Available online at www.sciencedirect.com
SCIENCE DIRECT®
Stud. Hist. Phil. Biol. & Biomed. Sci. 36 (2005) 373–395
Studies in History and Philosophy of Biological and Biomedical Sciences
www.elsevier.com/locate/shpsc

Beyond reduction: mechanisms, multifield integration and the unity of neuroscience

Carl F. Craver

Department of Philosophy, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA

Abstract

Philosophers of neuroscience have traditionally described interfield integration using reduction models. Such models describe formal inferential relations between theories at different levels. I argue against reduction and for a mechanistic model of interfield integration. According to the mechanistic model, different fields integrate their research by adding constraints on a multilevel description of a mechanism. Mechanistic integration may occur at a given level or in the effort to build a theory that oscillates among several levels. I develop this alternative model using a putative exemplar of reduction in contemporary neuroscience: the relationship between the psychological phenomena of learning and memory and the electrophysiological phenomenon known as Long-Term Potentiation. A new look at this historical episode reveals the relative virtues of the mechanistic model over reduction as an account of interfield integration.

Keywords: Mechanism; Reduction; Unity of science; Fields; Neuroscience; Memory; Long-term potentiation

E-mail address: ccraver@wustl.edu (C.F. Craver).

1369-8486/$ - see front matter
doi:10.1016/j.shpsc.2005.03.008

1. Introduction

Neuroscience is a multifield research program.[1] Its departments, journals, societies, and textbooks include perspectives from anatomy, biochemistry, computer science, developmental, evolutionary and molecular biology, electrophysiology, experimental psychology, ethology, pharmacology, psychiatry, and radiology to name just a few. The Society for Neuroscience (SfN) was founded in 1969 with the mission to:

> Advance the understanding of the brain and the nervous system by bringing together scientists of diverse backgrounds, by facilitating the integration of research directed at all levels of biological organization, and by encouraging translational research and the application of new scientific knowledge to develop improved disease treatments and cures.[2]

Understanding the structure of contemporary neuroscience requires understanding how these multiple fields, embodying distinct perspectives, techniques, and vocabularies, manage to integrate their work.

Most philosophers who have discussed interfield integration in neuroscience (e.g., Bickle, 1998, 2003; Churchland, 1986; Oppenheim and Putnam, 1958; Schaffner, 1993a,b) describe it using models of intertheoretic reduction. According to the 'classical' model of reduction, from which each of these authors' models descends, reduction is a species of deductive nomological explanation: one theory is reduced to another when it is possible to identify the theoretical terms of the first with those of the second and to literally derive the first from the second. On the assumption that fields and theories correspond, reduction then serves as a model of interfield integration as well.

There are many reasons why philosophers of neuroscience have found reduction attractive for thinking about interfield integration. First, the reduction relation can be defined precisely using formal logic (e.g., Schaffner, 1993a,b) or set theory (e.g., Bickle, 2003), and so the thesis that fields are integrated through reduction is clear and testable. Second, there is a long tradition of using reduction models in the philosophy of physics, chemistry, and biology, and it is natural to suggest that the models can be extended to the neurosciences. Finally, at least since Oppenheim and Putnam's manifesto (1958), reduction has been nearly synonymous with the explanatory unity of science: the unity of science is achieved by reducing the theories of all fields to the theories of the one field describing fundamental ontology. For these reasons, reduction seems natural as a model of interfield integration in the neurosciences.

[1] In the spirit of Darden & Maull (1977), I understand fields as groups of researchers related by common problems, techniques, and vocabularies. The boundaries between fields are fuzzy and change with time, but there is no pressing need to tidy them up for present purposes.

[2] Society for Neuroscience web page (2003) at http://web.sfn.org/content/AboutSfN1/Mission/mission.htm.

But reduction models have well rehearsed shortcomings. Most philosophers now accept that nonfundamental kinds are multiply realizable—that is, the same kind of phenomenon can be realized on different occasions by wildly different substances, parts, or organized processes. Arguments concerning multiple realization are largely responsible for the fact that non-reductive physicalism is now a standard view in the philosophy of mind (Putnam, 1975; Fodor, 1974, 1997). Furthermore, derivational models of explanation have received sustained attack for decades, and many now recognize the importance of non-derivational and especially causal forms of scientific explanation (e.g., Salmon, 1984, 1989). And finally, historians of science have identified numerous non-reductive forms of interfield integration including the development of interfield theories (Maull, 1977; Darden & Maull, 1977; Bechtel, 1986), interfield relations (Darden, 1991), and pidgin languages in the development of scientific techniques (Galison, 1987; see also Wylie, 2003). Reduction seems to survive as a model of interfield integration in neuroscience despite these shortcomings largely because there is no alternative of comparable scope and clarity.

The primary aim of this paper is to articulate such an alternative, to compare it to reduction, and to display some of its relative merits. My target is reduction as a model of interfield integration and the unity of neuroscience. There are further worries about reduction as a model of explanation or as a metaphysical thesis. Although such worries cannot be sidestepped entirely, my present project is limited to evaluating reduction as a model of interfield integration and the unity of neuroscience.[3]

This paper is partly historical and partly constructive. In the historical component (Sections 3–5), I consider the development of a prominent multifield research program in contemporary neuroscience: the learning and memory (LM) research program.[4] Several philosophers of neuroscience have described the LM research program as an exemplar of reduction. I argue, in contrast, that the development of the LM research program reveals at least three limitations of reduction models: first, that they neglect upward-looking aspects of interlevel interfield integration; second, that they ignore intralevel forms of interfield integration; and third, that they gloss over the fact that scientific progress has sometimes been achieved by abandoning reduction as an explanatory goal. Indeed, the LM research program needed to move beyond reduction to formulate an adequate explanation of learning and memory. Through the 1970s, as interdisciplinary fervor culminated in the SfN's foundation, the LM research program began to pursue multilevel and mechanistic forms of explanation. This mechanistic shift provided an abstract multilevel structure both onto and around which diverse fields could integrate their results. Scientists engaged in this pursuit integrate fields by adding constraints on mechanisms. The constructive portion of the paper examines this constraint-based and mechanistic form of scientific integration. In Section 6, I show how the pursuit of mechanisms scaffolds interfield integration at a given level. In Section 7, I illustrate how attention to mechanistic relations clarifies constraints on interlevel integration. I conclude that

[3] I focus on explanation and metaphysics in Craver (forthcoming).

[4] The phrase 'learning and memory' is problematic, since there is no single nontrivial feature shared by all instances of this conjunction, but it is nonetheless standard. And so I will adopt it.

mechanistic models of interfield integration are more historically accurate than reduction in describing the LM research program, that they reveal neglected varieties of interfield integration, and that they are more informative about the kinds of constraints needed to build the bridges from molecules to behavior.

2. Reductive views of interfield integration

There are many different reduction models (see, e.g., Nagel, 1949; Hooker, 1981; Schaffner, 1969, 1993a; Bickle, 1998), and there is no need to review the details of each. It will be more instructive for my purposes to develop a prototype of reduction. By prototype I mean a list of features that hold for many reduction models even if possibly no model has all of them. The mechanistic approach rejects each of these features.

First, reduction is an intertheoretic relationship. This means that it is a relationship between *theories*, rather than between items in the world, fields of science, or experimental techniques. Items in the vocabulary of the reduced theory are related via synthetic identities or partial definitions to items in the vocabulary of the reducing theory. While these terminological relations are intended to mirror relationships among items in the world or fields of science, the reduction model focuses attention on theoretical vocabularies and the effort to bring them into alignment. It is *terms* that are identified, it is *vocabularies* that are revised, and it is *meanings* that coevolve. To say that reduction is intertheoretic also means that it is a relationship *between* theories rather than within a single theory. The mechanistic model, in contrast, describes integration as occurring in the process of building a single theory.

Second, reduction is typically described as an interlevel relationship. Theories about phenomena at a higher level (e.g., gases, lightning, and life) are reduced to theories about phenomena at lower levels (e.g., molecules, electrons, and physiological systems). The levels relation is generally understood as a part–whole relation. For this reason, interlevel reduction has been called 'microreduction'. Precisely the same formal reduction models have been used to describe the relationship between a theory and its successor (e.g., between Newtonian mechanics and special relativity). However, microreduction and successional reduction are frequently different kinds of reduction (Nickles, 1973), serving different scientific purposes (Wimsatt, 1984 [1976]), and so they should be considered separately. My focus here is exclusively on microreduction. The mechanistic approach I develop below suggests a reasonable way to understand the level relationship and insists on recognizing interfield relations that oscillate upward and downward in a hierarchy of levels.

Third, reduction models are formally specified. Analyses of reduction often presuppose a formal analysis of theory structure. One begins by assuming that the theory can be axiomatized in first order predicate logic, articulated as a set-theoretic predicate, or represented as a partition of a multidimensional state space. The reduction relation is then spelled out using the apparatus of that formal system. This means that the presupposed view of theories is abstracted from the mechanistic structures in neuroscientific theories. It is these mechanistic structures, involving

causal, constitutive, spatial, and temporal relations, that scaffold[5] the instances of interfield integration in the LM research program. The mechanistic approach focuses its attention on the relationship between the phenomenon to be explained and the mechanism that explains it—that is, on the causal structure of the world (cf. Salmon, 1984). Fields are integrated when they collaborate in revealing a portion of the causal nexus.[6]

There are many ways to characterize the intertheoretic relationship involved in reduction, but most accounts require at least that the lower and higher level theories be homomorphic. This means that there should be some possible mapping from the items and patterns in one theory onto a set of items and patterns in the other. In classical reduction, the intertheoretic relation is deductive. Subject to certain restrictions and approximations, and coupled with identities (or partial identities) between the terms of the two theories, one literally derives the higher-level theory from the lower. The derivational requirement provides a clear regulative ideal—anything less than derivability is incomplete or sketchy. Other accounts of reduction require merely that one be able to form within the language of the reducing theory an 'equipotent image' of, or a structure homomorphic with, the reduced theory (see, e.g., Bickle, 1998). Efforts to spell out these weaker relations sacrifice some of the clarity and precision that makes classical reduction so attractive.

Finally, there is little agreement among advocates of reduction about whether or not the model should be taken as an empirical thesis that science exhibits a reductive trend from early work at higher levels to more mature work at lower levels. Oppenheim and Putnam clearly intended their reductive model this way, and Bickle (2003) argues that contemporary neuroscience continues to exhibit a reductive historical trend. Yet, many of the advocates of reduction admit that the model poorly describes the practice and history of science. Schaffner (1993a), for example, argues that reduction is 'peripheral' to scientific practice and should be regarded merely as a 'regulative ideal', that is, the goal state of an ideally complete explanation. Churchland, following Francis Crick, admits that reductions are rare and typically occur only after the interesting 'co-evolutionary' work has been done (see 1986, p. 285).

As I will argue over the next two sections, my own stance on the peripherality thesis is mixed. On the one hand, I agree with Churchland and Schaffner that reduction in the sense just developed is rare and largely peripheral to scientific practice. The LM research program exhibits no predominant trend from higher toward lower levels or toward explanations that comport with formal reduction models. In fact, peripherality is the primary drawback of reduction models: they ignore most of what is interesting about the development of the contemporary LM research program and the integration of fields in the search for mechanisms. Reductive explanations did appear in the early history of the LM research program. But the research program

[5] The term 'scaffolding' is loosely borrowed from Andy Clark (1999). I mean to suggest that the abstract structure of a mechanism (with its active, spatial, temporal, and constitutive organization) serves as a generic framework used to build explanations and theories in neuroscience.

[6] Schaffner (1993a) took some initial steps toward building a multilevel and causal view of reduction that inspired my approach in this paper.

eventually abandoned reductive aspirations in favor of the goal of building multi-level descriptions of mechanisms.

3. Technical background: Long-Term Potentiation and LM mechanisms

I will focus specifically on the protracted discovery of an electrophysiological phenomenon now known as Long-Term Potentiation (or LTP). LTP is the enhancement of a synapse resulting from rapid and repeated electrophysiological stimulation (a *tetanus*) to the pre-synaptic neuron. Many believe that this laboratory phenomenon demonstrates the existence of processes in the brain that might underlie synaptic plasticity during learning or memory.

LTP was first encountered in the hippocampus (shown in Figure 1a), a region within the brain's temporal lobes that is commonly associated with learning and memory. Damage to the hippocampus produces profound forward reaching (i.e., anterograde) amnesia in humans and in animal models. This is dramatically illustrated by the story of H. M. (see Scoville & Milner, 1957). In an effort to relieve life-threatening epileptic seizures, H. M. consented to experimental surgery that involved removing the hippocampus from both sides of his brain. When he awoke from the surgery, he was profoundly amnesic, but in a highly specific way. H. M. could still read and write, he could learn new skills, and he could be classically conditioned and primed. He could reportedly remember much of his childhood, and his I.Q. if anything increased after the surgery. His amnesia was fairly specific to declarative memories (memories for facts and events, episodic or semantic) as opposed to procedural memories (memories for skills or habits), leading researchers to conclude that there are separate brain systems for these types of learning.

LTP can be induced in many ways in synapses throughout the brain. Figure 1a shows electrodes stimulating the perforant path input to the hippocampus and recording from the dentate gyrus region of the hippocampus. Figure 1b depicts LTP at this synapse. The top shows the electrophysiological stimulus delivered to the pre-synaptic cell, and the bottom shows the post-synaptic electrophysiological response. Starting on the left, a test stimulus to the pre-synaptic cell in the perforant path (delivered at a rate of one per second) elicits a standard excitatory response in the postsynaptic cell in the dentate gyrus. After delivering a tetanus to the pre-synaptic cell (top center) with corresponding depolarization of the post-synaptic cell (bottom center), the same test stimulus as before produces a heightened response from the post-synaptic cell (as shown on the right). As shown below this diagram, this potentiation can last for hours, days, and perhaps weeks. The duration of LTP contributes to its plausibility as a potential LM mechanism. Another aspect of LTP's plausibility as an LM mechanism derives from the fact that it is necessary to depolarize both the pre-synaptic and the post-synaptic cell to induce LTP (see Figure 1c). As shown in the middle column on the bottom, LTP is not induced if one prevents the post-synaptic cell from depolarizing during the tetanus. When a voltage clamp delivers current to counteract post-synaptic depolarization, LTP is not induced (Figure 1c). This property of LTP has the same abstract structure as

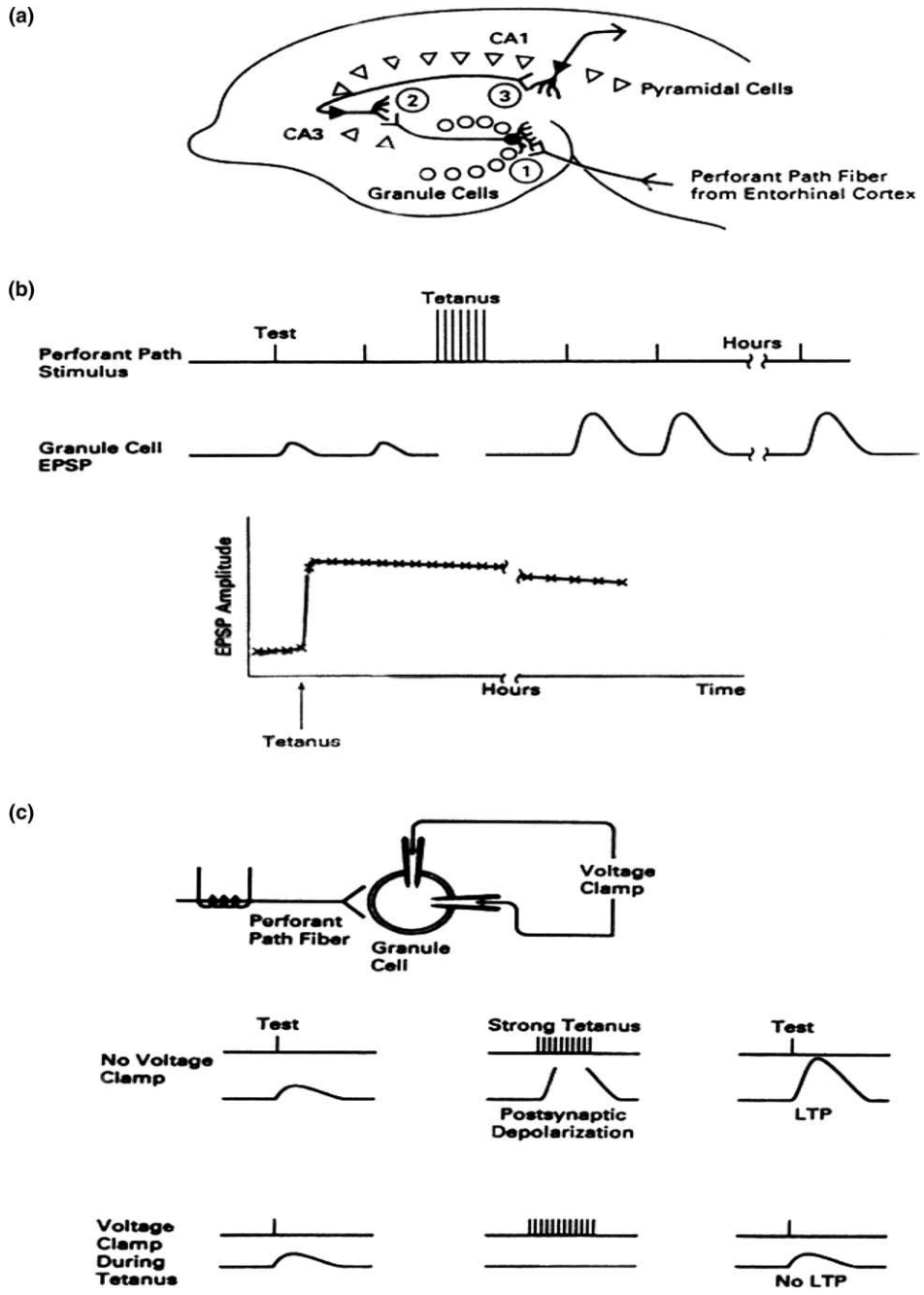

Fig. 1. Features of LTP at the synapse between the perforant path and the granule cells of the dentate gyrus (from Levitan & Kaczmarek, 1991, used with permission).

associationist views of learning (most notably, that of Donald Hebb, 1949): simultaneous co-activation of neurons (thoughts) strengthens their connection (association).

A standard scientific history of this episode (see, e.g., Squire & Kandel, 2000) portrays the discovery of LTP as an exemplar of reduction. This scientific history suggests that researchers started out believing that the hippocampus was an organ of memory and then went searching for evidence of plasticity in its synapses. The history then depicts progress in this research program as proceeding to ever deeper and ultimately molecular levels. But as we will now see, this story is a rational reconstruction, not an accurate history.[7]

4. LTP's origins: not top-down search but intralevel integration

Most neuroscientists date the discovery of LTP to a series of 1973 papers by Tim Bliss, Terje Lømo, and Tony Gardner-Medwin. These papers became a watershed in the development of the LTP research program. However, the story of LTP begins much earlier. In recovering the earlier history, we see the origins of LTP not as an instance of reduction, but rather as an instance of intralevel (and so nonreductive) interfield integration.

Electrophysiologists first produced and reported synaptic plasticity in the hippocampus in the 1950s. Plasticity was initially encountered as an experimental tool. Although hippocampal electrophysiologists now commonly experiment in *in vitro* slice preparations, these techniques did not exist until the 1970s. In the 1950s, researchers worked *in vivo*, inserting electrodes through the skulls of anaesthetized rabbits. They were pressed to produce as much data as possible before anesthesia, blood loss, and repeated stimulation conspired to run the neural responses down. As a trick of the trade, electrophysiologists knew that they could reawaken the signals by turning the stimulus rate knob to a higher frequency for a few seconds (see Andersen, 1991, 2003). But they typically ignored the phenomenon, only rarely mentioning it in print (see, e.g., Green & Adey, p. 250; Andersen, 1960a, p. 191; 1960b, p. 216; Andersen, Bruland, & Kaada, 1961; Gloor, Vera, & Sperti, 1964). None of these papers reports potentiation lasting longer than ten minutes, and none connects the phenomenon to learning or memory.

Why were these electrophysiologists investigating the hippocampus? It is reasonable to think that they were drawn to the hippocampus because of its link with learning and memory, suggesting a downward trend in the LM research program. However, although some neurologists had associated the temporal lobes with learning or memory prior to these early reports, only a minority had implicated the hippocampus specifically. The idea of localizing memory anywhere in the brain was unfashionable at this time. Karl Lashley, for example, had performed systematic lesion experiments in pigeons that apparently showed that the severity of learning and memory deficits depended not on the lesion's location, but rather on the volume of

[7] See Craver (2003) for a more detailed account.

brain tissue removed. Lashley suggested, and many agreed, that memory storage is distributed across the entire brain.

At the time, the hippocampus was commonly associated with other functions. Most neuroscience texts of the day associated it with olfaction, but other reports suggested that it might be involved in behavioral inhibition, emotion, fear, ingestive behavior, sexual activity, sleep, and respiration. Given such a wide range of phenomena, it is no wonder that the reports of hippocampal synaptic plasticity failed to mention any association with learning or memory. In some cases, they explicitly declined to speculate as to the 'true role' of the hippocampus, claiming that, 'many more physiological and behavioral studies will be needed before any systematic correlation with the anatomical structure can be attempted' (Cragg & Hammlyn, 1957, p. 483).

If these electrophysiologists were not looking for a neural correlate of memory, as the reductive reconstruction leads us to believe, then why were they looking in the hippocampus? There are many reasons. Some explored the hippocampus because of its long suspected role in the origins of epilepsy. Many electrophysiologists were interested in the hippocampus as a model for investigating the spread of electrical activity in circuits of neurons. Many researchers were attracted to the hippocampus because it was an anatomically well characterized experimental system. The rodent hippocampus is large and readily accessible through the skull, making it attractive for *in vivo* studies. Researchers also thought of the hippocampus as an evolutionarily primitive structure and one likely to reveal highly conserved patterns of neural organization.

In Oslo, where the first long-lasting form of hippocampal plasticity was reported, the hippocampus was seen as an especially attractive experimental model for integrating results from the fields of anatomy and electrophysiology. Anatomists used Golgi staining (which stains ten percent of the neurons, allowing one to see neural structures and organization clearly) and terminal degeneration studies (which kill the cell body and then trace synapses by staining for degenerating axons) to reveal neuronal locations, structures, and connections. Electrophysiologists used microelectrodes to study the propagation of neural excitation through the hippocampus. These anatomical and electrophysiological experiments were both directed at cells and their connections with one another. The approaches of these two fields did not bridge levels, but rather combined to reveal different aspects of precisely the same level. The goal was to understand how neural wiring diagrams and electrical activities are related in a well defined neural circuit. It is in the context of this interfield project that Lømo, and later Bliss, would encounter the phenomenon now known as LTP.

To recap: hippocampal synaptic plasticity was not discovered in the top-down search for the neural correlate of memory; rather, it was noticed in an intralevel research project that combined anatomical and electrophysiological perspectives. Such intralevel varieties of interfield integration are far more common than their interlevel counterparts, and they are not even within the purview of reductive models of interfield integration. As we will see in Sections 6 and 7, attention to the abstract structure of mechanisms and levels reveals constraints on mechanistic organization that act as loci for interfield integration both at a given level and between levels.

5. The mechanistic shift

How did LTP gain its association with learning and memory? This upward-looking phase of the research program, which continues to the present, only developed after the initial discovery of *long lasting* (i.e., greater than ten minutes) forms of potentiation in the hippocampus. Some neuroscientists in the 1950s and 1960s had a reductive view of the connection between synaptic changes and learning or memory. However, with the LTP watershed, these reductive aspirations were subtly replaced by mechanistic explanatory goals.[8] Reduction as an historical thesis (as Oppenheim and Putnam formulated it) is misleading because the science is both upward-looking and downward-looking at different times. The historical thesis also glosses over a major shift in the research program's explanatory goals that involved implicitly rejecting reduction as an explanatory goal in favor of a more ecumenical mechanistic objective.

Warren McCulloch and Walter Pitts, for example, had clear reductive aspirations:

> The 'all-or-none' law of nervous activity is sufficient to insure that the activity of any neuron may be represented as a proposition. Physiological relations existing among nervous activities correspond, of course, to relations among the propositions; and the utility of the representation depends upon the identity of these relations to relations among the propositions. To each reaction of any neuron there is a corresponding assertion of a simple proposition. This, in turn, implies either some other simple proposition or the disjunction or the conjunction, with or without negation, of similar propositions according to the configuration of the synapses upon and the threshold of the neuron in question. (McCulloch & Pitts, 1943).

In this explanatory schema, propositions are identified with all or nothing action potentials, and the interrelationships among action potentials in a network are literally identified with complex propositions (e.g., conjunctions, disjunctions, and negations) and with inferences among propositions (e.g., from the activation of two propositions separately to the activation of their conjunction). Changes in beliefs or inference patterns, as might be involved in learning or memory, might be represented as changes in connection strengths among the cells. It would be difficult to find a clearer statement of reductive goals.

In fact, Oppenheim and Putnam (1958) cite McCulloch and Pitts along with other mathematical biophysicists, as showing that phenomena at the 'level' of the whole organism (psychology) are reducible to phenomena at the 'level' of cells. They use this development as evidence for their historical thesis that science exhibits a trend towards reductive unity: 'In terms of such nerve nets it is possible to give hypothetical micro-reductions for *memory, exact thinking, distinguishing similarity or*

[8] Those who associate reduction merely with the search for mechanisms (e.g., Sarkar, 1992; Wimsatt, 1984 [1976]) are thus failing to mark a significant difference between the explanatory goals of LM research in different historical periods.

dissimilarity of stimulus patterns, abstracting of "essential" components of a stimulus pattern, recognition of shape regardless of form and of chord regardless of pitch (phenomena of great importance to Gestalt psychology), *purposeful behavior* as controlled by negative feedback, *adaptive behavior*, and *mental disorders*' (Oppenheim & Putnam, 1958, p. 20; italics in original). Something like the goal of reducing learning to synaptic changes is implicit in early reports of synaptic plasticity coming out of Oslo. The Oslo claim was not, following McCulloch and Pitts, that learning in organisms is identical to synaptic changes, but rather, that synaptic changes are a simple and primitive *type of* learning. Andersen and Lømo, for example, describe a relatively short-lasting form of potentiation as being 'of some interest in connection with', 'an indication of', and 'an example of' a synaptic learning process. Similar language can be found in the work of Sir J. C. Eccles, with whom Andersen worked from 1961 to 1963. Unlike the Oslo community, Eccles was specifically interested in neural correlates of learning and was looking for evidence of such changes in spinal reflex circuits. Eccles did manage to potentiate these synapses, but only by using physiologically implausible stimulation frequencies and intensities. Still, Eccles thought of this 'residual potentiation' in the spinal cord as a *kind of* memory. He claimed, for example, that 'disused synapses are capable of "learning" to operate more effectively' (Eccles, 1953, p. 209) in the sense that repeated stimulation (experience) strengthens the synapse (long-lasting change). Andersen and Lømo found potentiation in the brain, but it did not last long enough to be plausibly explanatorily relevant to learning in organisms (Andersen & Lømo, 1967, p. 410). For this reason, they describe their phenomenon as merely a simplified model or exemplar of memory—not a potential memory mechanism.

In 1966, Lømo published an abstract reporting a form of tetanus-induced plasticity in the hippocampus that would last for hours. And in 1968, when T.V.P. Bliss came to Oslo, he and Lømo teamed up to characterize the phenomenon and to argue that it was more than a laboratory curiosity or experimental artifact. This work culminated in the 1973 LTP watershed (Bliss & Lømo, 1973; Bliss & Gardner-Medwin, 1973; Bliss, Gardener-Medwin, & Lømo, 1973). It is widely appreciated that these three papers are the first to characterize the LTP phenomenon in detail and the first to suggest that it might be associated with learning or memory. It is not widely recognized that these three papers subtly reconceptualized the link between LTP and memory.

This mechanistic shift involved coming to see LTP not as *identical* to memory or as a *kind* of memory, but rather as a *component* in a multilevel memory mechanism. This mechanistic shift guided subsequent research by clarifying two basic goals. The first goal (the exclusive focus of Bliss and Lømo's, 1973 discussion section) was to discover, 'the mechanisms which might be responsible for long-lasting Potentiation'. Such molecular pursuits remain a mainstay of the LM research program to this day. The second goal was and is to evaluate the role of LTP in higher-level memory mechanisms. Bliss and Gardner-Medwin (1973) approached this topic by attempting to produce LTP in awake and behaving rabbits. Sleeping rabbits do not learn, and if LTP is a learning mechanism, it should be possible to produce it without anesthesia. After demonstrating that LTP could be produced in approximately normal physiological

conditions, Bliss, Gardner-Medwin, and Lømo (1973) appealed to LTP's theoretical plausibility as a synaptic model of memory and to LTP's physiological plausibility as a Hebbian learning mechanism. The three papers also, not coincidentally, appeal to results from multiple fields (ablation experiments, biochemical assays, EEG recordings, and psychiatric case studies) to argue for the potential relevance of the hippocampus to memory.

So, as of 1973, LTP was no longer identical to or an example of memory; it was a component in a multilevel memory mechanism. This shift in explanatory perspective clearly defined the goals of the LTP research program and at the same time, sketched a theory around the LTP phenomenon that could serve as a framework for integrating fields in the young neurosciences of memory. Anatomists, biochemists, electrophysiologists, mathematicians, psychologists, and psychiatrists could all potentially contribute to understanding either the mechanism of LTP or the memory mechanisms containing LTP as a component.

This revised history helps to dislodge three reductionist assumptions about the LM research program. First, the history of the LTP research program provides a clear historical counter example to those (such as Oppenheim & Putnam, 1958; or Bickle, 2003) who tout reduction as an empirical hypothesis. In the 1960s, the development of the LM research program was decisively upward: looking for a way to connect this synaptic phenomenon into a higher-level mechanism. Oppenheim and Putnam may have been right about the historical trends apparent in the 1950s, but the trend turned out not to be projectable onto the subsequent development of the research program. Bliss and Lømo are the first to articulate clearly this upward-looking connection of LTP to learning and memory. In the present stage of research, work procedes both downward and upward. For example, dominant areas of research involve both molecular techniques (e.g., gene targeting) and techniques for monitoring functional activation in the whole brain (e.g., PET and fMRI). The point is that research rarely shows a single direction through a hierarchy, but rather involves oscillating up and down within a hierarchy as new problems call out for solutions and as new techniques become available. It is only with a reductive bias antecedently in mind that one can see the history of the LM research program as evidence for an empirical trend toward a single lower level theory.

Second, the reductive focus on interlevel relationships distracts attention from interfield integration at a single level. Anatomical investigations of the cells of the hippocampus and electrophysiological research into the activities of those cells are not at different levels in any meaningful sense: both of the fields were concerned with neurons and hippocampal wiring diagrams. At least since Oppenheim and Putnam, reductionists have assumed that there is a one-to-one mapping among fields of science, scientific theories, and levels of organization. The mutlifield climate of mid-twentieth-century neuroscience, as exemplified in the LM research program, highlights just how restrictive this assumption is. Fields can span multiple levels or investigate phenomena at the same level. Furthermore, the goal of this interfield research program was not to translate the theories of anatomy into the theories of electrophysiology or to exhibit the theories as in any sense homomorphic to one another. Instead, the techniques and perspectives were uniquely able to reveal different aspects

of the same mechanism at a single level. Reduction models of interfield integration ignore this intralevel form of interfield integration.

Finally, the LM research program abandoned reduction as an explanatory goal and adopted a mechanistic approach. The origins of the LM research program span the epoch in which the neural sciences coalesced into Neuroscience. It was in this climate that the reductive goals of the LM research program were replaced by the goal of building descriptions of multilevel mechanisms. And in this climate, the goal of describing a multilevel mechanism provided an abstract structure that could accommodate the findings from diverse fields.

6. Mechanisms and intralevel integration

Researchers in the LM research program are not just building theories *simpliciter*; rather, they are building theories about mechanisms. Perhaps mechanisms can be described using formal accounts of theories—perhaps they can be axiomatized in predicate logic or reconstructed as set theoretic predicates. But such formal accounts of the structures of scientific theories gloss over the mechanistic structures crucial for understanding how these theories are constructed and evaluated. We cannot hope to understand the forces driving the co-evolution of fields in neuroscience without at least supplementing formal accounts of theory structure with attention to the mechanistic structures of neuroscientific theories.

What are mechanisms? Let us begin abstractly before considering an example. Mechanisms are collections of entities and activities organized together to do something (cf. Machamer, Darden, & Craver, 2000; Craver & Darden, 2001; Bechtel & Richardson, 1993; Glennan, 1996). The entities in LM mechanisms include things like pyramidal cells, neurotransmitters, hippocampi, and rats. The activities are the various things these entities do: pyramidal cells fire, neurotransmitters bind, hippocampi construct spatial maps, and rats learn. These entities and activities are organized such that they carry out some process, perform some function, or produce some end product. Call this the *phenomenon to be explained by the mechanism*.

In our present discussion, organization is the crucial feature of mechanisms. Different aspects of organization are more or less important in different kinds of mechanisms. One aspect of mechanistic organization is spatial. Different stages of the mechanism may be *compartmentalized* within some boundary or more or less *localized* within some 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 and upon those structured entities being *oriented* with respect to one another in particular ways. A second aspect of mechanistic organization is temporal. The activities in mechanisms are *ordered* as earlier or later than one another, and they have characteristic *rates* and *durations*. These spatial and temporal forms of organization underlie the active organization of mechanisms. It is because the entities have the spatial properties that they do, and because things happen when they do or as quickly as they do, that the mechanisms work. It is a useful, if mundane, fact about mechanisms that intervening to change the organization of

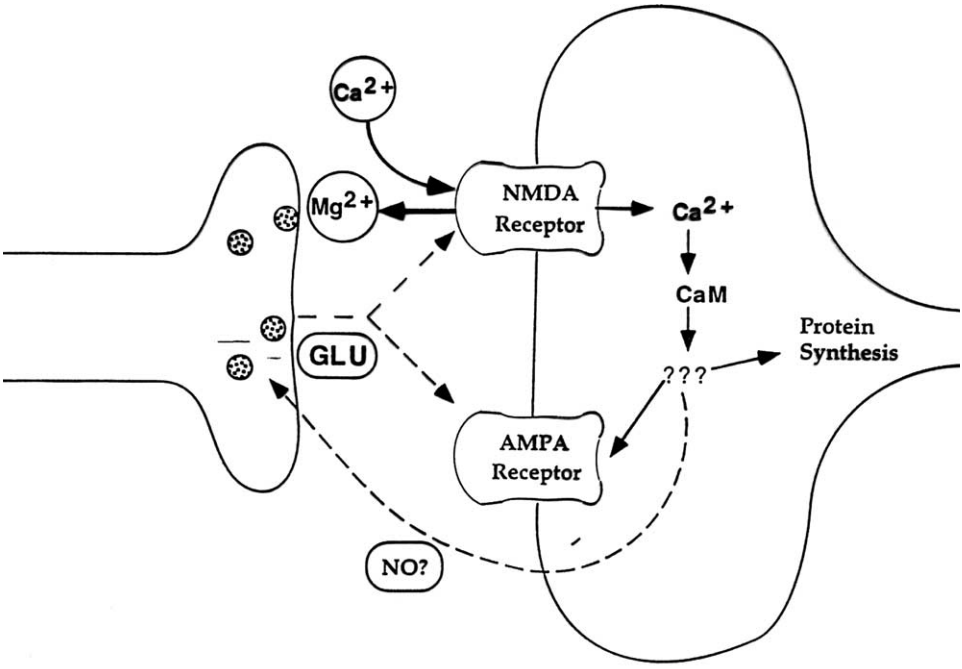

Fig. 2. Cartoon sketch of LTP mechanisms.

their components (by moving parts, changing their shapes, delaying their activities) can dramatically change how they behave.

The organization of mechanisms sustains their productive continuity—one stage of a mechanism gives rise to, drives, makes, or allows its successor in the order of productivity. Diagrams of mechanisms, such as the one in Figure 2, trace the active organization of the components. Each arrow shuttles from one region of the static diagram to another, constituting a stage of the mechanism's operation. The arrows trace the direction or flow of productivity through the mechanism. The goal in describing a mechanism—an alternative regulative ideal to the deductive ideals implicit in reduction—is to describe this productive flow, without gaps from the beginning of the mechanism to its end. When we show how a mechanism works, we describe the relevant parts of the mechanism and their relevant properties, and we show how their activities are organized together so that the working of the mechanism is transparent and unmysterious. As such, the search for mechanisms is driven by the goals of replacing black boxes (the question marks of Figure 2) and filler terms (e.g., filler action verbs such as 'causes', 'activates', 'inhibits', 'modulates') with descriptions of mechanisms.[9] This regulative ideal is often beyond the reach of

[9] Keil and Wilson's (2000) distinction between superficial and deep forms of explanation might be made sharper by noting that deep explanations describe productively continuous mechanisms while superficial forms have gaps and promissory notes.

contemporary neuroscience, where gappy or highly schematic descriptions of mechanisms are more common. But an ideal mechanistic explanation would fill these gaps and replace these filler terms with descriptions of productively continuous mechanisms.

As an example of a mechanism, consider a cartoon of the molecular and electrophysiological mechanisms of Long-Term Potentiation (LTP) as they were described in the 1980s (sketched in Figure 2). The pre-synaptic neuron (e.g., the perforant path from Figure 1) is on the left, and the post-synaptic neuron (e.g., the dentate gyrus of Figure 1) is on the right. When the pre-synaptic neuron is active, it releases glutamate. Glutamate binds to so-called NMDA (for N-methyl D-aspartate) receptors on the post-synaptic cell. The NMDA receptors change their conformation, exposing a pore in the cell membrane. If the post-synaptic cell is in its polarized resting state, the channel remains blocked by large Mg^{2+} ions. But if the post-synaptic cell is depolarized, these Mg^{2+} ions depart the channel, allowing Ca^{2+} to diffuse into the cell. The rising intracellular Ca^{2+} concentration regulates the biochemical pathways required for LTP. The NMDA receptor thus acts as a gate for the induction of LTP, ensuring that it is only induced when the pre- and post- synaptic cells are simultaneously active (i.e., ensuring that LTP is cooperative).

The changes induced in this biochemical pathway through the influx in Ca^{2+} are thought to result in several changes that could underlie LTP. In the short-term, this cascade might add new receptors to the membrane, or alter their sensitivity to glutamate, or change their ability to conduct ions. Such changes might account for the rapid induction of LTP. In the long-term, the cascade leads to the production of proteins used to alter the structure of the synapse. Some suspect that there is also a pre-synaptic component of this mechanism, whereby, for example, the pre-synaptic cell releases more glutamate with each action potential.[10]

The entities in this mechanism are glutamate molecules, neurons, NMDA receptors, and Ca^{2+} ions; the activities include binding, diffusing, and changing conformation. These entities and activities are organized in the induction and maintenance of LTP. The components are *temporally organized*, beginning with the release of glutamate and terminating in structural changes that strengthen the synapse. The rates and durations of the different stages are crucial for the working of the mechanism. Short-term modification of the AMPA receptors is thought to account for LTP's rapid induction, and long-term production of proteins is thought to account for LTP's persistence. The mechanism is also *spatially organized*. Stages of the mechanism are compartmentalized or localized in cells, membranes, and pores. Ignoring the question marks, the early stages of the mechanism are *actively organized*; they are causally organized with one another, exhibiting a variety of activities and interactions.[11] These activities depend crucially on the structures and orientations of the relevant entities; for example, the size of the pore and the complementary shapes of glutamate and the NMDA receptor allow the relevant activities to occur in the right places at

[10] For more on these mechanisms, see Frey & Morris (1998); Squire & Kandel (2000).

[11] Machamer, Darden, & Craver (2000) abandon talk of causation in favor of 'activities'. This has been discussed by Tabery (2004). Nothing in this paper turns on any particular account of the causal relation.

the right times. The productive continuity at both ends of this cartoon mechanism is interrupted by gaps in the middle—the question marks that must be filled in for the description of the mechanism to be complete. Sometimes these question marks are replaced by filler terms, such as 'leads to', 'produces', or 'regulates'. A contemporary account of this mechanism could fill in many (but not all) of these question marks and filler terms.

Mechanistic theory building typically proceeds through the piecemeal accumulation of constraints on the space of possible mechanisms for a given phenomenon. A constraint is a finding that either shapes the known boundaries of the space of possible mechanisms or changes the probability distribution over that space. Constraints involve identifying the relevant entities and activities and detailing their organization. Some constraints serve the negative function of excluding regions of the space of possible mechanisms, showing that some set of mechanisms is implausible or impossible given what is known about the parts and their organization. Some constraints suggest new regions of that space, as when the double helix famously suggested to Watson and Crick a mechanism of reproduction. Constraints on the space of possible mechanisms constitute evidence for evaluating descriptions of mechanisms and act as loci for interfield integration. It is probably fruitless to say in advance how many kinds of constraints there are or which constraints will be relevant to understanding a given mechanism; however, many of the constraints simply involve finding aspects of a mechanism's organization.

Scientists in different fields use different techniques to investigate different kinds of constraints on different components of the same mechanism. When the findings of two fields cooperate as constraints on a mechanism, the fields are, perhaps only rather locally, integrated. For example, in Oslo, anatomists provided spatial constraints on the mechanism by revealing where the neurons are, how they are connected to one another, and how they are organized into a circuit. Electrophysiologists, on the other hand, could study temporal constraints on the electrophysiological activities of the hippocampus by stimulating and recording at different points in this circuit. These fields concentrate on different components: anatomists focus on neurons and electrophysiologists focus on electrical activities such as action potentials, excitatory post-synaptic potentials, and LTP itself. Each of these varieties of constraint is used to fill in an abstract mechanistic structure with more detail as researchers try to describe a productively continuous mechanism. This variety of scientific integration is exclusively intralevel, and so it is a form of scientific integration that reduction cannot handle. For the kinds of interfield integration considered thus far, the phenomena are not at different levels in any interesting sense; rather, the fields are investigating different components or stages of the same mechanism. The terms describing the different constraints are not translated into one another. Nor are the different constraints identified with one another. Rather, the constraints open or close different portions of the space of possible mechanisms for a given phenomenon. Finally, the different constraints are usually not in any interesting sense homomorphic with one another; they may be as different as anatomy and electrophysiology. Recognizing these forms of interfield integration leads us to see just how myopic reduction models are when it comes to describing the unity of science: at best they focus on very special cases of

interlevel integration. But in multifield sciences such as neuroscience, abstract mechanistic structures scaffold interfield integration.

7. Mechanisms and interlevel integration

The mechanistic approach also has many advantages over reduction for thinking about interlevel forms of interfield integration. First, it provides a straightforward way to interpret the talk of 'levels', which is a problem almost entirely neglected in much of the literature on reduction. Second, it offers significantly more insight into what interlevel integration is, into the kinds of evidence by which interlevel bridges are evaluated, and into the forces driving the co-evolution of work at different levels. Finally, it dispenses with the weak requirement that phenomena at different levels should be homomorphic with one another and replaces it with the regulative ideal of describing a productively continuous lower-level mechanism. Consider the LM case again.

At least since 1973, when Bliss and Lømo linked LTP with learning and memory and with the molecular mechanisms in the synapse, the theory surrounding LTP has spanned multiple levels. Although there is room for dispute about exactly how many levels there are and how they are to be characterized, four levels figured prominently in the above story. At the top of the hierarchy is a kind of learning or memory or the performance of some learning and memory task, such as learning and remembering how to run a maze. Below that is the hippocampus generating spatial maps, consolidating information, or tutoring the cortex. Beneath the hippocampus in this hierarchy are the synapses between the cells of the hippocampus and their electrophysiological activities. And finally, at the present lowest level, are the activities of the molecules that make up the synapse, for example, the NMDA receptor activating and inactivating.

Talk of 'levels' is ambiguous, but it is not beyond repair.[12] One simply needs to be clear about which sense of 'level' is in play. The levels in the LM case are best thought of as levels of mechanisms, which satisfy the following two requirements: that the levels are related as parts to wholes and that the parts are components in a mechanism. The first requirement ensures, as most intuit, that things at higher levels are larger than things at lower levels, things at higher levels are composed of things at lower levels, and things at higher levels tend to exhibit different regularities from things at lower levels. The second requirement emphasizes that not just any parts will do. Irrelevant parts (i.e., parts within the spatial boundaries of a

[12] Wimsatt (1984 [1976]) has the most extended treatment of levels. He develops a prototype view, according to which levels are local maxima of regularity and predictability plotted against a logarithmic scale of size. Levels are (among other things) where the entities are, where theories are pitched, where causal interactions are contained, and where things of similar size are clustered. I argue against this wonderfully rich treatment of levels in Craver (forthcoming), suggesting that a prototype approach glosses over differences in the way that levels talk is used for different scientific ends.

mechanism but not a part of the mechanism), arbitrary parts (i.e., those that are merely spatial sub-sections of the whole), or gerrymandered parts (i.e., arbitrary summations of parts in either of these first two senses) are on this account excluded from the interlevel relation. To rule out such cases, containment or spatial relations must at least be supplemented with a componency relationship explicit in the second requirement.

The next question that must be addressed regarding levels is: what does it mean to integrate levels? The 1973 LTP watershed defined two integrative goals for the research program: one upward-looking and the other downward-looking. The upward-looking aspect of interlevel integration involves showing that an item is a component in a mechanism and describing its role in that mechanism. For example, one shows that the lower-level NMDA receptor is a component in the higher-level mechanism of LTP induction, or that the lower-level hippocampus is a component in the higher-level mechanisms of memory. In contrast, the downward-looking aspect of interlevel integration involves detailing lower-level mechanisms for a phenomenon. To integrate computation in the hippocampus with lower levels, one would show how the networks of cells are organized such that they execute the relevant computation in the hippocampus. These upward-and downward-looking tasks are carried out, like their intralevel counterparts, by placing constraints on the interlevel relationship. Consider some different interlevel constraints.

First, the interlevel relationship in mechanistic hierarchies is a relationship between a mechanism and its phenomenon. The character of the phenomenon to be explained by a mechanism places *accommodative constraints* on any mechanism that will produce it. (Bechtel & Richardson, 1993, call these 'phenomenal constraints'; Churchland & Sejnowski, 1992, describe them as a driving force in the co-evolution of levels.) The description of the mechanism must be responsive to changes in how the phenomenon is understood. The boundaries of the mechanism are fixed by reference to the phenomenon to be explained (Bechtel & Richardson, 1993; Glennan, 1996). Only relevant entities, activities, and properties are inside the mechanism. Changing the phenomenon to be explained, by definition, changes the most fundamental constraint on the mechanism: that it contains the relevant components for the given phenomenon.

Second, just as the description of a mechanism is constrained by discoveries concerning higher levels, so the higher-level character of the phenomenon often must be accommodated to findings about lower-level mechanisms. Memory is a classic example (see Churchland & Sejnowski, 1992). Researchers now recognize several kinds of learning and memory. There are short-term and long-term memories, echoic and iconic memories, and explicit and implicit memories, to name a few. Evidence for many of these distinctions comes from the fact that different kinds of memory can be 'dissociated' from one another: accidental or experimental damage to the brain (lower-level) can disrupt one kind of memory (higher-level) while leaving the other kinds intact. For example, philosophers of neuroscience have appealed to the case of H. M. (see Section 3) to illustrate the ability of neuroscientific research to drive conceptual revision in psychology (e.g., Bickle, 1998; Churchland, 1986). It is uninformative to characterize this relationship between a phenomenon and its mechanism as

'homomorphism', as weakened versions of classical reduction sometimes suggest. If homomorphism means merely 'similarity', then asserting that higher and lower levels are homomorphic tells us very little about their relationship to one another unless one is able to say in more detail in what respects the two levels are similar.

What the mechanistic account offers beyond homomorphism as a regulative ideal is an emphasis on how the relevant parts of a mechanism are organized in productive continuity from beginning to end. Suppose, for example, that we want to explain why LTP is cooperative, that is, why it can only be induced when the pre- and post-synaptic neurons are simultaneously active (the 'cooperativity' shown in Figure 1c). The explanation, as we have seen, crucially involves the Mg^{2+} gate in the NMDA receptor. When we know the spatial, temporal, and active organization of the NMDA receptor and its relations with its environment, we see why it can act as a 'gate' under these circumstances. The explanatory burden is borne not by the homomorphism between the gate and LTP, but by how the components of the mechanism work together (cf. Salmon, 1984, pp. 274–275). Depolarization of the pre-synaptic neuron is relevant because the active pre-synaptic cell releases glutamate, which activates the NMDA channel. Depolarization of the post-synaptic cell is relevant because it releases the Mg^{2+} blockade, clearing the channel and allowing an influx of Ca^{2+}. We explain LTP by laying bare hidden portions of the causal nexus, describing the parts of the mechanism, their activities, and their organization, not by establishing homomorphism between the two levels.

The search for homomorphic mechanisms can often be misleading. As we saw, part of the appeal of LTP as a learning mechanism had to do with its cooperativity and with the possible mapping between cooperativity and classical associationism. As the LTP–LM research program has progressed, however, this simplistic view has been abandoned. No one still believes that memories are stored in single synapses or in single cells. The dominant idea guiding present research is that memories are stored in distributed representations over large populations of neurons. Homomorphism between cellular activity and associations among ideas in this case was initially suggestive, but had to be abandoned as the mechanisms became more complex. Where reduction accentuates identity and homomorphism among higher and lower levels, the mechanistic account emphasizes that lower-level parts are components in a mechanism whose behavior is the phenomenon to be explained.

Third, the relationship between a mechanism and a phenomenon is also constrained by spatial and temporal aspects of each. Spatial features of the phenomenon, such as its locus of control (Bechtel & Richardson, 1993), its role in still higher-level mechanisms, and any spatial patterns involved in its operation (e.g., spatial coding) place constraints on the mechanism. For example, if LTP is found to be a post-synaptic phenomenon, one must look inside the post-synaptic neuron for its mechanisms, and the spatial structures of hippocampal representations will have to be accounted for by the dynamics of the component neurons and their connections. Temporal features of the phenomenon (e.g., its rate or duration) also place constraints on the mechanism. No LM mechanism, for example, will be complete if it cannot account for the known slope of learning curves, effects of reinforcement schedules, the persistence of memory, and rates of forgetting. To my knowledge,

none of these temporal features has been precisely explained by contemporary LM researchers (contra Bickle's suggestion that this represents an accomplished psychoneural reduction, 1998), but such details will have to be given if a proposed mechanism is to be explanatory for the LM phenomenon. Different fields focus on different kinds of interlevel constraint that can be integrated in a description of a multilevel mechanism.

A final set of constraints on interlevel relations arises from interlevel experiments.[13] Interlevel experiments are designed to establish a component's relevance to the *explanandum*. Interlevel experiments use the techniques of different fields to intervene into and to detect the activities of mechanisms at different levels. Such experiments can be bottom-up or top-down. Bottom-up experiments are those in which one intervenes to perturb the activities at lower levels and to detect the effects of that intervention at higher levels. Examples include ablation studies (e.g., removing or inhibiting the hippocampus or NMDA receptor and detecting effects at higher levels) and stimulation studies (e.g., intervening to electrically or pharmacologically stimulate a neuron, receptor, or brain region and detecting the effects at higher levels). The case of H. M. is a poignant example of the former: one intervenes to remove the hippocampus and then detects the consequences of this intervention on memory. Wilder Penfield's classic stimulation studies are an example of the latter. Penfield delivered weak electrical stimuli directly to the cortex in awake and responsive patients during elective surgery. When he stimulated the temporal lobes (which house the hippocampus), he wrote that his patients reported vivid apparent memories (Penfield, 1952).

Top-down experiments, in contrast, involve intervening at a higher level and then detecting effects at lower levels. There are many examples of top-down experiments in the LM research program. For example, Wilson and McNaughton (1993) used a top-down experiment to argue that the hippocampus acts as a spatial map. After recording from 50–100 hippocampal neurons at once while the rat moved freely on a surface, they were then able to use that data to predict the location of the rat in its environment. Such top-down experiments contrast with Bickle's view of reduction (2003), which focuses exclusively on bottom-up experiments.

Interlevel experiments frequently span multiple levels and require the expertise of researchers from several different fields. In one such bottom-up experiment (McHugh et al., 1996), researchers deleted the gene for the NMDA receptor and tested for effects at a number of different levels. They found that mice without the NMDA receptor lost LTP induction, had distorted spatial maps, and had difficulty solving mazes. Experiments of this sort draw on molecular biology (for the deletion), electrophysiology (to assay for LTP), higher-level electrophysiology and computation (for detecting and modeling the spatial map), and psychology (to monitor LM performance), piecing their findings into a single multilevel theory.

To conclude, I have suggested that interlevel relationships in multilevel theories are mechanistic relations. These mechanistic structures embody constraints on interlevel

[13] See Craver (2002) for a more detailed discussion.

relations. Findings at different levels constrain the links between levels accommodatively, spatially, temporally, and experimentally. Fields are integrated across levels when different fields identify constraints on mechanisms at different levels. This gradual process of accumulating constraints on mechanisms at multiple levels in no way resembles efforts to translate one theory to another or to create a homomorphic image of one in terms of another. Instead, different fields elaborate the multilevel mechanistic scaffold with a patchwork of constraints on its organization, thereby revealing different hints as to how the mechanism can and cannot be organized. Progress in understanding interlevel relations and co-evolution is more likely to be achieved if philosophers of neuroscience leave reduction behind and focus instead on the mechanistic structures that scaffold the unity of the neurosciences.

8. Conclusion

Reductive models of interfield integration describe it as an abstract relationship between theories at different levels and as a relationship that involves establishing homomorphism between the reduced and reducing theories. According to the mechanistic model of interfield integration, fields are integrated by adding constraints on the organization of a mechanism. Given the centrality of the search for mechanisms in contemporary neuroscience and beyond, it seems fair to say that reduction has a viable competitor, at least in discussions of multifield integration in neuroscience. It remains to be seen whether this mechanistic approach is also a viable competitor to reduction in discussions of explanation and metaphysics.

Acknowledgements

William Bechtel, Lindley Darden, Ilya Farber, Kenneth Schaffner, and Alison Wylie provided helpful comments on earlier drafts. Kim Haddix and Phil Valko provided editorial assistance. Any errors are my own.

References

Andersen, P. O. (1960a). Interhippocampal impulses II. Apical dendritic activation of CA1 neurons. *Acta Physiologica Scandinavica, 48*, 178–208.

Andersen, P. O. (1960b). Interhippocampal impulses III. Basal dendritic activation of CA3 neurons. *Acta Physiologica Scandinavica, 48*, 209–230.

Andersen, P. O. (1991). LTP—an exciting and continuing saga. In M. Baudry, & J. Davis (Eds.), *Long term potentiation* (pp. xiii–xvii). Cambridge, MA: MIT Press.

Andersen, P. O. (2003). A prelude to long-term potentiation. *Philosophical Transactions of the Royal Society of London, Series B, 358*, 613–615.

Andersen, P. O., Burland, H., & Kaada, B. R. (1961). Activation of the dentate area by septal stimulation. *Acta Physiologica Scandinavica, 51*, 17–28.

Andersen, P. O., & Lømo, T. (1967). Control of hippocampal output by afferent volley frequency. *Progress in Brain Research, 27*, 400–412.

Bechtel, W. (Ed.). (1986). Introduction. In idem, *Integrating scientific disciplines*. Dordrecht: Martinus Nijhoff.

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

Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge, MA: MIT Press.

Bickle, J. (2003). *Philosophy of neuroscience: A ruthlessly reductive approach*. Dordrecht: Kluwer Press.

Bliss, T. V. P., & Gardner-Medwin, A. R. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology, 232*, 357–374.

Bliss, T. V. P., Gardner-Medwin, A. R., & Lømo, T. (1973). Synaptic plasticity in the hippocampal formation. In G. B. Ansell, & P. B. Bradley (Eds.), *Macromolecules and behavior* (pp. 193–203). London: Macmillan.

Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology, 232*, 331–356.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Clark, A. (1999). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Cragg, B. G., & Hamlyn, L. H. (1957). Some commissural and septal connexions of the hippocampus in the rabbit: A combined histological and electrical study. *Journal of Physiology, 135*, 460–485.

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

Craver, C. F. (2003). The making of a memory mechanism. *Journal of the History of Biology, 36*, 153–195.

Craver, C.F., (forthcoming). *Explaining the brain*.

Craver, C. F., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. K. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 112–137). Pittsburgh: University of Pittsburgh Press.

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

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 44*, 43–64.

Eccles, J. C. (1953). *The neurophysiological basis of mind*. Oxford: Clarenden Press.

Eccles, J. C. (1964). *The physiology of synapses*. New York: Academic Press.

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

Fodor, J. A. (1997). Special sciences: Still autonomous after all these years. In J. E. Tomberlin (Ed.), *Mind, causation and world, 1997* (pp. 149–163). Philosophical Perspectives, 11. Malden, MA; Oxford: Blackwell.

Frey, U., & Morris, R. G. M. (1998). Synaptic tagging: Implications for late maintenance of hippocampal long-term potentiation. *Trends in the Neurosciences, 21*, 181–188.

Galison, P. (1987). *How experiments end*. Chicago: University of Chicago Press.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71.

Gloor, P., Vera, C. L., & Sperti, L. (1964). Electrophysiological studies of hippocampal neurons III: Responses of hippocampal neurons to repetitive perforant path volleys. *Electroencephalography and Clinical Neurophysiolology, 17*, 353–370.

Green, J. D., & Adey, W. R. (1956). Neurophysiological studies of hippocampal connections and excitability. *Electroencephalography and Clinical Neurophysiology, 8*, 245–262.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hooker, C. A. (1981). Towards a general theory of reduction. *Dialogue, 20*, 38–59, 201–236, 496–529.

Keil, F., & Wilson, R. (2000). The shadows and shallows of explanation. In F. C. Keil, & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 87–114). Cambridge, MA: MIT Press.

Lashley, K. S. (2000). In search of the engram. In R. Cummins, & D. D. Cummins (Eds.), *Minds, brains, and computers* (pp. 333–350). Malden, MA: Blackwell (First published 1950).

Levitan, I. B., & Kaczmarek, L. K. (1991). *The neuron: Cell and molecular biology*. Oxford: Oxford University Press.

Lømo, T. (1966). Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. *Acta Physiologica Scandinavica, 68*(Suppl. 277), 128.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 57*, 1–25.

Maull, N. (1977). Unifying science without reduction. *Studies in History and Philosophy of Science, 8*, 143–162.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 7*, 115–133.

McHugh, T. J., Blum, K. I., Tsien, J. Z., Tonegawa, S., & Wilson, M. A. (1996). Impaired hippocampal representation of space in CA1-Specific NMDAR1 knockout mice. *Cell, 87*, 1339–1349.

Nagel, E. (1949). The meaning of reduction in the natural sciences. In R. Stauffer (Ed.), *Science and civilization* (pp. 97–135). Madison: University of Wisconsin Press.

Nickles, T. (1973). Two concepts of inter-theoretic reduction. *Journal of Philosophy, 70*, 181–201.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind–body problem* (pp. 3–36). Minnesota Studies in the Philosophy of Science, II. Minneapolis: University of Minnesota Press.

Penfield, W. (1952). Memory mechanisms. *Archives of Neurology and Psychiatry, 67*, 178–191.

Putnam, H. (1975). Philosophy and our mental life. In idem, *Philosophical papers, Vol. 2. Mind, language, and reality* (pp. 291–303). Cambridge: Cambridge University Press.

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

Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher, & W. C. Salmon (Eds.), *Scientific explanation* (pp. 3–219). Minnesota Studies in the Philosophy of Science, XVIII. Minneapolis: University of Minnesota Press.

Sarkar, S. (1992). Models of reduction and categories of reductionism. *Synthese, 91*, 167–194.

Schaffner, K. (1969). The Watson–Crick model and reductionism. *British Journal for the Philosophy of Science, 20*, 325–348.

Schaffner, K. F. (1993a). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.

Schaffner, K. F. (1993b). Theory structure, reduction, and disciplinary integration in biology. *Biology and Philosophy, 8*, 319–347.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*, 11–20.

Society for Neuroscience. (2003). www.sfn.org.

Squire, L. R., & Kandel, E. R. (2000). *Memory*. New York: Scientific American Library (First published 1999).

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

Wilson, M. A., & McNaughton, B. (1993). Dynamics of the hippocampal ensemble code for space. *Science, 261*, 1055–1058.

Wimsatt, W. C. (1976). Reduction, levels of organization and the mind body problem. In G. Globus (Ed.), *Consciousness and the brain* (pp. 205–267). New York: Plenum Press.

Wimsatt, W. C. (1984). Reductive explanation: A functional account. In E. Sober (Ed.), *Conceptual issues in evolutionary biology* (pp. 369–385). Cambridge, MA: MIT Press (First published 1976).

Wylie, A. (2003). *Thinking with things*. San Francisco: University of California Press.