Manuscript for inclusion in Bickle, J. (Editor),
Oxford Handbook of Philosophy and Neuroscience

Molecules, Systems, and Behavior: Another View of Memory Consolidation

William Bechtel
Department of Philosophy and Interdisciplinary
Programs in Cognitive Science and Science Studies

From its genesis in the 1960s, the focus of inquiry in neuroscience has been on the cellular and molecular processes underlying neural activity. In this pursuit neuroscience has been enormously successful. Like any successful scientific inquiry, initial successes have raised new questions that inspire ongoing research. While there is still much that is not known about the molecular processes in brains, a great deal of very important knowledge has been secured, especially in the last 50 years. It has also attracted the attention of a number of philosophers, some of whom have viewed it as evidence for a ruthlessly reductionistic program that will eventually explain how mental processes are performed in the brain in purely molecular terms. As neuroscience developed, however, there emerged a smaller group of researchers who focused on systems, behavioral, and cognitive neuroscience. These investigators have also made impressive advances in the last 50 years and they have been the focus of an even larger group of philosophers, who have appealed to systems level understanding of the brain as providing the appropriate point of connection to the information processing accounts advanced in psychology.

Both the philosophers who appeal to cellular and molecular neuroscience and those who focus on systems and behavioral neuroscience are, from the point of view of mental activity, reductionists. But their construal of reduction differs, and it is those differences that are the focus of this chapter. John Bickle (1998; 2003; 2006) has been the primary champion of reduction of mental function to cellular and molecular processes and he identifies his approach as *ruthless reduction*. The alternative approach, which appeals first to systems neuroscience to understand mental phenomena, I will characterize as *mechanistic reduction*. Both approaches, as I will show, recognize that much is to be learned by understanding the cellular and molecular processes in the brain. Where they differ is in their conception of what is required to explain mental phenomena. The mechanistic approach emphasizes the need to identify all (or at least the major) operating parts of the mechanism responsible for the phenomenon of interest and to understand the way they are organized and how their operations are orchestrated to realize the phenomenon. For most mental phenomena these working parts will be characterized in terms of brain regions or cell populations and it is their integration into networks that constitute the mechanisms responsible for the phenomena. The cellular and molecular processes targeted by the ruthless reductionist are typically not operating parts of these mechanisms, but are parts of their operating parts (or of the operating parts of the operating parts, etc.). They are themselves mechanisms at a lower-level of organization. Understanding them is extremely important for the project of understanding the mechanisms responsible for the cognitive behavior, but it is important to maintain a focus on the level at which they enter the explanatory account.

Focusing not on the appropriate philosophical analysis but on neuroscience itself, the contention of this chapter is that systems, behavioral, and cognitive neuroscience have a critical explanatory role to play. Whereas on the ruthless reductionist account, their role is primarily heuristic and preparatory for the ultimate account in terms of cellular and molecular processes,

the mechanistic reductionist holds that there are real causal processes at different levels of organization, and that appropriate research techniques are required to identify these causal processes and relate them in a mechanistic account. These tools are different from those developed in cellular and molecular neuroscience and require different skills for their application and interpretation of results. Such differences in research techniques and the explanations that researchers are seeking are often the basis for divisions into disciplines in science. In the last section of the paper I will explore the implications of this for the neurosciences, showing how the Society for Neuroscience is predominately focused on the cellular and molecular processes and that investigations of systems and behavioral processes are often on the fringe of neuroscience proper and frequently pursued more directly in parts of psychology and increasingly in distinct fields that have adopted the names *behavioral neuroscience* and *cognitive neuroscience*.

1. Memory Consolidation: From Behavior to Molecular Processes

To make the discussion of contrasting accounts of reduction concrete, I will focus on a particular mental phenomenon, memory consolidation. This is the phenomenon that for a period of time after an episode in which a new memory is encoded, the memory remains labile and is relatively easily disrupted. Some time later, after the processes referred to as consolidation are completed, the encoding acquires the robust and enduring characteristics of long-term memories. The phenomenon was already reported by Quintillian in the first century AD[1] (Dudai & Eisenberg, 2004) but it did not become the object of experimental investigation until the end of the 19th century with the research of Georg Elias Müller and Alfons Pilzecker in Göttingen (1900; see Lechner, Squire, & Byrne, 1999, for detailed analysis), who named the phenomenon *consolidation*. Only a decade earlier Hermann Ebbinghaus (1885) had pioneered a technique for studying the time course of memory in which he trained himself on lists of nonsense words and measured forgetting in terms of the number of retraining trials required before he could recite the list without error. Müller and Pilzecker adapted Ebbinghaus' procedures by testing subjects other than themselves and using syllable pairs, one of which would serve as the cue for the recall of the other. Their subjects reported a strong tendency for syllable pairs to come to mind between training trials despite efforts to suppress them. Müller and Pilzecker invoked the term *perseveration* for this phenomenon and appealed to it to explain a pattern of error they had observed—when subjects made incorrect responses, they often responded with other items that had been on the studied list, which may have become associated with the cue on subsequent rehearsals. This effect diminished the longer the interval between study and test. More significantly, Müller and Pilzecker proposed that perseveration served the function of helping establish representations of the syllables in memory and strengthen the connections between them, for which they used the term *consolidate*:

> The experience of everyday life shows that perseverative tendencies of different parts of a train of thought can be weakened considerably by turning one's attention with energy to a different matter. . . . One can question, however, whether the effect of such an intense mental occupation . . . , immediately following reading of a list, simply reduces the

[1] Dudai (2004, p. 52) quotes Quintillian as reporting the "curious fact . . . that the interval of a single night will greatly increase the strength of the memory," and as raising the possibility that " . . . the power of recollection . . . undergoes a process of ripening and maturing during the time which intervenes."

> frequency with which syllables of this list spontaneously regain conscious awareness. One might deem that the perseverative tendencies of syllables of a previously read list might also serve to consolidate the associations between these syllables . . . and that accordingly, weakening of the perseverative tendencies of syllables from a previously read list, due to other intense mental occupations, might have the additional effect of obstructing the associations between these syllables (Müller & Pilzecker, 1900, p. 68; translated in Lechner, Squire, & Byrne, 1999, p. 79).

In additional experiments they established that imposing another strenuous mental activity immediately after a learning and test episode would diminish later recall, but not if it was delayed even a few minutes.

William McDougall (1901) and William Burnham (1903) applied the results of Müller and Pilzecker to explain what Burnham referred to as *retroactive amnesia*, the loss of memory for the period immediately preceding shock or injury. Nearly 50 years later, Carl Duncan (1949) developed a procedure to create such amnesia experimentally using electroconvulsive shock in rats. As Lechner et al. note, explanations of retrograde inhibition in terms of inhibition of consolidation were generally superseded in psychological studies of memory by accounts in which subsequently encountered items were taken to interfere with the memory. Consolidation continued to figure in more biologically oriented studies of memory such as those inspired by Donald Hebb's (1949) account of gradual increases in cortical connections between neurons that spike together. Hebb's idea fostered an interest in a possible role for protein synthesis in establishing memories. Agranoff, Davis, and Brink (1966), for example, demonstrated that puromycin, a protein synthesis inhibitor, administered either just before or immediately after a training sequence in shock avoidance in goldfish had no effect on learning during a training session but eliminated all learning as measured four days later. This provided support for the hypothesis that protein synthesis is responsible for longer term memory formation but not short-term or working memory.

The path of research just described reflects a progression from behavioral level studies relating shock or injury to amnesia to the identification of a molecular agent that brings about the same effect. Before following this path further, however, let me note what became the classic behavioral level study of memory processes. In 1953 William Scoville performed a bilateral medial temporal lobe resection on a 27 year old patient, H.M., who had experienced epileptic seizures since he was 10, presumably as a result of an accident in which he was knocked down by a bicycle. The surgery was successful in reducing the seizures, but resulted in anterograde amnesia as well as temporally graded retrograde amnesia (i.e., memory loss is greater for more recently acquired memories than for ones acquired longer ago) for several years prior to the surgery (Scoville & Milner, 1957). The combination of graded retrospective amnesia with severe anterograde amnesia had been noted many decades earlier by Théodule-Armand Ribot (1882) and is expressed in Ribot's law ("Progressive destruction advances progressively from the unstable to the stable") and provided strong support for the idea that a process of consolidation was required to stabilize memories.

What H.M. made clear was that the medial temporal lobe, especially the hippocampus, plays an important role in the consolidation process. Research with H.M. also helped to differentiate types of long-term memory. While H.M. has no memories for either episodes in his life (episodic memory) since the accident or for information presented to him since then (semantic memory), he has been able to acquire new skills (albeit denying that he has ever

performed the skill prior to the test occurrence). This is taken to show that the hippocampus plays a critical role in encoding of episodic and semantic memories, but not in what is usually terms *procedural* or *implicit* memory (Squire & Knowlton, 1995).

At approximately the same time as H.M. led researchers to focus on the hippocampus as involved in encoding of long-term explicit memories, a different line of research resulted in identifying in the hippocampus the process of long-term potentiation (LTP), the persisting enhancement in the response of a post-synaptic cells to a input from a pre-synaptic cell when the post-synaptic cell has readily spiked after inputs from the pre-synaptic cells. Although it is often reported that Tim Bliss and Terje Lømo initiated research on LTP in 1973 in the wake of the memory deficits found in H.M. "to see whether the synapses between neurons in the hippocampus had the capability of storing information" (Squire & Kandel, 2000, p. 110), Carl Craver (2003) has shown that this line of research had a different origin. Already in the 1950s and 1960s investigators, who were frustrated that electrical responses from cells in hippocampal preparations soon diminished, found that responsiveness could be revived by supplying a brief tetanus. From this grew numerous reports of increased potentiation of cell response in various areas of the hippocampus after tetanus. Moreover, during this period there was no direct linkage between the hippocampus and memory (if there had been, Scoville would probably not have been willing to remove it in H.M.). Rather, as Craver describes, the hippocampus was associated with a host of mental phenomena, such as olfaction, emotion, and autonomic regulation, sleep and respiration, sexual behavior, etc. One of the major reasons to study the hippocampus was that it was thought to be involved in epileptic seizures, and the application of a tetanus to hippocampal cells was conceived as comparable to the repeated electrical stimulations thought to facilitate seizures. Another factor that Craver argues was particularly influential on Lømo's studying the hippocampus was that it was an anatomically well-characterized structure that was thought to provide a simpler model of cortical networks. It was thus well-suited for studying synaptic mechanisms. In particular, it was possible to study monosynaptic pathways and focus directly on processes occurring at a given synapse.

While the hippocampus was not viewed at the time as having a specific role in memory, researchers such as Per Andersen, Lømo's dissertation director, were very much interested in understanding learning, viewed as a process of neural plasticity. In an abstract for a conference of the Scandinavian Physiological Society, Lømo (1966, p. 128) reported an increase in "synaptic plasticity" that "may last for hours." Lømo was later joined by Bliss, and together they produced evidence for long enduring LTP (Bliss & Lømo, 1973), although as Craver notes, there was great variability in the phenomena they elicited and significant efforts in subsequent years were required to identify the conditions that regularly produced the phenomenon. One major aspect of this process was developing tissue slice preparations in which the relevant circuits could be isolated, electrodes inserted to stimulate the pre-synaptic cell and record from the post-synaptic cell, and investigated with a variety of probes, especially chemical ones.

Although they began their research for other reasons, in their 1973 paper Bliss and Lømo proposed that the cellular phenomenon of LTP was a process involved in learning and memory. Demonstrating and regularizing the phenomenon of LTP was only the first step, though, in developing a cellular and molecular account of learning and memory. The next step involved discovering the mechanism through which LTP occurred. Bliss and Lømo themselves, took a stand on what became one of the central issues concerning LTP—whether it involved changes in the excitability of the post-synaptic cell, changes in the pre-synaptic cell, or changes in synaptic efficacy. They adopted the latter view. Discovering the mechanism required more than

determining the locus of the phenomenon; it also required identifying the parts and operations of the mechanism and their organization. The major operations in a mechanism were discovered over the ensuing years and involve a number of molecules. In brief, when glutamate, the neurotransmitter that activates neurons exhibiting LTP, binds to N-methyl-D-aspartate (NMDA) receptors, they undergo a shape change that exposes pores in the membrane. These pores remain blocked unless the post-synaptic cell generates an action potential, in which case Ca^{++} ions flow into the cell and initiate a chemical cascade the alters the response properties of α-amino-3-hydroxy-5-methyl-4-isoxazole proprionic acid (AMPA) receptors, which also binds glutamate but in addition regulate the flow of Na^{+} and K^{+} ions that determine the voltage across the membrane of the post-synaptic cell. Beyond this immediate change in the post-synaptic cell, there is also a longer-term change that results in the synthesis of new proteins that alter such things as the dendritic spines on which receptors are lodged and the response properties of the AMPA receptors. The process of protein synthesis requires activation of DNA in the neuron's nucleus and the engagement of the protein synthesis machinery in the cytoplasm. Communication with the nucleus is realized by protein kinase A (PKA) that is released by the binding of cyclic adenosine monophosphate (cAMP) with cAMP-dependent-PKA molecules. In the nucleus PKA serves to phosphorylate cyclic AMP response element binding protein (CREB), which in turn activates genes that then initiate the synthesis of two proteins. One of these creates a positive feedback loop by destroying the regulatory subunits that bind with PKA to suppress its activity while the other facilitates an increase in the number of active receptors in post-synaptic dendrites, thereby producing a structural change that alters the response of the post-synaptic cell to neurotransmitters.

This knowledge of the chemical processes involved in LTP has enabled researchers to begin to identify the particular genes involved in the overall activity. Thus, Silva and Bickle (this volume) describe gene knock-out experiments in which mutant mice are created wherein activators of CREB are non-functional although the mice are otherwise normal. They showed that the mice which acquire normal associations between cues and shocks in the immediate setting (freezing when presented with the cue) that are evidenced in behavior when they are tested one hour later, fail to demonstrate learning when tested 24 hours later. Similar results were demonstrated in social recognition tasks and provide powerful evidence for the role of CREB in these behaviors.

2. Ruthless Reductionism

The path of research described in the previous section exhibits a progression from behavioral level characterization of memory consolidation to identification of important components in the process at progressively lower levels. For the ruthless reductionist, this represents the form of explanatory advance that marks important progress in science. Bickle (2006) characterizes the project of the ruthless reductionist with the aphorism "intervene cellularly/molecularly and track behaviorally" which he then explicates:

- intervene *causally* at the level of cellular activity or molecular pathways within specific neurons (e.g., via genetically engineered mutant animals, as in the case study described in the previous section);
- then track the effects of these interventions under controlled experimental conditions using behavioral protocols well accepted within experimental psychology. (p. 425).

The strong claim in Bickle's account of reduction is the appeal to intervening *molecularly*. He articulates the standard for success in terms of finding cellular or molecular components on

which intervention affects the phenomenon of interest: "One only claims a successful *explanation*, a successful *search for a cellular or molecular mechanism*, or a successful *reduction*, of a psychological kind when one successfully intervenes at the lower level and then measures a statistically significant behavioral difference" (p. 425).

Bickle contrasts his ruthless reduction approach to that of philosophers who have appealed to cognitive neuroscience as the appropriate locus in neuroscience for explaining cognitive functions. He focuses the difference on the question of whether one can drop multiple levels down in explaining cognitive function:

> Many philosophers will still wonder how current neuroscience proposes to step across so many "levels" in a single bound. Between the behavioral and the molecular-biological levels lie (sic) (at least) the cellular, the neuroanatomical, the circuit (neuron networks), the regional, the systems (including the motor system, to generate measurable behavior), and perhaps even the information-bearing and -processing. Must not reductive "bridges" be laid between all these intermediaries before we can claim "mind-to-molecular pathway reductions"? And is not *cognitive neuroscience*—the branch of the discipline that at least some philosophers can claim familiarity with—having enough trouble "bridging" the higher levels to warrant reasonable worries about whether neuroscience will ever pull off the entire reduction? (p. 412).

Appealing to examples such as that described above on memory consolidation, Bickle embraces the project of molecular neuroscientists to "bridge the behavioral to the molecular pathway levels *directly*" (p. 414).

3. Mechanisms and Mechanistic Reduction

A rather different account of reduction, one which takes very seriously the role of intervening levels that the ruthless reductionist disparages, results from focusing on what constitutes explanation in neuroscience. Whereas many philosophers of science, following the lead of the logical positivists in the early 20th century, have emphasized the role of laws and viewed scientific explanation as involving logical deductions of the phenomena to be explained from these laws, neuroscientists (and more generally biologists and psychologists), tend to appeal to mechanisms for explanation. Thus, to explain a phenomenon such as memory consolidation, they seek to describe the mechanism responsible for producing it. Recently a number of philosophers have sought to explicate the notion of mechanism to which these scientists have appealed and the way in which scientists develop and test models of mechanisms (Bechtel & Richardson, 1993; Glennan, 1996; 2002; Machamer, Darden, & Craver, 2000.) My preferred characterization is that a mechanism is "a structure performing a function in virtue of its components parts, component operations, and their organization" (Bechtel & Abrahamsen, 2005; Bechtel, 2006). In this account I differentiate parts and operations (Machamer, Darden, and Craver similarly differentiate entities and activities) insofar as parts are structurally characterized whereas operations are functionally identified as doing something (typically, altering themselves or something else in the process). Different research techniques are employed in decomposing a mechanism structurally into its component parts than are used to decompose it functionally into its component operations. Staining, for example, provides a means to identify neuron membranes, but electrodes are required to detect the electrical potential across the membrane. Ultimately, however, the parts of interests are those that perform the

operations (which, following Craver, I will refer to as *working parts*), and accordingly it is important to link operations with parts (which I refer to as *localization*).

Insofar as a mechanistic explanation requires decomposing the mechanism into its working parts, and these are, in a well delineated sense, at a lower level than the mechanism as a whole, mechanistic explanation is reductionistic. But, critically, the working parts into which the mechanism is decomposed are just one level lower than the mechanism as a whole. These are the parts that are organized and whose operations are orchestrated so as to realize the phenomenon in question. In many cases the working parts of a mechanism will themselves be mechanisms consisting of their own working parts. Accordingly, the process of mechanistic explanation can often be iterated and researchers may end up dealing with multiple levels of organization. But it is important to attend to what phenomenon is being explained at each level of decomposition. The second round of decomposition is concerned with explaining how the working parts are constituted and organized so as perform their operations.

From the mechanistic perspective, there is an important difference between intervening on a part, and intervening on a part of a part. In order to understand the mechanism responsible for the phenomenon of interest researchers must identify the various working parts of that mechanism and determine how they are organized to realize the phenomenon of interest. The working parts within these working parts are not themselves working parts of the first mechanism as they do not *directly* contribute to the phenomenon for which the first mechanism is responsible. They do so only as they are organized into the working parts of that mechanism.

This does not negate the importance of the parts of the parts and understanding what operations these subparts perform, but to make clear what the explanatory target is when focusing on these subparts—explaining how the parts of the larger mechanism perform their operations and not directly explaining the phenomenon realized in the larger mechanism. It is the whole organized part that plays a causal role in the top-level mechanism, not each of the sub-parts taken individually, which contribute to the whole mechanism only insofar as they enable the part to perform its operation.

One strategy for developing and testing accounts of mechanisms shares much with Bickle's characterization of ruthless reductionistic research. Both utilize the strategy of intervening causally within the responsible mechanism and detecting the effects of intervention in the behavior of the whole system. The difference lies in the ruthless reductionist's emphasis on intervening *cellularly*/*molecularly*; the mechanistic perspective in contrast focuses on intervening on whatever the components are that perform the operations that figure in the functioning of the mechanism. One version of this strategy for the mechanist involves removing or disabling a component from the mechanism to determine how the mechanism operates without the component, thereby potentially gaining insight into what the component contributes to the normal functioning of the mechanism. The other involves stimulating a component to determine how the hyper operation of the component affects the overall mechanism. (Yet other major strategy for the mechanist is to record, e.g., through single cell recording or functional neuroimaging, changes internal to the mechanism as it functions under various conditions. For discussion, see Bechtel, in press-a; in press-b.)

Having identified differences between ruthless reduction and mechanistic reduction, I turn to the import that difference has to the practice of neuroscience. There is no doubt that the research that Bickle cites as exemplary of ruthless reduction has provided important information about processes at work at some level of decomposition within the mechanism of memory consolidation. Moreover, information secured at such lower-levels often places serious

constraints on the accounts at higher levels. For example, if we learn that the subcomponents of a mechanism do not permit the component of the mechanism to perform the operation that is being assigned to it by a higher-level decomposition, then the higher-level decomposition must be revised. But there is a major risk inherent in the strategy of ruthless reduction. Insofar as it focuses exclusively on a sub-component (or a sub-sub-component) and relates it directly to behavior, it risks ignoring the other components of the mechanism and the organization which enables the components to work together to produce the phenomenon of interest. What is gained by paying attention to each level of organization identified in the course of decomposition is an understanding of the variety of roles that must be performed in producing the overall phenomenon.

Craver (2007) insists that a necessary condition on a good mechanistic explanation is that it account for the productive continuity between what he calls the start-up and termination conditions on the operation of a mechanism. Accounts that leave gaps in the characterization of the processes he counts only as sketches of mechanisms. While I would argue that complete productive continuity is more than is needed in practice for good explanations (as well as resist the imposition of linearity that results from specifying start-up and termination conditions), the idea of productive continuity helps focus our attention on the need not just to identify a single component but to identify at least many of the parts and operations so that it is possible to conceptualize how the operations couple together to realize the phenomenon. Often such understanding is realized in mechanistic research by simulating the operation of the whole mechanism, either mentally or using physical or computational models. Only by including the major operations in the account can a researcher simulate the overall production of the phenomenon.

The shortcoming of the ruthless reductionist's approach is that it focuses on only one or a few sub-components within the mechanism and fails to consider how those sub-components are related to other in the realization of the phenomenon in question. The process of new protein synthesis activated by CREB is, it appears, an important operation in the process of memory consolidation (it may even be the *crux point*). But it is not the whole process. Other components, both within the hippocampus and elsewhere in the brain, also play a role in memory consolidation. One serious shortcoming of ignoring these other factors is the propensity to ascribe too much to the factors that are considered. (The legacy of focusing on genes as responsible for traits, ignoring the variety of factors involves in the regulation and expression of genes, illustrates the risks involved. Researchers often claim success in explaining biological traits as soon as they identify a responsible gene, failing to recognize that often many factors, including other genes as well as environmental processes, are required to generate the trait.)

4. Memory Consolidation and Reconsolidation: From Molecular Processes Back to Systems

While great progress was being made in articulating the mechanisms of LTP, other advances were also being made with respect to the mechanisms involved in memory consolidation. Some of this research has been local to the hippocampus itself, although focusing not the molecular processes within cells but the neuroarchitecture of the hippocampus, which is distinctive and idiosyncratic. The hippocampal formation is organized as a loop involving a number of different regions. As shown in the schematic representation in Figure 1, information from neocortical association areas in temporal, parietal, and frontal cortex enters the hippocampal formation via the parahippocampal gyrus and the perirhinal cortex. From these areas it is projected to the entorhinal cortex (EC), from which it is processed through a loop,

ending up again in the entorhinal cortex, from which it is sent back out to the neocortical areas. The critical processing loop from the entorhinal cortex proceeds either directly on a pathway to the CA3 fields or via a route through the dentate gyrus (DG), which is unusual for having an unusually small number of neurons active in response to a given input. CA3's pyramidal cells are highly interconnected via recurrent connections (indicated by a small looping arrow); these cells also send activations forward to the pyramidal cells in CA1 via the small CA2 area (not shown). From CA1 activations are passed back to the EC directly or via an indirect pathway through the subiculum.

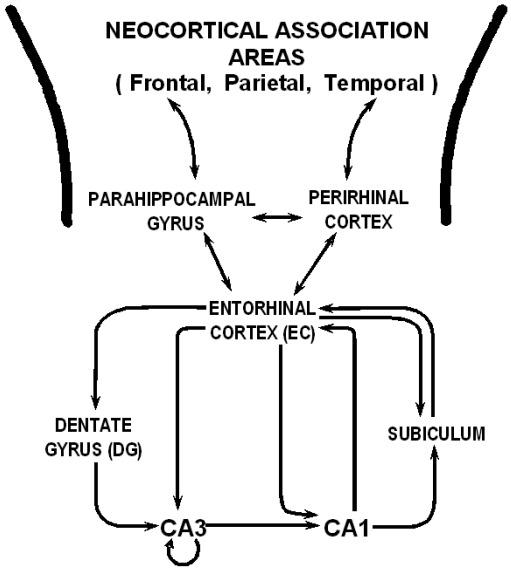

Figure 1. Schematic diagram of the hippocampal system. Information from widespread areas of neocortex converge on the parahippocampal region (parahippocampal gyrus, perirhinal cortex, and entorhinal cortex, EC) to be funneled into the processing loops of the hippocampal formation. The tightest loop runs from EC into the core areas of the hippocampus (CA1 and CA3) and back; the loop through the dentate gyrus and the recurrent connections in CA3 are also important; and the subiculum, which is not part of the hippocampus proper, provides an alternative return to EC. Not shown are a number of subcortical inputs and details of pathways and their synapses.

The distinctive architecture of the hippocampus has intrigued theorists as to its functional significance. The overall loop structure suggests that this may figure in the hippocampus's contribution to the ability, lost in H.M., to encode information over a several year period until a more permanent engram is laid down outside the hippocampus and as playing a role in laying down that engram. As Edmund Rolls (1995; Rolls & Treves, 1998) conceptualized the activity of the hippocampus, it needed to be able both to differentiate new events the organism encountered so it could develop new memories for them from events that were variations on previous events that would then be recalled. With these demands in mind, he developed a model with the same structure as the actual hippocampus of a rat but fewer units for each structure. Thus, he employed 1,000 units each in his model of DG and CA3, whereas in the rat there are approximately 10^6 granule cells in the DG and 300,000 in CA3. To capture the sparseness of the coding found in the DG-CA3 pathway, he allowed each artificial DG units to activate only 4 CA3 units. The interconnections between CA3 cells were modeled by recurrent connections through which CA3

cells provided input to themselves and other cells. In the pathway from CA3 to CA1 each CA3 unit sent a projection to 60 of the 1,000 CA1 units, and in the pathway from CA1 to EC each CA1 unit sent a projection to 200 of the 600 EC units.

Rolls employed his network to store 100 random patterns that were presented just once each as inputs to EC; storage was accomplished by adjusting weights between various units of the network by means of Hebbian learning. The architectural differences just noted affected how each part of the network functioned. Thus, after training, competitive learning in the pathways from EC to DG and from EC and CA3 to CA1 determined the activation patterns in DG and CA1; the pathways from EC to CA3 and from CA1 to EC functioned as pattern associators, whereas the very sparse connections from DG to CA3 served to differentiate patterns. After getting initial activation directly from EC and from EC via DG, CA3 ran 15 cycles on its recurrent connections to function as an autoassociative network. Activation then passed from CA3 to CA1 and from there back to EC, completing the processing loop.

Rolls measured the success of the model by its ability, when just part of each pattern was presented as input to the EC units, to regenerate the whole pattern on the same units (now as an output pattern). When the partial pattern was similar enough to the complete pattern—a correlation of just 0.40 was sufficient—the network could regenerate the complete pattern perfectly. In addition to obtaining this impressive overall result, Rolls was able to pin down some crucial design decisions. He already expected from the mathematical analysis and from the example of the rat hippocampus that varying the number of connections and sparseness would provide both separation and completion capacities. He also expected the recurrent connections in CA3 to be important, and demonstrated that turning them off eliminated CA3's ability to complete patterns (the pathways beyond it could partially but not completely compensate).

The functions that Rolls tries to account for at the level of the hippocampus itself are very different than those the ruthless reductionist localizes within individual neurons but are ones still critical to memory consolidation. The Hebbian learning that Rolls assumes in his model requires a mechanism such as the enhanced protein synthesis initiated by CREB, but the overall model captures features of memory consolidation not explained at the molecular level involving, as it does, communication between populations of neurons.[2]

Memory consolidation is not limited to activity within the hippocampus. McGaugh (2000), for example, examine the contribution of the amygdala to memory consolidation, concluding that it plays an important role in allowing emotional arousal to modulate memory strength. This is shown, for example, by the fact that β-Adrenergic receptor agonists infused into the basolateral nucleus of the amygdala enhance memory whereas antagonists block the effects of systematically applied dexamethasone, which usually enhances memory. Since inactivating the amygdala with lidocane infusions before retention tests does not impair enhanced memory, he concludes that it is not the locus of the memory. McGaugh concludes:

---

[2] In terms of its differentiation and categorization capacities Rolls' network exhibits some of the same capacities as a neural network developed by Lynch and Granger (1989) to model olfaction that was based on the anatomy of the olfactory bulb, piriform cortex and the entorhinal cortex and used a learning rule motivated by research on LTP. Bickle (1995) appeals to this model in support of his claim that there have already been successful reductions of psychological theories to purely neuroscientific theories. Without addressing the claim that the reducing theory is *purely* neuroscientific, I would note that the reduction does not simply involve cellular and molecular analysis (as his more recent ruthless reductionism advocates), but an explanation in terms of a network (mechanism) involving multiple interacting components.

> It is clear from these findings that memory consolidation involves interactions among neural systems, as well as cellular changes within specific systems, and that amygdala is critical for modulating consolidation in other brain regions (p. 249).

The fact that patients such as H.M. retain memories acquired several years before damage to their hippocampus indicates that the ultimate result of consolidation involves encoding of information in parts of the brain other than the hippocampus. Lashley's (1950) failure to identify areas in neocortex which would eradicate particular memories when lesioned led many to turn away from seeking memory encodings (*engrams*) in neocortex. The reemergence of neural network modeling in psychology in the 1980s suggested an explanation for the failure to find engrams—in neural network models information is often encoded in a very distributed fashion in which there is no single locus for the engram (accordingly, Rumelhart & McClelland, 1986, referred to such modeling as parallel distributed processing). These neural network models also employed a means of training such distributed representations through procedures involving gradual error reduction (Rumelhart, Hinton, & Williams, 1986a; 1986b). One consequence of such distributed representations is that they are subject to catastrophic interference as the learning of new information alters the connections that maintained the previously learned information (McCloskey & Cohen, 1989; Ratcliff, 1990). This was initially construed as a shortcoming of these models, but McClelland, McNaughton, and O'Reilly (1995) turned it to an advantage by suggesting that it explains why consolidation must be a drawn-out process. They proposed that by having the hippocampus serve to reinstate patterns repeatedly over time it was possible for the connections in neocortex to develop so as to acquire new information without losing information acquired previously. They developed a computational model to simulate the gradual training of the neocortex via the hippocampus and showed that when this training procedure was interrupted, their simulated cortical networks exhibited retrograde amnesia much like that found in human and animal research.

The proposed interaction of hippocampus and cortical areas in consolidating memories is not just a theoretical model. Research on sleeping rats by Jones Leonard, McNaughton, and Barnes (1987) showed that induction of hippocampal LTP is blocked during non-REM sleep while Bramham and Screbo (1989) demonstrated that it did not impair maintenance of previously induced LTP. This led Wilson and McNaughton (1994) to explore activity in the hippocampus during non-REM sleep. Using multi-electrode recording from CA1 cells they established that cells that fired synchronously during a maze-learning training episode also produced correlated firing during non-REM sleep.[3] Hoffman and McNaughton (2002) have gone on to show that during periods of quiet wakefulness following a sequential reaching task cells in posterior parietal, motor, and somatosensory cortices (but not in prefrontal cortex) that produced correlated or sequential firing during the task continued to exhibit correlated or sequential firing. The authors are cautious in claiming that this reactivation of coordinated firing patterns is part of

[3] While these researchers were investigating non-REM sleep, others such as Elizabeth Hennevin (see Hennevin, Hars, Maho, & Bloch, 1995) focused on REM sleep and established an increase in REM sleep following learning and amnesia when animals or humans are deprived of REM sleep after learning. György Buzsáki (1998) proposes that both REM and non-REM sleep are critical for memory formation: during REM sleep updated information is provided to CA3 from neocortical areas and bursting activity critical for synaptic plasticity and long-term memory consolidation is realized during non-REM sleep.

memory consolidation and how it relates to the synchronized activity in the hippocampus, but clearly anticipate a connection:

> At present, the mechanisms leading to the observed widespread memory trace reactivation remain unknown, and the necessity of coherent memory trace reactivation for memory consolidation remains to be demonstrated. Nevertheless, the observation that memory trace reactivation is temporally ordered and concurrent across large areas of the primate neocortex is a critical prerequisite for this process to function as a mechanism for memory consolidation (p. 2073).

Further evidence of a need to focus on the whole mechanism involved in consolidation, not just local operations, is found in the recently renewed interest in what is called *reconsolidation*. Memory consolidation has often been construed as a one-time event, occurring in a period of approximately 24 hours (often much shorter) immediately after learning. Already in 1968 two studies suggested that after a memory is recollected, it must be reconsolidated or it will be lost. Schneider and Sherman (1968) were addressing the question of whether electroconvulsive shock produced amnesia when applied immediately after a single trial in which rats were given a foot shock for stepping off a platform was due to the connection with the learning or the arousal created by the foot shock. They investigated this by applying the electroshock after a second foot shock applied 30 seconds later (while the rats remained on the shock grids). This still produced amnesia but the rats seemed to recover what they learned previously when tested a day later. The researchers then tried a six hour interval between the first and second foot shock (the second accompanied by electroconvulsive shock) and found that it produced amnesia for the learning that was associated with the first foot shock that was contingent on stepping off the platform. Moreover, the rats did not exhibit recovery when tested later. Misanin, Miller, and Lewis (1968) also demonstrated loss of memory when they applied electroconvulsive shock to rats immediately after presentation of a conditioned stimulus for which the rats had learned a response 24 hours earlier. These results initiated a period of investigation into amnesia produced after exposure to cues that might re-elicit the initial learning using not just electroconvulsive shock but also hypothermia and inhibition of protein synthesis to block memory consolidation. These studies seemed to indicate that the amnesia was dependent upon reactivating the memory traces and that reactivated memories were once again vulnerable and required consolidation (Lewis, Bregman, & Mahan, 1972). Other studies indicated that when memories were re-elicited it was also possible to enhance the memory by applying electrical stimulation to the mesencephalic reticular formation (MRF), which also served to enhance memory when applied during learning episodes (DeVietti, Conger, & Kirkpatrick, 1977).

This line of research was largely eclipsed as the prevailing interest in the 1980s and 1990s focused on the molecular mechanisms of LTP emphasized by ruthless reductionists but were rekindled in part as a result of a review article by Susan Sara (2000). She encountered the phenomenon of memory reconsolidation herself in the course of a study on the effects of a NMDA receptor antagonist on rats performing a well-learned maze task. The antagonist had no effect on the immediate trials, but when rats were tested again 24 hours later, they exhibited amnesia. In the period since Sara's review the phenomenon has again become the subject of multiple investigations. Many of the new studies applied protein synthesis inhibitors such as anisomycin to a range of species ranging from invertebrates to humans. When administered along with a retrieval task, usually a day after training involving punishment or aversive stimuli,

the protein inhibitor caused behavior no longer to be guided by the memory of the recalled item. Nader, Schafe, and LeDoux (2000), for example, trained rats to associate a tone with a foot shock and respond by immediate freezing. Microinfusion of anisomycin into the lateral and basal nuclei of the amygdala could block either consolidation if applied immediately after training, or immediately after retrieval, whether the retrieval was 1 or 14 days after conditioning. The investigators proposed the reconsolidation was required even for memories that had previously been well-consolidated.

Some have viewed reconsolidation as employing the same mechanism as is responsible for consolidation to begin with, but substantial differences have also been observed in the two processes. For example, neither the central amygdala nucleus, required for acquisition of conditioned taste aversion, nor the basolateral amygdala, required for extinction, is required for reconsolidation (Bahar, Dorfman, & Dudai, 2004). Likewise, the hippocampus is required for consolidation, but not reconsolidation of passive avoidance in the rat (Taubenfeld, Milekic, Monti, & Alberini, 2001). At present findings such as these are suggestive of differences between the mechanisms involved in consolidation and reconsolidation, but a full understanding of the parts and operations in either mechanism are still to be worked out.[4]

Findings such as those I have discussed in this section, many of which resulted from studies taking off from the research on protein synthesis in memory consolidation, point to the limitations of the ruthless reductionist approach of focusing solely on the molecular constituents and processes of protein synthesis in post-synaptic cells, even when tested in behaving systems. What these findings point to is that there are many additional working parts in the mechanisms responsible for memory consolidation. The molecular cascade which Bickle emphasizes does may play a role, indeed, an important role, in the overall process. But that role is situated within one of the constituent working parts of a much larger mechanism. Mechanistic reduction emphasizes the need to focus on identifying all the major components of the mechanism, determining their operations, and understanding how they are organized in the realization of memory consolidation. Ruthless reduction only identifies lower-level constituents of one component part of the mechanism.

5. Disciplinary Settings: Neuroscience versus Systems, Behavioral, and Cognitive Neuroscience

While the term *neuroscience* is often applied retrospectively to research on the brain that began to be pursued in earnest in the 19th century, its origins are in the 1960s and with a research program that focused on the physics and chemistry of the brain, especially its biophysics. The primary endeavor for which the term *neuroscience* was employed indeed reflects the project of ruthless reductionism, a narrowing of focus to the molecular processes within individual neurons. In this section I will begin by describing the disciplinary setting in which neuroscience itself developed, and then show how this project represented a radically narrowing of the research pursuits of inquiries into the structure and function of the brain that had preceded it. These broader aspirations are maintained in fields which append adjectives such as *systems*, *behavioral*,

[4] Bickle (2005) has himself appealed to the details of research on memory reconsolidation in responding to criticisms of ruthless reduction advanced by Looren de Jong and Schouten (2005) and has emphasized the role investigations at the cell and molecular level have played in recent research on memory reconsolidation. I agree with him that the interventions are often molecular in nature, but maintain that the challenge in understanding memory reconsolidation, as with memory consolidation itself, is to understand the full mechanism involved, not just the individual molecular components.

and *cognitive* to the core term *neuroscience*. It is to the fields named using these adjectives that one must look for the realization of the older investigations in neuroanatomy and neurophysiology and the promise of understanding how cognition is realized in the brain.

The primary inspiration for the development of the interdisciplinary field of neuroscience came from advances in biochemistry, biophysics, and molecular biology in the middle decades of the 20th century. The principal instigator for bringing these disciplines together was Francis Schmitt, who created what he termed the Neurosciences Research Program (NRP) in 1962. Schmitt's own background was in biophysics where he had contributed to deciphering the action potential in the giant squid axon as well as applying techniques such as electron microscopy to identifying structural components of neurons as well as muscle fibers. After moving to MIT in 1941 to head the Department of Biology and Biological Engineering, Schmitt established a large laboratory devoted to biophysical studies of collagen, muscle fibers, and neurons. Earlier in his career he had worked with Joseph Erlanger and Herbert Gasser at Washington University in St. Louis, who had pioneered techniques for recording from individual neurons and discovered that different neurons conducted impulses at different speeds (see Erlanger & Gasser, 1937, for an overview of their investigations). At MIT he attempted to understand the processes within neurons that gave rise to action potentials. While working at the Marine Biological Station in Viña del Mar, Chile, Schmitt encountered the large squid, *Dosidicus gigas*, with a giant axon ranging up to 4 mm diameter. He studied both the constitution of its axoplasm and the mechanism of the action potential, studies which he then continued at MIT using the smaller squid *Loligo pealii*, found off New England (Schmitt, 1959).

In 1955 Schmitt became highly involved with creating scientific institutions to support biophysics and with financial support from the Biophysical and Biophysical Chemistry Study Section of NIH he organized a four-week long summer program in 1958 entitled Intensive Study Program in Biophysics at the University of Colorado, Boulder. This brought together 200 biologists, chemists, physicists, psychologists, and engineers and culminated in the publication of *Biophysical Science: A Study Program* (Oncley, Schmitt, Williams, Rosenberg, & Bolt, 1959) that helped to delineate the scope and chart the research agenda for the new enterprise. Schmitt's own interest in nerve transmission led him to focus on the question of how chemical processes could operate sufficiently rapidly to support information retrieval. He was inspired by a meeting with Manfred Eigen to begin to consider ways in which elementary-charge carriers might be transported within neurofilaments and organized a symposium on the topic at M.I.T. in the spring of 1960 and another the following spring on macromolecules and memory function (Schmitt, 1962). From the experience with these two symposia, Schmitt began to formulate plans for an interdisciplinary institute focused on what he called *mental biophysics* or *biophysics of the mind* which would

> Investigate the "wet and dry" biophysics of central nervous system function, i.e., to study the physical basis of long-term memory, learning, and other components of conscious, cognitive behavior, by effective utilization of the biophysical and biochemical sciences, from the physical chemistry of neuronal and glial constituents (wet biophysics) through bioelectric studies (moist biophysics) to studies of fast transfer of element charged particles, organized microfields, stochastic models, and applications of computer science (dry biophysics).[5]

[5] From a 1962 grant proposal quoted by Swazey (1975, p. 532).

Schmitt identified nine basic disciplines comprising mental biophysics: solid-state physics, quantum chemistry, chemical physics, biochemistry, ultrastructure (electron microscopy and x-ray diffraction), molecular electronics, computer science, biomathematics, and literature research. Conspicuously absent from this list are neuroanatomy (above the ultrastructure level) and neurophysiology. Despite the inclusion of the word *mental* and the major developments that had transpired in psychology and linguistics in the previous decade (the advent of information processing models in psychology and of transformational grammars in linguistics), these disciplines were also absent.

By September 1961 Schmitt was referring to his project as the *Mens Project*, reflecting a growing interest in the traditional phenomena associated with mind. The range of disciplines expanded, but mostly in the direction of including disciplines such as neurophysiology, neuroanatomy, and neurology (psychology, though, did make it into the list of 25 fields that might be included in the project). When Schmitt convened an organizational meeting of eleven researchers in early 1962, most were from physics, biophysics, biochemistry, and molecular biology and one neurologist. The closes to psychology or neurophysiology was Heinrich Klüver, who was identified as representing biological psychology. Klüver was perhaps best known for his work with Paul Bucy that focused attention of the range of deficits that followed removal of the temporal lobe in monkeys. Klüver was led to this research as a result of his interest in determining whether removing the temporal lobe would block the hallucinations induced by mescaline. While the lesion had no affect on the organism's response to mescaline, it did impair the monkey's ability to identify objects by sight alone. Klüver and Bucy (1938) characterized this condition as *psychic blindness* or *visual agnosia*. Although these results played an important role in understanding temporal lobe function and provided the foundation for later research by Karl Pribram and Mortimer Mishkin, they hardly represented the mainstream of work in psychology, which by the 1960s was increasingly influenced by researchers emphasizing information processing accounts (Miller, Galanter, & Pribram, 1960).

On the basis of this meeting, Schmitt made a proposal to NIH and another to NASA to fund the *Neurosciences Research Program* (NPR). It is pretty clear that while the phenomena the NPR sought to consider were ones traditionally associated with psychology, the approaches it promoted were to be drawn from biophysics and other closely allied sciences. Figure 2, from the 1963 program report to the NIH, conveys the way the project was conceived: the first 27 Associates are situated within a triangle in which physics, chemistry, and biology form the points and the edges are labeled physical chemistry, biophysics, and biochemistry.

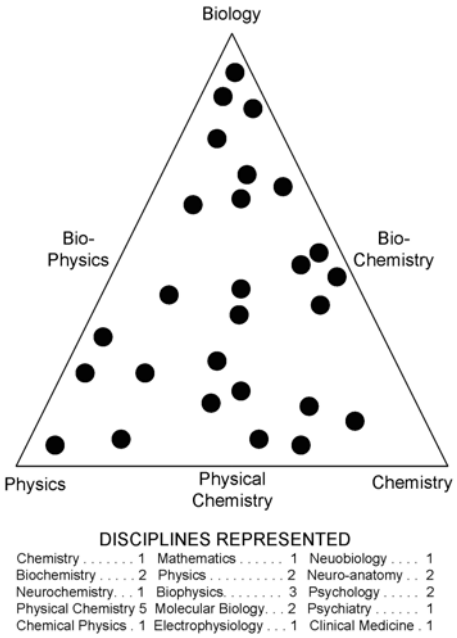

Figure 2. A chart depicting the distribution of base fields of Associates of the Neurosciences Research Program, redrawn from a 1963 Progress Report to the National Institutes of Health.

One of the main functions of the NPR was to host meetings that would foster exchange. Some of these were limited to associates; Schmitt described the first meeting of the associates: it "lasted for five days, and we were tutoring each other day and night." In addition, the NRP sponsored several three-day Work Sessions a year, in which senior scientists in the chosen fields each would educate others about his or her field. The chairperson for each of these Work Sessions was charged with writing a synthetic overview of the field and the discussions at the meeting, which was then published in *The Neurosciences Research Program Bulletin* (MIT Press also published an annual anthology of material from the Bulletin under the title *Neurosciences Research Symposium Summaries*). The Work Sessions for the first year give a flavor of the endeavors of the NRP:

Report Title	*Chair*
Cell Membranes	A. L. Lehninger
Information Storage and Processing in Biomolecular Systems	M. Eigen and L. C. M. DeMaeyer
From Brain Structure and Functions to Memory	W. J. H. Nauta
Glial Cells	R. Galambos

The NRP became a spur behind the creation of the Society for Neuroscience in 1970, which has grown in 35 years from 500 members to more than 30,000. Although the society does include researchers who work at the systems and behavioral level, the predominant emphasis remains on cellular and molecular neuroscience and, as Bickle has pointed out, it reflects the approach of the ruthless reductionist. Nonetheless, the term *neuroscience* (or sometimes *neurobiology*) is often taken to have a much broader scope, one that is more reflective of traditions of investigating the

brain that were largely set aside by the founders of the Society of Neuroscience. After briefly noting highlights of this history, I will focus on the recent continuation of this history in systems, behavior, and cognitive neuroscience.

The study of the structure and function of the brain has a long history, but the 19th century proved particularly important in development of neuroanatomy and neurophysiology at both the neuronal and whole brain levels as well as the development of neuropathology. The last part of the century witnessed a controversy over whether neurons were discrete cells separated by a gap between them or whether there was continuity between cells either in their axons or dendrites. This was resolved with the introduction of a silver nitrate stain by Golgi and its use by Cajal to demonstrate the individuality of neurons, a conclusion notorious rejected by Golgi (Mundale, 2001). The electrical properties of neurons were demonstrated earlier in the century by Emil du Bois-Reymond (1848-1884) and his student Julius Bernstein, who at the beginning of the 20th century advanced a proposal to explain electrical conduction in nerves as a result of the influx of potassium ions (Bernstein, 1912). One direction this research led was into the study of the mechanisms underlying the action potential and of LTP. Another was in the direction of recording from individual neurons and attempting to relate action potentials in neurons to external stimuli. This technique complemented approaches in the 19th century to studying brain regions either by analyzing deficits when areas were damaged or lesioned or by microstimulating neurons to determine the effects of their activity (Ferrier, 1876). By the early 20th century these techniques had revealed that many brain areas were laid out in topographical maps. Damage to selected regions revealed that different parts of the visual field are mapped onto different parts of the striate cortex (Holmes, 1918) whereas stimulation of areas near the central sulcus revealed the organization of somatosensory and motor cortex (Penfield & Boldrey, 1937).

As important as lesion and stimulation studies were, it was the ability to record from individual neurons while manipulating sensory inputs that provided the greatest advances in understanding the large-scale organization of the brain. Following upon Stephen Kuffler's (1952) discovery of retinal cells that responded to the contrast between the stimulation in the center of their receptive fields and that in the surround, Hubel and Wiesel (1962; 1968) initiated their studies of cells in LGN and Brodmann's area 17 (subsequently known as V1), in the latter of which they found cells that responded to bars of light. They also identified other areas with topographical maps of the visual field (which came to be known as V2 and V3) and other researchers, especially Semir Zeki (1973; 1974), identified still other maps in areas that responded selectively to the color or motion of the stimulus. By the 1990s 32 brain areas had been identified as principally involved in visual processing, taking up over 1/3 of the cortex in the macaque monkey (Felleman & van Essen, 1991; for a historical review of this research, see Bechtel, in press-b). Similar mapping of brain regions has been done for other sensory and motor areas and even in frontal cortex where there is less ability to correlate activity with sensory stimuli or motor responses (Carmichael & Price, 1996). These efforts at identifying brain regions responsive to differing types of information, and increasingly on identifying the connectivity patterns linking them, constitutes the project of systems neuroscience. While pursued by investigators who align themselves with neuroscience, system neuroscience is recognized even by them as at the periphery of neuroscience. Thus, van Essen and Gallant, in introducing their work on the organization and functioning of different visual areas in *Neuron*, a leading neuroscience journal, comment:

> These explorations involve mainly physiological and behavioral techniques that are quite different from the cellular and molecular techniques most familiar to this journal's readership. Nonetheless, we hope that a review of recent progress in understanding visual cortex will interest a broad spectrum of neuroscientists who share the ultimate objective of attaining a continuum of explanations of brain function, from the most molecular to the most cognitive levels (van Essen & Gallant, 1994, p. 1)

Historically, the term *behavioral neuroscience* reflects a development out of a tradition in psychology referred to as *physiological psychology*. Since physiological processes involved in behavior could often be studied most directly in non-human animals, physiological psychology was often linked with comparative psychology, and in 1940 the *Journal of Comparative Psychology* was renamed the *Journal of Comparative and Physiological Psychology* and became the flagship journal for psychological studies that emphasized the biological basis of behavior. By the 1980s comparative and physiological psychology went separate ways, and the journal split into two journals, one of which was named *Behavioral Neuroscience*. In an editorial announcing the split, Richard Thompson, who had been the editor of the combined journal and the first editor of *Behavioral Neuroscience*, commented on the scope of the field of behavioral neuroscience:

> Traditionally, physiological psychology emphasized certain approaches and techniques, particularly lesions, electrical stimulation, and electrophysiological recording. Study of the biological bases of behavior is now much broader and includes genetic factors, hormonal influences, neurotransmitter and chemical factors, neuroanatomical substrates, effects of drugs, developmental processes, and environmental factors, in addition to more traditional approaches. All these variables act ultimately through the nervous system . . . The contemporary meaning of the term 'behavioral neuroscience' is almost as broad as 'behavior' itself. Behavioral neuroscience is the field concerned with the biological substrates of behavior. Judging by its current rate of development, behavioral neuroscience could well become a dominant field of science in the future (Thompson, 1983, p. 3).

Cognitive neuroscience is the most recent of these three specialized domains of neuroscience, developing in the late 1980s. Part of the thrust for its creation stemmed from a perceived gulf between brain researchers and psychologists, which some researchers sought to overcome. In this spirit, brain researcher Joseph E. LeDoux and psychologist William Hirst edited a book, *Mind and Brain: Dialogues in Cognitive Neuroscience*, published in 1986, which drew together neuroscientists and psychologists to review the state of the art in their own discipline and respond to the corresponding reviews of the other on four topics: perception, attention, memory, and emotion. These reviews reveal both the emergence of a desire of some practitioners in each area to be able to draw upon the resources of the other, but also the large differences between the investigations pursued. As just one example of the differences, while most cognitive psychologists conducted their studies on adult humans, most neuroscience research was conducted on other species, where ethical constraints did not prevent insertion of electrodes into brains or making experimental lesions.

The term *cognitive neuroscience* was introduced somewhat earlier when Michael Gazzaniga established a Cognitive Neuroscience Institute with $500,000 provided by the Alfred P. Sloan Foundation. The grant was awarded to Gazzaniga and George Miller to attempt to draw

inferences about cognitive function in normal individuals from deficits that followed brain injury. It was the only grant with a neural component in the ten year, $17.4 million initiative the Sloan Foundation began in 1977 to support cognitive science, itself a fledging interdisciplinary endeavor involving principally psychology, linguistics, and artificial intelligence.[6] In 1986 the James S. McDonnell Foundation announced an initiative "to develop some specific programs to support research linking the biological with the behavioral sciences" and one month later Gazzaniga submitted a proposal to fund cognitive neuroscience. The foundation put together a panel to explore how to proceed (in addition to Miller and Gazzaniga, the panel included Marcus Raichle, Michael Posner, Terry Sejnowski, Gordon Shepherd, Emilio Bizzi, and Steven Hillyard; only Shepherd reflected the cellular and molecular focus typical in neuroscience). Concluding that there were "too many open questions, theoretical tangles and potential misunderstandings separating the two critical specialties—neuroscience and psychology—to proceed immediately," the panel elected first to constitute study groups. By 1988 the Foundation was ready to fund a summer institute in cognitive neuroscience and in 1989 it began a ten-year collaboration with the Pew Charitable Trusts to fund centers and individual research in cognitive neuroscience. In 1989 Gazzaniga established the *Journal of Cognitive Neuroscience* and in 1994 co-founded the Cognitive Neuroscience Society, providing institutional identity to the new initiative. It is noteworthy that, in contrast to the Society for Neuroscience, with over 37,500 members, the Cognitive Neuroscience Society currently has approximately 2000 members.

Above I identified a growing interest in brain researchers and psychologists in integrating their results, but what most advanced the endeavor of cognitive neuroscience was the development of a new research technique that permitted linking cognitive operations to brain regions. This involved on the one hand non-invasive means of measuring brain activity (actually, blood flow, which is assumed to be linked to brain activity) through either positron emission tomography (PET) or magnetic resonance imaging (MRI) and on the other hand ways of relating brain activity to operations in cognitive tasks. The most influential of the strategies for relating brain activity to cognitive tasks was provided by Michael Posner, a cognitive psychologist hired as part of a requirement in a very large grant from the McDonnell Foundation to Washington University in St. Louis for the study of higher brain function, where Marcus Raichle was developing PET to study blood flow in the brain. Posner adapted a technique first introduced by the 19th century Dutch psychologists Cornelius Donders for using the difference in time required to perform two cognitive tasks differing in that one required an additional operation to determine the time required for that operation. Posner proposed measuring not increased time but increased blood flow between two tasks differing in one operation. In a pioneering study, Petersen, Fox, Posner, Mintun, and Raichle (1988) subtracted blood flow measured when subjects read nouns aloud from that measured when they generated a verb related to the noun and spoke that verb.

---

[6] Although the roots of the collaboration between these fields go back at least to a September 1956 Symposium on Information Theory held at MIT in which Chomsky presented his work on transformation grammars, Newell and Simon introduced the first AI program, Logic Theorist, and George Miller identified informational limitations on cognitive processes involving the magic number 7, the term *cognitive science* itself first appears in two books published in 1975 (Bobrow & Collins, 1975; Norman & Rumelhart, 1975). The Sloan initiative was a major factor in developing cognitive science, fostering the creation of interdisciplinary programs at many institutions. One of these institutions, UCSD, used part of its grant to sponsor a conference, the LaJolla Conference on Cognitive Science, in 1979, which became the founding conference of the Cognitive Science Society. Two years earlier a new journal, *Cognitive Science*, was founded and subsequently was published by the Cognitive Science Society (for an account of the history of cognitive science, see Bechtel, Abrahamsen, & Graham, 1998).

The regions of increased blood flow, left dorsolateral prefrontal cortex, anterior cingulate, and right inferior lateral cerebellum, were taken to be involved in semantic processing.

In many ways, cognitive neuroscience can be viewed as the human counterpart of systems and behavioral neuroscience and all three are counterpoised to cellular and molecular neuroscience. Their focus is not on individual neurons and the molecular processes within them, but on the interactions of multiple brain regions. Sometimes it appears that the main interest in these endeavors is to determine where in the brain cognitive activities occur, but in fact in most research the goal is to identify the various parts of the brain that are active in a given task, what each is contributing, and how these areas are organized so as to coordinate their operations in the performance of an overall task. In other words, they are engaged in identifying the mechanisms responsible for such tasks as identifying an object, detecting and responding to motion, processing language, encoding or retrieving memories, etc. Researchers pursuing such research are fully aware that the brain regions they are studying are comprised of neurons and that processes such as the generation of action potentials and the altering of communication at synapses are crucial to the behavior of the brain regions. For systems and cognitive neuroscientists, however, those activities are at a lower-level of organization; operations at that lower level can explain how the components they study perform their operations, but not how those operations together perform the overall cognitive task, which is their objective.

6. Conclusions

The two positions I have discussed in this paper, ruthless reduction and mechanistic reduction, are both reductionist in that both recognize the importance of seeking knowledge of brain processes at different levels of organization in order to understand cognitive function. They are united in standing opposed to the attempts to divorce psychology and cognitive science from being constrained by our rapidly growing knowledge of brain processes. The principled arguments some philosophers have offered for psychology and cognitive science developing explanations ignoring the brain sciences are, on both views, a prescription for disaster.

The two accounts also agree that information about molecular and cellular processes, such as information about the molecular cascade involved in LTP and the genetic processes involved in synthesizing new proteins is also of potentially great relevance to understanding memory consolidation. Bickle quote approvingly the following passage from Kandel, Schwartz, and Jessell's (2000, pp. 3-4) textbook, the leading textbook in neuroscience:

> This book . . . describes how neural science is attempting to link molecules to mind—how proteins responsible for the activities of individual nerve cells are related to the complexities of neural processes. Today it is possible to link the molecular dynamics of individual nerve cells to representations of perceptual and motor acts in the brain and to relate these internal mechanisms to observable behavior.

The language of linkages employed by Kandel et al. in this passage is instructive. Both ruthless reduction and mechanistic reduction accommodate linkages. Where they part company is on Bickle's gloss of this passage: "These "links" are nothing less than *reductions* of psychological concepts and kinds to molecular-biological mechanisms and pathways." The molecular processes in cells to which Bickle proposes to reduce psychological concepts such as memory consolidation are, on the mechanistic reductionists account, processes within operating components of the relevant mechanism. They do not, on their own, explain the phenomenon of

memory consolidation. Such an explanation requires identifying the full range of brain areas involved in memory consolidation and the operations each performs. The processes to which the ruthless reductionist appeals only figure in the subsequent attempts to explain these operating parts.

Although I have argued that Bickle is wrong in defending the ruthless reductionist approach to explanation, he has, I have indicated, correctly captured the ethos in mainstream neuroscience, as represented by the Society for Neuroscience. Since its inception in the 1960s, it has represented an attempt to explain mental phenomena at the cell and molecular level. As a result, systems neuroscience has only found a home at the fringe of neuroscience, and behavioral and cognitive neuroscience are pursued in enterprises outside the scope of neuroscience as characterized by the Society for Neuroscience. But it is to these projects that researchers are increasingly looking for the explanation of phenomena such as memory consolidation. The cellular and molecular accounts are parts of a subordinate (but certainly not unimportant) explanation of how the parts of the mechanism operate.

References

Agranoff, B. W., Davis, R. E., & Brink, J. J. (1966). Chemical studies on memory fixation in goldfish. *Brain Research, 1*, 303-309.

Bahar, A., Dorfman, N., & Dudai, Y. (2004). Amygdalar circuits required for either consolidation or extinction of taste aversion memory are not required for reconsolidation. *European Journal of Neuroscience, 19*, 1115-1118.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (in press-a). The epistemology of evidence in cognitive neuroscience. In R. Skipper, C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson & R. C. Richardson (Eds.), *Philosophy and the Life Sciences: A Reader*. Cambridge, MA: MIT Press.

Bechtel, W. (in press-b). *Mental mechanisms: Philosophical perspectives on the sciences of cognition and the brain*. London: Routledge.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421-441.

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

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

Bernstein, J. (1912). *Elektrobiologie. Die Lehre von den elektrischen Vorgängen im Organismus auf moderner Grundlage dargestellt* Braunschweig: Vieweg & Sohn.

Bickle, J. (1995). Psychoneural reduction of the genuinely cognitive: Some accomplished facts. *Philosophical Psychology, 8*, 265-285.

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

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer.

Bickle, J. (2005). Molecular neuroscience to my rescue (again): Reply to Looren de Jong & Schouten. *Philosophical Psychology, 18*, 487-494.

Bickle, J. (2006). Reducing mind to molecular pathways: explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese, 151*(411-434).

Bliss, T. V. P., & Lømo, T. (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*, 331-356.

Bobrow, D. G., & Collins, A. F. (1975). *Representation and understanding: Studies in cognitive science*. New York: Academic Press.

Bramham, C. R., & Srebo, B. (1989). Synaptic plasticity in the hippocampus is modulated by behavioral state. *Brain Research, 493*, 74-86.

Burnham, W. H. (1903). Retroactive amnesia: Illustrative cases and a tentative explanation. *American Journal of Psychology, 14*, 382-396.

Buzsáki, G. (1998). Memory consolidation during sleep: A neurophysiological perspective. *Journal of Sleep Research, 7*, 17-23.

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology, 371*, 179-207.

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

Craver, C. (2007). *Explaining the brain: What a science of the mind-brain could be*. New York: Oxford University Press.

DeVietti, T. L., Conger, G. L., & Kirkpatrick, B. R. (1977). Comparison of the enhancement gradients of retention obtained with stimulation of the mesencephalic reticular formation after training or memory reactivation. *Physiology and Behavior, 19*, 549-554.

Du Bois-Reymond, E. (1848-1884). *Untersuchungen über thierische Elektricität*. Berlin: Reimer.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55*, 51-86.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron, 44*, 93-100.

Duncan, C. P. (1949). The retroactive effect of electroshock on learning. *Journal of Comparative and Physiological Psychology, 42*, 32-44.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker & Humblot.

Erlanger, J., & Gasser, H. S. (1937). *Electrical signs of nervous activity*. Philadelphia: University of Pennsylvania Press.

Felleman, D. J., & van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*, 1-47.

Ferrier, D. (1876). *The functions of the brain*. London: Smith, Elder, and Company.

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

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

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

Hennevin, E., Hars, B., Maho, C., & Bloch, V. (1995). Processing of learned information in paradoxical sleep: relevance for memory. *Behavioural Brain Research, 69*, 125-135.

Hoffman, K. L., & McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science, 297*, 2070-2073.

Holmes, G. M. (1918). Disturbances of visual orientation. *The British Journal of Ophthalmology, 2*, 449-468.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106-154.
Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215-243.
Jones Leonard, B., McNaughton, B. L., & Barnes, C. A. (1987). Suppression of hippocampal synaptic activity during slow-wave sleep. *Brain Research, 425*, 174-177.
Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neuroscience* (4th ed.). New York: McGraw-Hill.
Klüver, H., & Bucy, P. (1938). An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness". *The Journal of Psychology, 5*, 33-54.
Kuffler, S. W. (1952). Neurons in the retina: Organization, inhibition and excitatory problems. *Cold Spring Harbor Symposia on Quantitative Biology, 17*, 281-292.
Lashley, K. S. (1950). In search of the engram. *Symposia of the Society for Experimental Biology, IV. Physiological Mechanisms in Animal Behaviour*, 454-482.
Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation--remembering Müller and Pilzecker. *Learning and Memory, 6*, 77-87.
LeDoux, J. E., & Hirst, W. (Eds.). (1986). *Mind and brain: Dialogues in cognitive neuroscience*. Cambridge: Cambridge University Press.
Lewis, D. J., Bregman, N. J., & Mahan, J. (1972). Cue-dependent amnesia in rats. *Journal of Comparative and Physiological Psychology, 81*, 243-247.
Lømo, T. (1966). Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. *Acta Physiologica Scandinavica, 68*, 128.
Looren de Jong, H., & Schouten, M. (2005). Ruthless reductionism: A review essay of John Bickle's Philosophy and neuroscience: A ruthlessly reductive account *Philosophical Psychology, 18*, 473-486.
Lynch, G., & Granger, R. (1989). Simulation and analysis of a simple cortical network. *Psychology of Learning and Motivation, 23*, 205-241.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.
McClelland, J. L., McNaughton, B., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419-457.
McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109-165). New York: Academic.
McDougall, W. (1901). Experimentelle Beitraege zur Lehre vom Gedaechtnis: von G.E. Mueller und A. Pilzecker. *Mind 10*, 388-394.
McGaugh, J. L. (2000). Memory--A century of consolidation. *Science, 287*, 248--251.
Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt.
Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of consolidated memory trace. *Science, 160*, 554-555.

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane. Ergänzungsband, 1*, 1-300.

Mundale, J. (2001). Neuroanatomical foundations of cognition: Connecting the neuronal level with the study of higher brain areas. In W. Bechtel, P. Mandik, J. Mundale & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences*. Oxford: Blackwell.

Nader, K., Schafe, G. E., & LeDoux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature 406*, 722-726.

Norman, D. A., & Rumelhart, D. E. (1975). *Explorations in cognition*. San Francisco: Freeman.

Oncley, J., Schmitt, F. O., Williams, R. C., Rosenberg, M. D., & Bolt, R. H. (Eds.). (1959). *Biophysical science: A study program*. New York: Wiley.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain, 60*, 389-443.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(18 February), 585-588.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review, 97*, 285-308.

Ribot, T.-A. (1882). *Diseases of memory: An essay in the positive psychology* (W. H. Smith, Trans.). New York: Appelton-Century-Crofts.

Rolls, E. T. (1995). A model of the operation of the hippocampus and entorhinal cortex in memory. *International Journal of Neural Systems, 6, supplement*, 51-70.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function*. Oxford: Oxford University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature, 323*, 533-536.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Sara, S. J. (2000). Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learning and Memory, 7*, 73-84.

Schmitt, F. O. (1959). Molecular organization of the nerve fiber. *Reviews of Modern Physics, 31*, 455-465.

Schmitt, F. O. (Ed.). (1962). *Macromolecular specificity and biological memory*. Cambridge, MA: MIT Press.

Schneider, A. M., & Sherman, W. (1968). Amnesia: a function of the temporal relation of foot-shock to electroconvulsive shock. *Science, 159*, 219-221.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*, 11-21.

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

Squire, L. R., & Knowlton, B. J. (1995). Memory, hippocampus, and brain systems. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 825-837). Cambridge, MA: MIT Press.

Swazey, J. P. (1975). Forging a neuroscience community: A brief history of the neurosciences research program. In F. G. Worden, J. P. Swazey & G. Adelman (Eds.), *The neurosciences: Paths of discovery* (pp. 529-546). Cambridge, MA: MIT Press.

Taubenfeld, S. M., Milekic, M. H., Monti, B., & Alberini, C. M. (2001). The consolidation of new but not reactivated memory requires hippocampal C/EBP$_{\beta}$. *Nature Reviews Neuroscience, 4*, 813-818.

Thompson, R. F. (1983). Editorial. *Behavioral Neuroscience, 97*, 3.

van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*, 1-10.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science, 265*, 676-679.

Zeki, S. M. (1973). Colour coding of the rhesus monkey prestriate cortex. *Brain Research, 53*, 422-427.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology, 236*, 549-573.