MAN AS MACHINE: A REVIEW OF *MEMORY AND THE COMPUTATIONAL BRAIN: WHY COGNITIVE SCIENCE WILL TRANSFORM NEUROSCIENCE* BY C. R. GALLISTEL AND A. P. KING

John W. Donahoe
University of Massachusetts at Amherst

Some years ago I was pleased to attend a colloquium at my university presented by the first author, the psychologist Randy Gallistel. After describing some behavioral studies with rats, he concluded that the animals had *counted* their responses. Following the talk, I ventured that surely he was speaking metaphorically when he said that the rats were counting. To my surprise, he assured me that he was not; he meant it literally! The book under review, co-authored with the computer scientist Adam King, describes the thinking behind this and related assertions about the behavioral and neural processes involved in complex behavior, especially memory. In the authors' words:

> There must be an addressable read/write memory mechanism in brains that encodes information received by the brain into symbols (write), locates the information received by the brain (addresses), and transports it to computational machinery that makes productive use of the information (reads). (p. viii)

In short, living organisms are made in the image of a digital computer of the sort described by John von Neumann and envisioned by Alan Turing. Note that the claim is not that behavior *may* be so conceived, but that it *must* be:

> If one believes that the brain is an organ of computation—and we take that to be the core belief of cognitive scientists—then to understand the brain one must understand computation and how it may be physically implemented. (p. 125)

After reading the book, I now better understand the authors' theoretical position but continue to find their arguments unpersuasive. This review contrasts their views with a biobehavioral approach that is informed by experimental analyses of behavior and neuroscience. The authors and the reviewer agree on at least one important point: "All contemporary attempts to spell out how computations are realized in the brain are speculative" (p. 170).

Before moving to the specific content of the book and examining the reasoning behind the claims, a few preliminary comments are called for. First, the book is a "point-of-view" book—as are, unavoidably, all books to some extent.

AUTHOR'S NOTE: Please address correspondence to the author at:
jdonahoe@psych.umass.edu

This one is avowedly and explicitly so. As a result, findings and interpretations that provide alternative accounts of complex behavior in terms of well-understood simpler processes are typically given short shrift. In their hands, Occam's razor has become a blunt instrument. However, because the authors are too knowledgeable (and clever) to ignore competing formulations altogether, the alternatives are sometimes mentioned *en passant* before being dismissed. Second, the review divides the book into two main sections. Section one (Chapters 1 through 9) is largely conceptual in nature with empirical content introduced primarily to illustrate the conceptual points. Section two (Chapters 10 through 16) is largely empirical and intended to exemplify and support the earlier conceptual claims. Finally, I would note that the prose is eminently clear and carefully written throughout.

The Conceptual Foundations

The book opens with a commitment to an information-processing view of behavior/brain science, a view espoused earlier in cognitive science (Atkinson & Shiffrin, 1968).[1] Organisms are said to have access to "information" contained in the neural inputs initiated by stimuli, where information is used in the technical sense of Claude Shannon's information theory. On this view, information is a *change* in uncertainty when an estimate of the probability of an event following an observation is compared to its probability before the observation. To the extent that the current observation changes that probability, the observation is informative. To apply the information metric, two sets of values are needed—all possible states and the probabilities of each state prior to the observation. As an example, the states might be whether a light or a tone occurred in a Pavlovian conditioning experiment. Bayes' inverse-probability theorem provides a measure of the probability of the state following the current observation. Some fifty years ago, psychologists thought that this formulation showed promise as a general approach to behavior (e.g., Attneave, 1959; Bakan, 1953). In fact, the relation between information theory and Bayes' theorem was included in a graduate course in experimental design that I taught in the 1960s. However, it became apparent that although this formulation might be useful for guiding the behavior of theorists (cf. Chamberlin, 1880; Platt, 1964), it was inadequate for guiding the ongoing behavior of organisms. When a state is encountered for the first time, neither its prior existence (by definition) nor its *a priori* probability could be known. The second deficiency may be tolerated because it can be shown that, over repeated occurrences in at least simpler cases, the probability estimate of a state will converge to the population value regardless of its initial value (LaPlace's principle of insufficient reason). The first deficiency is fatal because it presupposes that the organism has foreknowledge of all the states that could ever occur. The authors appreciate this difficulty: "It. . .implies that the receiver has a representation of the

---

[1] This review follows the practice of citing only references that were not included in the book.

source probabilities, which is itself a controversial assumption in behavioral neuroscience and cognitive psychology" (p. 10). Nevertheless, the authors take the position that "communication between the brain and the world is only possible, in a rigorous sense, if the brain is assumed to have a representation of possible states of the world and their probabilities" (p. x). In communication theory, the possible states are known to the theorist (e.g., the letters of the alphabet) but all the events that an organism may ever encounter in its lifetime are unknowable. The authors make an effort to address the problem by considering a state as the presence or absence of a nerve impulse (or of a particular interval between nerve impulses) in a given neuron, but this presupposes that such events constitute the critical information. On this point, the authors concede: "Generally speaking, neuroscientists do not know what it is about trains of action potentials that carries the information, nor exactly what information is being communicated" (p. 21). In spite of this uncertainty, they proceed heedless of Sherlock Holmes's admonition, "It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts." Given the centrality of information theory in their formulation, it is surprising that it does not resurface in any serious fashion until much later in the book (Chapter 13) where the events are external stimuli, not neural impulses.

The remaining chapters in this section of the review reflect the authors' (pre)conception that organisms are information-processing machines in the mode of a digital computer: "Our hunch is that information transmission and processing in the brain is. . .essentially digital" (p. 24). If this "hunch" is correct, then the behavioral/neural capabilities of organisms are appropriately constrained by principles that have been developed in computer science. These principles are explicated in Chapters 3 through 9.

Among these principles is that the inputs to the processor must be in the proper form for use by the computational system. An especially egregious nonbehavioral example of a failure to comply with this computing principle was the 1998 demise of the $125 million Mars orbiter, whose inputs to the processor were in English units instead of International Standard units (pounds force instead of newtons), which the processor required. An important corollary of this principle is that, because different events may be coded in different ways by their inputs, different processors are required for these inputs. As stated later in *Memory and the Computational Brain* (*MCB*), this implication leads to the conclusion that there can be no single general processing system:

> For computational reasons learning mechanisms must have a problem-specific structure, because the structure of a learning mechanism must reflect the computational aspects of the problem of extracting a given kind of representation from a given kind of data. Learning mechanisms are problem-specific modules—organs of learning. (p. 219)

At this point the linguist Noam Chomsky is cited with approval.

As the preceding comments indicate, events in the world are said to have "representations" within the information-processing system: "The entities in the

brain that represent possible messages are symbols" (p. 55) and "symbols. . .are physical entities in a physically realized representational system" (p. 72). In computational work in artificial intelligence, symbols have traditionally been thought to be important to simulate complex behavior (Newell & Simon, 1976), although this assumption is now often questioned (Brooks, 1990; Dawson, 1998). If a symbol merely indicates that different events in the world produce distinguishable physical events within the nervous system and that these intra-organismic events lead to other physical consequences within the nervous system and ultimately to observable behavior, then they are at worst conceptually superfluous. They function as shorthand terms for what are undoubtedly complex neural sequelae. If a symbol refers a stable state stored in a specific location in the nervous system from which it may be retrieved, then the formulation becomes problematic. No one doubts that experienced organisms can behave such that a single behavior (e.g., saying "dog") may be guided by any one of many inputs (e.g., seeing chihuahuas or mastiffs); this is the province of what is conventionally called concept formation. However, it is a different matter to claim that there resides in the nervous system a stable neuronal array whose pattern of activity constitutes the concept and the output of which is transmitted to a processor where it is integrated with other symbols to guide behavior.

Inferences based on behavioral observations about the cognitive processes that are said to underlie behavior have proven notoriously difficult to sustain (Minsky & Papert, 1988). These include such distinctions as between propositional and imaginal coding (Anderson, 1978) and between serial and parallel processing (Townsend, 1972). Even in the most thoroughly studied form of human behavior, verbal behavior, the fundamental construct of a lexicon has encountered difficulties. For example, a person who has suffered damage to a particular region of the brain may be able to appropriately use the words "up" and "down" to describe emotional states but not to indicate the locations of objects (see Donahoe, 1991 for this and other examples). Are "up" and "down" stored as symbols in a lexicon, or not? More generally, are there representations of these words? A seemingly attractive tactic to preserve the notion of a lexicon is to claim that polysemic words are stored in multiple places in the lexicon, one related to mood and a second related to location in the cases of "up" and "down." The context then disambiguates which entry in the lexicon is accessed. But this tactic undermines the very notion of a single stable entity and approximates the position that different events in the world simply produce distinguishable physical events within the nervous system. Put briefly, the issue is whether symbols and representations are inferences from behavioral observations that are simply convenient terms for the *effects* of more fundamental biobehavioral processes (i.e., they are instrumental fictions) or whether they are among the *causes* of behavior. *MCB* takes the latter view. On this issue the authors observe:

> Roughly speaking the more committed theorists are to building psychological theory on neurobiological foundations, the more skeptical they are about hypotheses that there are symbols and symbol processing operations in the brain. (p. 101)

However, instead of sounding a cautionary note for *MCB*, this state of affairs is taken as indicating that the information-processing approach is largely unknown to such theorists.

The last of the fundamental concepts considered here concerns the nature of memory. As already noted, a "read/write memory" is seen as central to the information-processing approach presented in *MCB*. Read/write memory is shorthand for the following process: Experience with the environment leads to the storage of symbols (write). These stored symbols are then retrieved at an appropriate later time (read) and they, together with other symbols retrieved through the direct action of the contemporaneous environment, provide inputs to the processor. The processor then computes the output and behavior occurs. In keeping with the computer model, each stored memory can be retrieved because it has an address (a location in the brain), and that location also contains the address of a related memory. Thus one memory can lead to another. The read/write view of memory is contrasted with what might be called a write-only memory, in which the products of experience are stored and then later retrieved if the contemporaneous environment contains events that were present when the memory was originally stored. In the terminology of computer science, this latter form of memory employs a "look-up table" in which the appropriate output for every experienced input is stored. Such a memory system is said to constitute a "finite-state automaton" because it can process only the finite number of inputs to which it has been exposed. Write-only memory is held to have two major limitations. First, the look-up table becomes unmanageably large if every output must be stored for every conceivable input. In the authors' words, a finite-state memory becomes "overwhelmed by the infinitude of the possible" (p. 177). Second, a finite-state memory is inflexible because the organism can behave only in ways already found to be appropriate for that particular input. Write-only memory is said to be the only sort of memory consistent with associationist psychology and—by extension— with behavior analysis, which is not distinguished from associationism by *MCB*. Associationist psychology is not the only target of this criticism:

> Contemporary neuroscience is committed to the thesis that the brain had the functional architecture of a finite-state automaton rather than that of a Turing machine: it lacks a read/write memory. (p. 176)

MCB greatly underestimates both the economy and fruitfulness of write-only memory as it is realized in living organisms. With regard to economy, stimulus generalization permits the effects of specific environments to transcend their particular features. At the behavioral level of analysis, the processes involved in stimulus generalization are quite well understood when subjected to experimental analysis (Guttman & Kalish, 1956). Any newly encountered configuration of environmental stimuli likely contains stimulus elements that have previously occurred when some behavior was reinforced. As a result, new environments (i.e., new combinations of previously sensed stimuli) guide the occurrence of *new combinations* of previously selected responses (Bickel & Etzel, 1985; Donahoe & Wessells, 1980). These new mixtures of behavior then become subject to whatever

contingencies of reinforcement prevail in the new environment. In this way, response topographies can also change. At the neural level, the processes mediating stimulus generalization are well known for simpler stimuli (Thompson, 1965) and appear to be similar for more complex stimuli (DeValois & DeValois, 1988). These processes have been simulated in neural networks using coarse coding in which stimuli activate overlapping groups of units. The behavior that is guided by the stimulus is also not well characterized as a response of a single topography. The contingencies of reinforcement are most commonly satisfied by a range of responses, not a single response. That is, what is selected is an operant. The outcome of selection by reinforcement is typically that a *range* of stimuli come to guide a *rang*e of responses. Reinforcement affects the relation between classes of environmental and behavioral events (Skinner, 1935). Thus, the outcome of selection by reinforcement is also more fruitful than characterized by *MCB*'s view of write-only memory. The net effect of selection by reinforcement of relations between classes of stimuli and responses is that the functional dimensions of the look-up table become much larger than the particular environmental and behavioral events that were previously experienced.

Be that as it may, *MCB* is onto something important when it asserts that not all memory can be understood as the expression of previously selected environment–behavior relations. If I were asked "What did you do yesterday afternoon?" the correct response is "I worked on the review of *Memory and the Computational Brain*." But I have not been asked this question before and, hence, could not have received a prior reinforcer for the response. Moreover, the correct answer to the question changes with time (at least I hope so!). How is this example of memory to be understood? *MCB* proposes that the question initiates computations in which stored memories are retrieved together with the addresses of other stored memories and the contents of these new addresses are then accessed and processed until the sequence produces the normatively correct response. In this way, the results of experience in a prior environment (yesterday afternoon's) that differ from today's environment may affect current behavior. The processor has access not only to current inputs (write) but also to the effects of prior inputs (read)—in short, a read/write memory. (This specific example is an instance of what is conventionally labeled "episodic memory"—not "semantic memory," to which the account is most often applied; however, the basic point remains the same.) A read/write memory attempts to solve an important problem in memory, but in so doing it raises other problems: How do the addresses coordinated with the symbols activated by the present environment come to point to the addresses of appropriate other symbols stored as the result of earlier experience? These addresses are needed to support "indirect addressing," "linked lists," "pointers," and the like that allow the read process to continue. In computer programs the required addresses can be specified by the programmer, but how is the on-line programming needed for memory accomplished, and how is it implemented in the brain? *MCB* is silent on such questions.

The last point considered here is *MCB*'s view of an alternative, neurally-inspired approach to memory—neural networks. Neural networks are

interconnected units of neuron-like elements whose strengths of connections are modified by the simulated experience of the network (Donahoe & Palmer, 1989; Rumelhart & McClelland, 1986). Neural networks of interest have input units activated by external events, output units whose activations constitute the behavior of the network, and interior units (so-called "hidden" units) that intervene between input and output units. Most often, the units in neural networks are sub-symbolic with behavior emerging from which symbols are sometimes inferred. The relative slowness of neurons in comparison to the logic gates of digital computers is circumvented in read/write memory through the postulated storage of informationally compact symbols. Any adequate treatment of memory must engage the speed issue because tasks, such as the perception of complex stimuli and memory for vast sets of heterogeneous events, can be parsed more rapidly by living organisms than by even sophisticated programs implemented on modern computers. Neural networks attempt to meet this challenge through the simultaneous action of many units working in concert, that is, through units acting in *parallel*. The generally high levels of "spontaneous" activity observed in cortical neurons are consistent with this view. Of the neural-network approach to the speed problem, *MCB* comments "Massive parallelism. . .[is]. . .mostly a cloak for ignorance" (p. 174) and "Contemporary neuroscience is committed to the thesis that the brain had the functional architecture of a finite-state automaton rather than that of a Turing machine: it lacks a read/write memory" (p. 176). Whether neural networks are, in fact, committed to a finite-state memory is a matter to which I shall return.

Implications of the Conceptual Foundations

MCB's commitment to a digital-computer architecture leads the requirements for a functioning read/write memory to be described as follows:

> The architecture of the computing machine must make provision for values to be retrieved from physically different locations in memory and brought to the processing machinery that implements the primitive. . .functions. (p. 284)

> We need to first extract one value, hold it in a memory register while we extract the second value, then bring the two sequentially extracted values to the machinery that can realize the specified. . .function, and then out the value thus obtained. . . (p. 285)

The difficulty is that current neuroscience provides no support for such requirements. This state of affairs does not discourage *MCB*. *MCB*'s reaction is not to question their formulation but to posit that neuroscience has not been asking the right questions, and that when the right questions are asked the digital-computer machine architecture will be uncovered. The theoretical tail is ever more vigorously wagging the experimental dog. Let us briefly assess the relevant current knowledge in computer science, neuroscience, and behavior to see whether *MCB* has fully appreciated the implications of this knowledge for memory.

Neural Networks

First, are neural networks—which posit a richly interconnected set of neuron-like units—Turing-machine competent, or are they restricted to a finite-state conception of memory? Neural networks can be Turing-machine capable. If a neural network has recurrent connections (i.e., feed-back connections whereby units later in the network can affect the activity of more "upstream" units) then a neural network can have the functionality of a Turing machine. In this way, the effects of previous activity can influence current neural activity. This has been proven (Siegelmann & Sontag, 1991) for networks that have units with the nonlinear thresholds needed for nontrivially modifying the strengths of connections to hidden units (Rumelhart & McClelland, 1986) and which occur in neurons as shown by the Goldman equation (Junge, 1981). Furthermore, neuroanatomy has documented that rich sets of connections (cortical tracts) exist whereby activity in, say, motor-association areas may affect activity in sensory-association areas (Fuster, 2008; Fuster, Bodner, & Kroger, 2000).[2]

Feed-Back Connections

Next, do biobehavioral accounts of memory exploit the presence of feed-back connections? Again, the answer is yes. Biobehavioral accounts distinguish between memory that may be understood when the present environment evokes behavior that has been previously conditioned to that or similar stimuli and memory in which the present environment does not contain such stimuli. This former type of memory is termed "reminding" and the latter "remembering" (Donahoe & Palmer, 1993/2005; Palmer, 1991). In *MCB*'s terms, these distinctions correspond to the functionality of finite-state and Turing machines, respectively. The salient characteristics of remembering may be traced back to comments by B. F. Skinner:

> When a person recalls something he once saw, or engages in fantasy, or dreams a dream, surely he is not under the control of a current stimulus. Is he not then seeing a copy? Again, we must turn to his environmental history for an answer. After hearing a piece of music several times, a person may hear it when it is not being played, though probably not as richly or as clearly. So far as we know, he is simply doing in the absence of the music some of the things he did in its presence. Similarly, when a person sees a person or place in his imagination, he may simply be doing what he does in the presence of the person or place. Both "reminiscing" and "remembering" once meant "being mindful of again" or "bringing again to mind"—in other words, seeing again as one once saw. . . (Skinner, 1974)

---

[2] Turing-capable computation has also been proven for neural networks more generally (Hyötyniemi, 1996) and for finite-state machines with multiple write capabilities as well as production systems (Minsky, 1967). Productions are roughly analogous to discriminated operants in which the productions are edited by a process that is roughly analogous to the blocking of conditioning.

Experimental evidence supports the biobehavioral account (Cabeza & Nyberg, 2000; Farah, 1989; Farah & Aguirre, 1999). In a particularly elegant experiment, monkeys were first conditioned using a matching-to-sample procedure in which computer-generated sample stimuli were very briefly presented to one side of a fixation point. The sample stimulus was followed after a short delay by two differing comparison stimuli (Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999). Responding was reinforced if the correct comparison stimulus for that sample stimulus was selected. Technically, the task may be described as a delayed arbitrary matching-to sample task (see Figure 1).
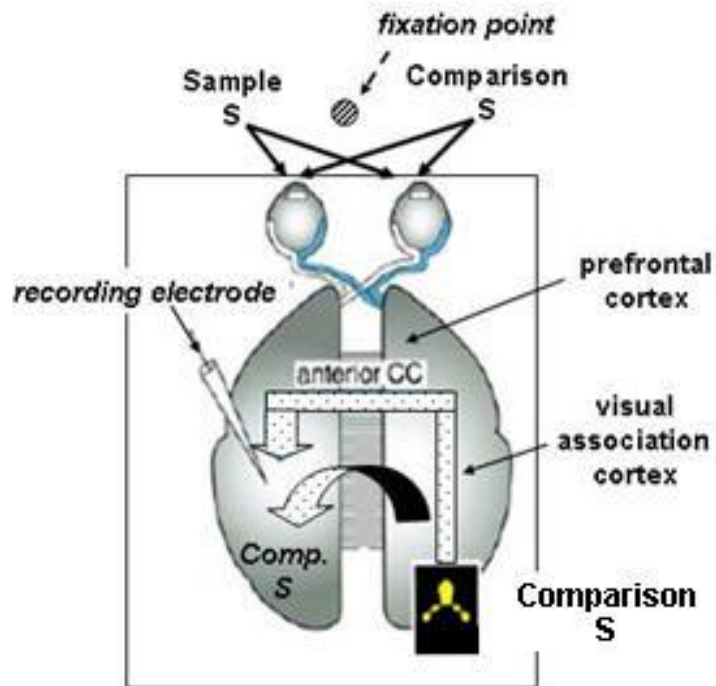

Figure 1. Arbitrary delayed matching-to-sample procedure showing the brain regions whose neurons were activated either directly or indirectly by the sample and comparison stimuli. The activated brain areas were restricted by presenting the sample stimulus briefly only when the eyes were foveating the fixation point. See the text for a more complete description of the procedure (Figure adapted from Tomita, et. al., 1999).

With an intact brain, the activity initiated by sample and comparison stimuli occurred in neurons in both the right and left sensory-association cortices. Because of the partial crossing of the optic nerve at the optic chiasma, stimuli to the left of the fixation point were projected comparatively directly to the right visual cortex (inferotemporal lobe) and then, via a feed-forward transcortical tract, through the posterior corpus callosum (CC) to the left visual cortex. Thus neurons in both sensory-association areas were activated by the visual stimuli in the intact brain. There is also a less direct pathway by which sensory-association neurons can be

activated by the visual stimuli: A transcortical tract arising from neurons in the sensory-association cortex projects to neurons in the ipsilateral (same-side) motor-association cortex (prefrontal cortex). Of particular importance for present purposes, the motor-association area is the origin of *feed-back* tracts from neurons in the prefrontal cortex along a transcortical tract traversing the anterior CC and activating neurons in the contralateral sensory-association cortex. Thus there are both relatively direct and more indirect means by which neurons in one sensory-association cortex are normally activated by visual stimuli regardless of the side of the fixation point on which the visual stimulus appears.

After behavioral responding had stabilized at high accuracy levels, the posterior CC was cut. This prevented sample stimuli briefly presented to one side of the fixation point from directly activating neurons in the ipsilateral primary visual cortex. However, it did not prevent neurons in the ipsilateral sensory-association cortex from being indirectly activated by neurons in the contralateral motor-association cortex via the feed-back tract coursing through the anterior CC (see Figure 1). Behavioral evidence that this feed-back pathway was effective is shown by the fact that matching-to-sample performance did not deteriorate after transection of the posterior CC: An intact feed-back tract was sufficient to mediate correct performance. Further evidence that the feed-back pathway was critical to performance was found when tracts in the anterior CC were cut. Without this feed-back, performance fell to chance levels. Thus, in an intact animal, any stimulus that can activate the appropriate neurons in the motor-association cortex permits the organism to remember the stimulus even though that stimulus is no longer present. In humans, verbal stimuli are especially well suited to fulfill this function. Verbal stimuli can generate activity in the motor-association area coincident with the activity generated by other contemporaneous stimuli. For example, if a visual stimulus prompts the subject to overtly or covertly respond verbally to the stimulus, then the motor activity generated by the verbal response can partially overlap with the activity generated by the visual stimulus in the motor-association cortex. Then, on some later occasion, the verbal stimulus may evoke, through feed-back connections, some of the same neural activity in sensory-association areas that was originally produced by the visual stimulus. Feed-back connections in the brain are competent to implement the functional capability of a Turing machine, and they do so without retrieving stored information to a read/write memory for processing (for a more complete interpretation of remembering, see Donahoe & Palmer, 1993). *MCB* holds that

> Brains must possess a mechanism for carrying a large amount of information forward in time in a computationally accessible form. . . .As of this writing, neuroscience knows of no plausible mechanism for carrying a large amount of information forward over indefinitely long intervals in a form that makes it accessible to the mechanisms. . .that are at the heart of computation. (p. 287)

The feed-back pathways just described accomplish this goal.

In addition to behavioral observations in the matching-to-sample experiment, Miyashita's group monitored the firing of individual neurons in the sensory-association cortex. These findings are shown in Figure 2.
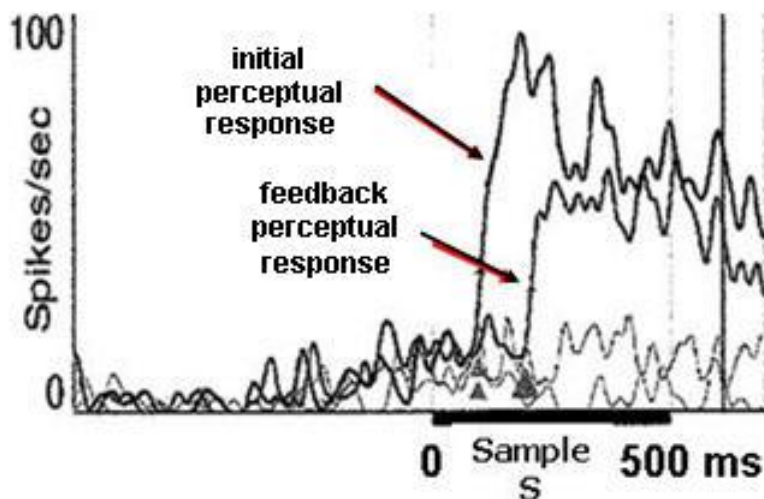

Figure 2. The number of spikes per second evoked in a sensory-association neuron in the inferotemporal cortex by a sample stimulus when tracts in corpus callosum were intact (initial perceptual response) and when only the posterior tracts arising from the contralateral motor-association area (prefrontal cortex) were intact (feed-back perceptual response). Adapted from Tomita, et. al., 1999.

Note two aspects of these findings especially. First, the perceptual response mediated by feed-back fibers in the anterior CC arising from neurons in the contralateral motor-association cortex arrived approximately 200 ms after the feed-forward perceptual response mediated by posterior fibers arising from neurons in the contralateral sensory-association cortex. The more circuitous route to neurons in the sensory-association cortex took a longer time to traverse. Second, the temporal pattern of firing of the sensory-association neuron by the feed-back pathway corresponded closely to the pattern produced by the feed-forward pathway. Thus the neuron responded similarly to the remembered and the perceived stimulus. It is important to keep in mind that the similarities in the temporal firing pattern of the neuron via feed-forward and feed-back pathways reflects interactions between that neuron and neighboring neurons with which it also makes synaptic contact. The same is true of the firing patterns of neurons in the motor-association cortex, whose axons give rise to the feed-back pathway. The temporal pattern of firing does not reflect a purely endogenous process. Whatever the nature of these interactions, they enable the organism, in Skinner's words, "to see again as. . .[it]. . .once saw" (Skinner, 1974). Note that in this conception of memory, memories do not exist except when they are occurring. What exist are

synaptic efficacies that permit the present environment to generate the remembered activity. A memory is not stored in some fixed location in the brain. Instead, only the capability for memory exists, ready to be realized when environmental events set in motion the behavioral and neural processes subsumed by memory.

MCB's criticisms of feed-back pathways (recurrent connections) in neural networks did not consider the possibility that these pathways might carry already highly processed neural signals that could be activated an indefinitely long time after they were originally established. Feed-back mechanisms considered in *MCB* are largely restricted to time-limited, reverberating circuits. Be that as it may, a strength of *MCB*'s discussion is its emphasis on the temporal pattern of firing of neurons. The role of inter-spike intervals is a relatively neglected topic in neuroscience. Variations in temporal firing patterns are known to occur, but whether they are appreciated by the organism is not yet clear (e.g., Richmond, 2009; Wiener, Oram, Liu, & Richmond, 2001). The opportunistic character of natural selection encourages the conjecture that they may well be. Much of the work on temporal patterns of firing cited in *MCB* is with the sensory neurons of insects, but technology is increasingly permitting comparable observations in more complex neural systems. As an example, very recent work with cortical neurons has shown that a single cortical neuron can respond differentially to various sequences of temporal inputs to its dendrites (Branco, Clark, & Gausser, 2010). Such a neuron provides a basis for the detection of different features of environmental stimuli; that is, the neuron becomes functionally a "place cell" for those features.

Synaptic Mechanisms of Neural Plasticity

MCB is skeptical of the ability of neural systems with feed-back pathways to function as Turing machines. Unfortunately, this skepticism has encouraged a companion effort to undermine long-term potentiation (LTP) as the major mechanism of synaptic plasticity. Some such mechanism is needed to enable the effective functioning of feed-back pathways:

> Neuroscientists justify the study of LTP by citing associative learning as an established psychological fact. While psychologists justify associative learning models by citing "associative" LTP as an established neurobiological fact, both groups and the progress of cognitive neuroscience would benefit from a fuller appreciation of the shakiness of the experimental foundations. . .to which each field appeals. (p. 179)

This is not the place to survey the vast LTP literature. However, a brief overview is warranted as an amendatory supplement to the account given in *MCB*, whose most recent LTP-related citation is more than 10 years old. The primary excitatory transmitter in the cortex is glutamate. Glutamate acts on two types of receptors in the post-synaptic membrane—a fast responding receptor (the AMPA receptor) and a slower responding receptor (the NMDA receptor; see Figure 3 for a

schematic representation of the major steps in the production of long-lasting changes in synaptic efficacy).
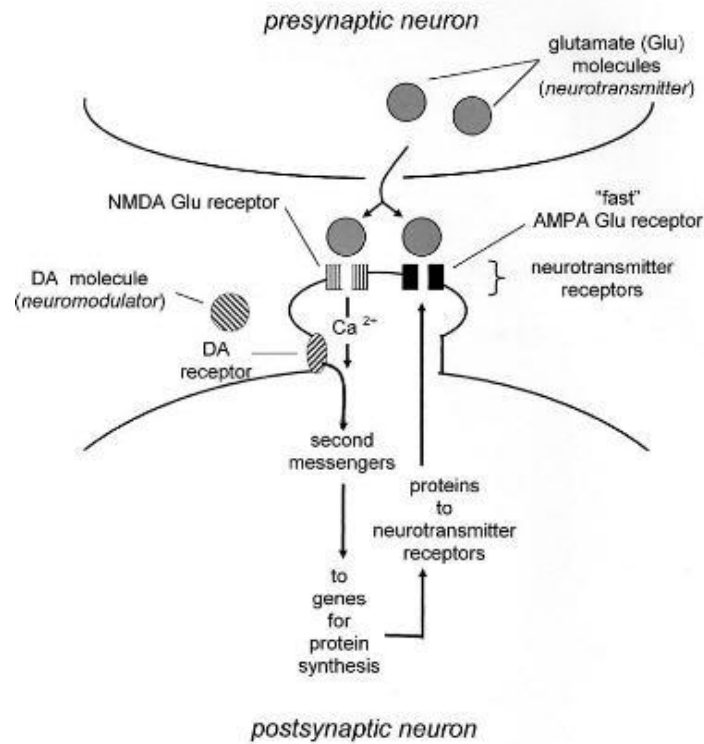

Figure 3. Schematic representation of the effects of the neurotransmitter (glutamate) on the glutamate receptors (AMPA and NMDA) that play a critical role in the development of long-term potentiation (LTP). If activation of these receptors is coincident with the release of the neuromodulator dopamine (DA) due to the occurrence of a reinforcer, then a long lasting increase in synaptic efficacy occurs between the presynaptic and postsynaptic neuron.

If the action of the AMPA receptor sufficiently changes the voltage across the cell membrane (depolarizes the cell), then the NMDA receptor allows the influx of calcium unto the cell. One effect of calcium influx is to place a "molecular tag" specifically on recently active AMPA receptors. The tag sensitizes these receptors to the action of glutamate for a few hours (so-called early LTP). If a reinforcer occurs, the neuromodulator dopamine (DA) is released concurrently. Because of the widespread distribution of DA fibers and their method of release of DA (volumetric release from varicosities along the axon), many synapses can be affected simultaneously. If the DA receptor is activated shortly after the NMDA receptor, then the tagged AMPA receptors ultimately become permanently more sensitive to glutamate. If the DA receptor is not activated, the tagged AMPA

receptors ultimately become less sensitive to glutamate (i.e., long-term depression or LTD occurs). DA initiates a chain of intracellular events (second messengers) that stimulate the synthesis of proteins through their effects on DNA. Protein synthesis may take several hours. The newly synthesized proteins migrate down the dendrite indiscriminately, and when they contact tagged AMPA receptors they change the configuration of these receptors. The specific receptors become permanently more responsive to the effects of glutamate (i.e., late LTP occurs). Although the understanding of LTP is clearly incomplete and important factors have been omitted from this brief summary, the current experimental evidence is not well characterized as "shaky" (for relevant references see Frey, 1997; Redondo, et. al., 2010; Sweatt, 2010).

MCB not only questions LTP as the major mechanism of synaptic plasticity in the "real" nervous system but also a commonly used technique for altering the strengths of connections between units in the artificial nervous system of neural networks. This technique is called back-propagation (Rumelhart, Hinton, & Williams, 1986). Back-propagation is rejected as biologically implausible, and I agree. However, there are other techniques for adjusting the strengths of connections in neural networks that are not subject to this criticism, notably some forms of reinforcement learning (Donahoe, 1997; cf. Sutton & Barto, 1998). In addition, even those who employ back-propagation agree that conclusions about the capabilities of neural networks are generally not dependent on the use of this technique (Rogers & McClelland, 2004). *MCB*'s reluctance to contemplate a system of memory based on connections between neurons (or units) has prompted an effort to undermine all of the major factors on which such an alternative account depends. Instead of LTP, *MCB* hypothesizes that an as-yet-unknown intracellular molecular process is responsible for storing memories. This is redolent of a much earlier and rejected proposal (Gaito, 1963, 1976) for which there remain rare proponents (Hameroff & Penrose, 1996) and no experimental evidence.

Concluding Comments

The treatment of memory in *MCB* is informed—but also strongly colored—by the conviction that memory is the product of computation by a machine having the architecture of a digital computer. This belief arises primarily from the formal tenets of computation theory, which hold that such an architecture is required—or, at least, highly likely—to implement a read/write memory. Read/write memory, or its functional equivalent, is needed for remembering. In remembering and other complex behavior such as problem solving, of which remembering is a special case (Donahoe & Palmer, 1993), current behavior is guided not only by the present environment but also by the effects of prior environments. Darwin cautioned biologists to be wary of formal restrictions after the Scots engineer Fleeming Jenkin (1867) "proved" that natural selection could not underlie evolution because rare beneficial variants would be swamped by more numerous, but less beneficial, variants (neither Darwin nor Jenkins knew of the gene, of course). Major contemporary contributors to the literature on memory have also expressed

reservations about the digital computer as a model for memory (e.g., McClelland, 2009), a judgment that this review shares. In defense of its theoretical commitment, *MCB* challenges even widely held views of brain function. As an example: "The synaptic plasticity hypothesis. . .fundamentally misrepresents the nature of what it is we should be looking for" (p. 278).

The theoretical orientation of *MCB* also affects the instances of memory that are examined in detail—navigating between food sources and the nest by desert ants and honey bees, and retrieving seeds cached in hundreds of different locations by certain species of birds. These examples hardly leap to mind when one thinks of memory, but they are well-chosen from a computational perspective. That is, algorithms can be readily implemented in digital computers to determine efficient paths for ants to traverse using dead reckoning and for bees to fly using solar positioning. Further, the locations of hundreds of seed caches can be readily stored in the memory of a digital computer where they are available to be retrieved later. Such phenomena may not be paradigmatic of memory, but biobehavioral science must ultimately find a place for them. Toward that goal, an alternative interpretation of seed-caching by birds is offered that makes no appeal to memory whatsoever.

The core aspect of caching that must be explained is how a bird, such as Clark's nutcracker, can retrieve seeds with substantial accuracy—perhaps 70%—that have been inserted below the surface many weeks before at hundreds of different locations over an area perhaps as large as 400 square kilometers. This remarkable behavior is interpreted by *MCB*, and by others, as a prodigious feat of memory. There is, however, a possible alternative account of major aspects of this behavior that does not involve memory at all. The visual system of birds and other animals is sensitive to a range of spatial frequencies, which—to an approximation—may be thought of as bars differing in width, spacing, and orientation (DeValois & DeValois, 1988; Donahoe & Palmer, 1993). The two eyes of birds (except birds of prey) have essentially non-overlapping fields of vision because their eyes are located on the sides of their heads. Further, the optic tracts are completely crossed so that the two eyes innervate different sides of the brain. This is an effective arrangement for the detection of predators over a very large field of view, but it has another consequence. When a bird views its environment, the particular combination of spatial frequencies that it senses with the two eyes specifies the bird's location in that environment. If there is a stable preference for caching when certain combinations of spatial frequencies are sensed, then that combination can evoke cache-recovering behavior at a later time when those combinations are again encountered. According to this interpretation of cache retrieval, the bird does not remember the location of the cache; the location reminds the bird of caching. The natural environment in which Clark's nutcracker caches seeds is at altitudes near the tree line, where trees are relatively sparse, thereby providing many sites with distinctive combinations of spatial frequencies. Furthermore, the trees are evergreens so that the spatial frequencies are relatively unchanged during the winter when the environment may be snow-covered and when reasonably accurate cache-retrieval still occurs. Laboratory studies of cache

retrieval are consistent with the proposed account. Landmarks nearest the location where the cache is buried are most important for retrieval, and changing their position most disrupts retrieval (Goodyear & Kamil, 2004). Near objects most precisely specify location because they provide the most discriminative spatial frequencies: Displacement of near objects changes the spatial frequencies more than equal absolute displacements of more distant objects. Again, the alternative account requires no memory whatsoever! It simply requires an innate preference for caching under specific visual conditions (and the preference should vary among birds to limit competition). The ability to locate caches based on visual cues is not specific to seed-caching birds but is present and similarly affected in other avian species, for example pigeons, although to a lesser degree (Jones, Antoniadis, Shettleworth, & Kamil, 2002). Thus, the exploitation of spatial frequencies to specify location does not require a great evolutionary leap by seed-caching birds; retrieval of caches requires only a preference for certain combinations of spatial frequencies that evoke caching. A complete account of cache retrieval undoubtedly involves additional variables, but the basic finding may be the result of processes that do not involve remembering—and, hence, a read/write memory.

The alternative account of the retrieval of caches illustrates a final general concern with the approach taken by *MCB*—an inadequate appreciation of the full contribution of the environment to behavior. In several places *MCB* refers to "the poverty of the stimulus," a phrase acknowledged as originating with Chomsky. Chomsky's phrase was prompted by his contention that the environmental events encountered during exposure to language were not sufficient to account for language acquisition. Chomsky used the claim to argue that an endogenous process compensated for the paucity of experience—a language-acquisition device implementing universal grammar. *MCB* similarly uses the poverty-of-the-stimulus argument to justify inferences about underlying memory processes—storage, retrieval, and read/write memory as implemented by a digital computer. When Chomsky originally made his claim, comparatively little was known about the details of language acquisition. The situation is much different now, when the verbal environment has been shown to provide a much richer source of stimulation to promote language acquisition (e.g., Donahoe & Palmer, 1993; Jusczyk, 1997; Lewis & Elman, 2002). Nevertheless, Chomsky's claim persists and is used to motivate conjectures about a variety of presumed internal processes in other contexts. Poverty-of-the-stimulus is an instance of "functional autonomy," a phrase introduced by the social psychologist Gordon Allport to describe concepts whose justification has disappeared but which nevertheless endure. One should exhaust the explanatory variance in the environment and its interactions with behavior before seeking internal events as the causes of behavior (Gibson, 1979). Sometimes it is not the stimulus that is impoverished but the theorist's conception of it. *MCB* identifies important problems that biobehavioral science must confront, but its proposed solutions should be regarded with extreme caution.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*, 249-277.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes, In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). San Francisco: Freeman.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results*. New York: Holt.

Bakan, D. (1953). Learning and the principle of inverse probability. *Psychological Review*, *60*, 360-370.

Bickel, W. K., & Etzel, B. C. (1985). The quantal nature of controlling stimulus–response relations as measured in tests of stimulus generalization. *Journal of the Experimental Analysis of Behavior*, *44*, 245-270.

Branco, T., Clark, B. A., & Gausser, M. (2010). Discrimination of temporal input sequences in cortical neurons. *Science*, Published online 12 August 2010; 10.1126.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*, 3-15.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*, 1-47.

Chamberlin, T. C. (1880). The method of multiple working hypotheses: With this method the dangers of parental affection for a favorite theory can be circumvented. *Science*, *15*, 92-96.

Dawson. M. R. W. (1998). *Understanding cognitive science*. Oxford, UK: Blackwell Publishers.

DeValois, R. L., & DeValois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

Donahoe, J. W. (1991). The selectionist approach to verbal behavior. Potential contributions of neuropsychology and connectionism. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 119-150). Reno, NV: Context Press.

Donahoe, J. W. (1997). Selection networks: Simulation of plasticity through reinforcement learning. In J. W. Donahoe & V. P. Dorsel (Eds.), *Neural-network models of cognition: Biobehavioral foundations*. New York: Elsevier Science Press.

Donahoe, J. W., & Palmer, D. C. (1989). The interpretation of complex human behavior: Some reactions to *Parallel Distributed Processing* by J. L. McClelland & D. E. Rumelhart (Eds.). *Journal of the Experimental Analysis of Behavior*, *51*, 399-416.

Donahoe, J. W., & Palmer, D. C. (1993). *Learning and complex behavior*. Boston, MA: Allyn & Bacon. (Reprinted by Ledgetop Publishers, 2005).

Donahoe, J. W., & Wessells, M. G. (1980). *Learning, language, and memory*. New York: Harper & Row.

Farah, M. J. (1989). The neural basis of visual imagery. *Trends in Neuroscience*, *10*, 395-399.

Farah, M. J., & Aguirre, G. K. (1999). Imaging visual recognition: PET and fMRI studies of functional anatomy of human visual recognition. *Trends in Cognitive Sciences*, *3*, 179-186.

Frey, U. (1997). Cellular mechanisms of long-term potentiation: Late maintenance. In J. W. Donahoe & V. P. Dorsel (Eds.), *Neural-network approaches to cognition: Biobehavioral foundations* (pp. 105-128). Amsterdam: Elsevier Science Press.

Fuster, J. (2008). *The prefrontal cortex* (fourth edition). New York: Academic Press.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*, 347-351.

Gaito, J. (1963). DNA and RNA as memory molecules. *Psychological Review*, *70*, 471-480.

Gaito, J. (1976). Molecular psychobiology of memory: Its appearance, contributions, and decline. *Physiological Psychology*, *4*, 476-484.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Goodyear, A. J., & Kamil, A. C. (2004). Clark's nutcrackers (*Nucifraga columbiana*) and the effects of goal–landmark distance on overshadowing. *Journal of Comparative Psychology*, *118*, 258–264.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*, 79-88.

Hameroff, S., & Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness? In S. R. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *Toward a science of consciousness: The first Tucson discussions and debates* (pp. 507-540). Cambridge, MA: MIT Press.

Hyötyniemi, H. (1996). Turing machines are recurrent neural networks. In J. Alander, T. Honkela, & M. Jakobsson (Eds.), *Proceedings of STeP'96 publications of the Finnish Artificial Intelligence Society* (pp. 13-24), Espo, Finland.

Jenkin, F. (1867). *The north British review*. (Reprinted in D. L. Hull, *Darwin and his critics*. Cambridge: Harvard University Press).

Jones, J. E., Antoniadis, E., Shettleworth, S. J., & Kamil, A. C. (2002). A comparative study of geometric rule learning by nutcrackers (*Nucifraga columbiana*), pigeons (*Columba livia*), and jackdaws (*Corvus monedula*). *Journal of Comparative Psychology*, *116*, 350-356.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Junge, D. (1981). *Nerve and muscle excitation* (2nd ed.). Sunderland, MA: Sinauer Associates.

Lewis, J., & Elman, J. (2002). Learnability and the statistical structure of language: Poverty of the stimulus arguments revisited. In *Proceedings of the 26th annual Boston University conference on language development* (pp. 359-370). Somerville, MA: Cascadilla Press.

McClelland, J. L. (2009). Is a machine realization of truly human-like intelligence achievable? *Cognitive Computation*, *1*, 17-21.

Minsky, M. (1967). *Computation: Finite and infinite machines*. Englewood Cliffs, NJ: Prentice-Hall.

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons* (expanded edition). Cambridge, MA: MIT Press.

Newell, N., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*, 113-126.

Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 261-279). Reno, NV: Context Press.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347-353.

Redondo, R. L., Okuno, H., Spooner, P. A., Frenguelli, B. G., & Morris, R. G. M. (2010). Synaptic tagging and capture: Differential role of distinct calcium/calmodulin kinases in protein synthesis-dependent long-term potentiation. *The Journal of Neuroscience*, *30*, 4981-4989.

Richmond, B. J. (2009). Stochasticity, spikes and decoding: Sufficiency and utility of order statistics. *Biological Cybernetics*, *100*, 447-457.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: Bradford MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318-362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Vol. 1* (pp. 110-146). Cambridge, MA: MIT Press.

Siegelmann, H. T., & Sontag, E. D. (1991). Turing computability with neural nets. *Applied Mathematics Letters*, *6*, 77-80.

Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *Journal of General Psychology*, *5*, 427-458.

Skinner, B. F. (1974). *About behaviorism*. New York: Alfred A. Knopf.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction (adaptive computation and machine learning)*. Cambridge, MA: Bradford MIT Press..

Sweatt, J. D. (2010). *Mechanisms of memory* (2nd ed.). San Diego, CA: Elsevier Academic Press.

Thompson, R. F. (1965). The neural basis in stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 154-178). Stanford, CA: Stanford University Press.

Tomita H., Ohbayashi, M., Nakahara, K., Hasegawa I., & Miyashita Y. (1999).Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*, 699-703.

Townsend, J. T. (1972). Some results concerning the indentifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, *25*, 168-199.

Wiener, M. C., Oram, M. W., Liu, Z., & Richmond, B. J. (2001). Consistency of encoding in monkey visual cortex. *Journal of Neuroscience*, *21*, 8210-8221.