure
Orientation
4. Temporal constraints
Order
Rate
Duration
Frequency
5. Hierarchical constraints
Integration of levels

sion of mechanisms at multiple levels and through the gradual integration of the components at each of these levels. This discovery process has been guided by a sketch, replete with black boxes, of how this mechanism is organized. Certain of these black boxes are starting to open with the accumulation of constraints on how they are to be filled in. We are beginning to understand the details of LTP, and perhaps more importantly, we are beginning to evaluate more precisely the role of LTP within the context of the spatial memory mechanism. The same is true for spatial map formation in the hippocampus and the opening and closing of the NMDA receptor.

This discovery process is guided by constraints on the organization of the mechanism that are summarized in table 6.1. These constraints come from many different specialties within neurobiology: neuroanatomists, clinical psychologists, electrophysiologists, and molecular neurobiologists are all contributing different constraints on the emerging organization of this hierarchical mechanistic structure. Discussions of scientific discovery in neurobiology should proceed by attending to the organizational structure of mechanisms. The product shapes the process of discovery. In understanding both the product and the process we come to see more clearly what is involved in understanding phenomena by describing mechanisms.

NOTES

This chapter is based on work supported by the National Science Foundation under grant SBR-9817942. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation. Craver's work was also partly funded by the Committee on Cognitive Studies Postdoctoral Fellowship in the Department of Philosophy at the University of Maryland, College Park. We thank German Barrionuevo, Nancy Hall, Tetsuji Iseda, Peter Machamer, Gualtiero Piccinini, Rob Skipper, and Nathan Urban for useful comments on earlier drafts of this chapter and Payman Farsaii for help as an undergraduate research assistant.

1. Although the notion of "hierarchy" is often associated with the control or governance of things at lower levels by those at higher levels, this should in no way be associated with the notions of "level" and "hierarchy" explicated here.
2. Skipper (1999) develops a mechanism schema for selection mechanisms.
3. Schaffner (1993) has suggested that the "bulk" of theories in the biomedical sciences be seen as "overlapping interlevel temporal models of varying scope." Although we are sympathetic with the direction of Schaffner's thinking here, we prefer to think of such theories as schematic multilevel descriptions of mechanisms. Our discussion of spatial, temporal, and hierarchical constraints on descriptions of mechanisms is intended to exhibit the additional content of an explicit emphasis on mechanisms over and above less specific talk of "theories" or "models." Both mechanism schemata and their components (the placeholders for entities and activities) can have widely varying scope, from near-terrestrial universality (such as the mechanism of protein synthesis) to mechanisms, entities, or activities that are found only in some parts of some highly specific strains of organisms. (More on scope can be found in Darden's [1996] review of Schaffner [1993].) Because the sense of "level" articulated in the second section is explicitly defined in terms of "componency" and hence "mechanism," emphasis on mechanisms in the structure of these theories also brings with it a sensible interpretation of their multilevel character. For discussions of multilevel mechanism schemata as theories in neurobiology, see Craver (1998, 10–48; 2001).
4. Additional details concerning this hypothesized mechanism continue to emerge. See Barinaga (1999) and Shi et al. (1999) on the delivery and activation of AMPA receptors in LTP.
5. For an alternative schematic account of experiments, see Lederberg (1995).

REFERENCES

- Barinaga, M. 1999. "New Clues to How Neurons Strengthen Their Connections." *Science* 284: 1755–57.
- Bechtel, W., and R. C. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, N.J.: Princeton University Press.
- Bogen, J., and J. Woodward. 1998. "Saving the Phenomena." *Philosophical Review* 97: 303–52.

- Chapuis, N., M. Durup, and C. Thinus-Blanc. 1987. "The Role of Exploratory Experience in a Shortcut in Golden Hamsters (*Mesocricetus auratus*)."*Animal Learning and Behavior* 15: 174–78.
- Craver, C. F. 1998. "Neural Mechanisms: On the Structure, Function, and Development of Theories in Neurobiology." Doctoral dissertation, Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, Pennsylvania.
- . 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science*, March.
- Darden, L. 1991. *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.
- . 1992. "Strategies for Anomaly Resolution." In R. Giere, ed., *Cognitive Models of Science*. Minnesota Studies in the Philosophy of Science, vol. 15. Minneapolis: University of Minnesota Press, 251–73.
- . 1996. "Generalizations in Biology." *Studies in History and Philosophy of Science* 28: 409–19.
- Darden, L., and M. Cook. 1994. "Reasoning Strategies in Molecular Biology: Abstractions, Scans, and Anomalies." In D. Hull, M. Forbes, and R. M. Burian, eds., *PSA 1994*, vol. 2. East Lansing, Mich.: Philosophy of Science Association, 179–91.
- Engert, F., and T. Bonhoeffer. 1999. "Dendritic Spine Changes Associated with Hippocampal Long-Term Synaptic Plasticity." *Nature* 399: 66–70.
- Frey, U., and R. G. Morris. 1998. "Synaptic Tagging: Implications for Late Maintenance of Hippocampal Long-Term Potentiation." *Trends in Neuroscience* 21: 181–88.
- Hacking, I. 1988. "On the Stability of the Laboratory Sciences." *Journal of Philosophy* 85: 507–14.
- . 1992. "The Self-Vindication of the Laboratory Sciences." In A. Pickering, ed., *Science as Practice and Culture*. Chicago: University of Chicago Press, 29–64.
- Hebb, D. O. 1949. *The Organization of Behavior*. New York: Wiley.
- Kuno, M. 1995. *The Synapse: Function, Plasticity, and Neurotrophism*. Oxford: Oxford University Press.
- Lederberg, J. S. 1995. "Notes on Systematic Hypothesis Generation and Application to Disciplined Brainstorming." In *Working Notes: Symposium: Systematic Methods of Scientific Discovery*. AAAI Spring Symposium Series. American Association for Artificial Intelligence, Stanford University, 97–98.
- Machamer, P., L. Darden, and C. F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67: 1–25.
- Maletic-Savatic, M., R. Malinow, and K. Svoboda. 1999. "Rapid Dendritic Morphogenesis in CA1 Hippocampal Dendrites Induced by Synaptic Activity." *Science* 283: 1923–26.
- Malinow, R. 1998. "Silencing the Controversy in LTP?" *Neuron* 21: 1226–27.
- McHugh, T. J., K. I. Blum, J. Z. Tsien, S. Tonegawa, and M. A. Wilson. 1996. "Impaired Hippocampal Representation of Space in CA1-Specific NMDAR1 Knockout Mice." *Cell* 87: 1339–49.

- Morris, R. G. M., P. Garrud, J. N. P. Rawlins, and J. O'Keefe. 1982. "Place Navigation Impaired in Rats with Hippocampal Lesions." *Nature* 297: 681–83.
- O'Keefe, J., and J. Dostrovsky. 1971. "The Hippocampus as a Spatial Map: Preliminary Evidence from Unit Activity in the Freely Moving Rat." *Brain Research* 34: 171–75.
- Olton, D. S., and R. J. Samuelson. 1976. "Remembrances of Places Passed: Spatial Memory in Rats." *Journal of Experimental Psychology: Animal Behavior Processes* 2: 97–116.
- Rotenberg, A., M. Mayford, R. D. Hawkins, E. R. Kandel, and R. U. Muller. 1996. "Mice Expressing Activated CaMKII Lack Low Frequency LTP and Do Not Form Stable Place Cells in the CA1 Region of the Hippocampus." *Cell* 87: 1351–61.
- Roush, W. 1997. "New Knockout Mice Point to Molecular Basis of Memory." *Science* 275: 32–33.
- Schaffner, K. 1993. *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.
- Scoville, W. B., and B. Millner. 1957. "Loss of Recent Memory after Bilateral Hippocampal Lesions." *Journal of Neurology, Neurosurgery, and Psychiatry* 20: 11–20.
- Sherry, D., and S. Healy. 1998. "Neural Mechanisms of Spatial Representation." In S. Healy, ed., *Spatial Representation in Animals*. Oxford: Oxford University Press, 133–57.
- Shi, S., Y. Hayashi, R. S. Petralia, S. H. Zaman, R. J. Wenthold, K. Svoboda, and R. Malinow. 1999. "Rapid Spine Delivery and Redistribution of AMPA Receptors after Synaptic NMDA Receptor Activation." *Science* 284: 1811–16.
- Skipper, R. 1999. "Selection and the Extent of Explanatory Unification." *Philosophy of Science* 66: S196–S209.
- Stevens, C. F. 1996. "Spatial Learning and Memory: The Beginning of a Dream." *Cell* 87: 1147–48.
- Tolman, E. 1948. "Cognitive Maps in Rats and Men." *Psychological Review* 55: 189–208.
- Tolman, E., and C. Honzick. 1930. "Introduction and Removal of Reward and Maze Performance in Rats." *University of California Publications in Psychology* 4: 257–75.
- Tsien, J. Z., D. F. Chen, D. Gerber, C. Tom, E. Mercer, D. Anderson, M. Mayford, and E. R. Kandel. 1996a. "Subregion- and Cell Type-Restricted Gene Knockout in Mouse Brain." *Cell* 87: 1317–26.
- Tsien, J. Z., P. T. Huerta, and S. Tonegawa. 1996b. "The Essential Role of Hippocampal CA1 NMDA Receptor-Dependent Synaptic Plasticity in Spatial Memory." *Cell* 87: 1327–38.
- Wilson, M. A., and B. McNaughton. 1993. "Dynamics of the Hippocampal Ensemble Code for Space." *Science* 261: 1055–58.
- Zola-Morgan, S., and L. Squire. 1993. "Neuroanatomy of Memory." *Annual Review of Neuroscience* 16: 547–63.

