This paper, “The neurological Basis for the Computational theory of Mind” to appear in the Festschrift for Jerry Fodor is copyrighted by Charles Gallistel.
© Charles Gallistel

The Neurobiological Bases for the Computational Theory of Mind

C.R. Gallistel

When we were young, Jerry and I, so, too, was the computational theory of mind, the central doctrine of cognitive science. It broke the theoretical fetters imposed by the mindless behaviorism that had dominated psychology, philosophy and behavioral neuroscience for decades. Jerry was making major contributions to cognitive science by spelling out its implications. Some of the implications seemed to me, then and now, to be obvious, but only after Jerry had spelled them out.

One such implication was that the mind must possess (unconscious) symbols and (equally unconscious) rules for manipulating them. That is, it must have a language of thought (Fodor 1975), just as do computing machines. Because the symbols are, on the one hand, objects of principled manipulation—they are the stuff of computation—and because, on the other hand, some of them refer to things outside the mind, it follows that the language of thought has a syntax and a semantics. When Jerry pointed this out, it seemed to me beyond reasonable dispute, although it has in fact been disputed, even unto the present day (Aydede 1977; Laurence and Margolis 1997; Schneider 2009).

What I found thought provoking about Jerry's insight was just what connectionists objected to: its neuroscientific implications. If one believed in the computational theory of mind, then the symbols and the machinery for manipulating them must have a material realization in the brain.

The problem was that no one knew what it might be. If there were symbols, then they must reside in memory, because the basic function of a symbol in a computing machine is to carry information forward in time in a computationally accessible form (Gallistel and King 2010). Most neuroscientists were and are unshakably committed to the hypothesis that memories consist of experientially altered synaptic conductances. I have been told in all sincerity that this hypothesis could not be false. Karl Popper would turn in his grave. There is, however, a problem with this hypothesis: synaptic conductances are ill suited to function as symbols (Gallistel and King 2010). Anyone who doubts this should ask the first neuroscientist they can corner to explain to them how the brain could write a number into a synapse, or into a set of synapses. Then, step back and watch the hands wave. In the unlikely event of an intelligible answer, ask next how the brain operates on the symbols written into synapses. How, for example, does it add the number encoded in one synapse (or set of synapses) to the number encoded in a different synapse (or set…) to generate yet another synapse (or set…) that encodes the sum?

Connectionists disliked the language of thought hypothesis because it was not readily reconcilable with what neuroscientists told them was the material realization of memory. However, as a behavioral neuroscientist, I knew how flimsy the evidence for the synaptic theory of memory was and how strongly neuroscientists' belief in it was undergirded by the associative theory of learning. The associative theory had—still has—enormous intuitive appeal: One of Lila Gleitman's many bon mots is that, "Empiricism is innate." Moreover, psychologists assured neuroscientists that the associative theory had dominated philosophical and psychological thinking for centuries, which is an historic truth, at least as regards Anglophone thought. So, how could this theory be wrong? As a psychologist who had focused on the theory of learning since my undergraduate days in the laboratory of Tony Deutsch, I knew how profoundly flawed the theory was—and is.

For me, the synaptic theory of memory rested on circularly reinforcing set of false beliefs: The neuroscientists' belief in the synaptic theory of memory was sustained in no small measure by the fact that it accorded with the psychologists' associative theory of learning. The psychologists' belief in the associative theory of learning was sustained in no small measure by its accord with what neuroscientists took to be the material realization of memory. Knowing this circular system of false beliefs, I was not tempted to follow where the connectionists wanted to lead, which was back to a murky, computationally hopeless associationism, motivated mostly by a misinformed assessment of what neuroscientists really new about the material realization of memory—which was nothing.

So, if symbols don't reside in altered synapses, where do they reside? I have argued against the synaptic theory of memory for decades, with no noticeable impact on the neuroscience community. My audiences always pester me with this question: If memory is not enduring changes in synaptic conductances, then what is its physical realization? I used to answer that I had no idea, but my audiences did not find that an appealing answer. Nor did I. Some years back, I began to have some ideas, but I was loath to put them in print, for fear they would further enhance my reputation for preposterous speculation. Now, however, very exciting experimental work in behavioral and systems neuroscience, which is just appearing, provides empirical support for at least the general thrust of these ideas.

Where to Find the Symbols

We have been looking in the wrong place—for both the symbols and the machinery that operates on them. The symbols are not in the synapses, and the machinery that operates on them is not (primarily) in the neural circuits. The symbols are in molecules inside the neurons, and the machinery that operates on them is intracellular molecular machinery.

On this view, each neuron is a computational machine. It takes in information through its dendrites, processes that information with complex information processing machinery implemented at the molecular level within the neuron itself, and, at least sometimes, it then generates a signal that carries the encoded results of its processing to other neurons by means of a patterned train of nerve impulses. On other occasions, it may only update its memories and not send out any signal.

Because symbolic memory is an indispensable component of any computing machine (Gallistel and King 2010), the molecular-level information processing machinery inside each neuron has, as one of its most basic constituents, molecules whose form may be altered by experience in such a way as to encode acquired information, information that has been conveyed to the cell through its synaptic inputs. These intracellular memory molecules carry the acquired information forward in a form that makes it accessible to the molecular computing machinery.

Insofar as neuroscientists are also biologists, they have known for decades where in the brain they *could* find materially realized symbols and computational machinery that operates on them. But they have assumed—without ever discussing the possibility—that what they knew about the genetic machinery in every cell was not relevant to the question of the material basis of memory. Anyone familiar with the rudiments of molecular biology knows that (most) codons[1] are symbols for amino acids. And, they know that the sequence of codons between a start codon and a stop codon is a symbol for a protein. If they have been following the evo-devo literature, they know that some proteins represent highly abstract aspects of organic form, such as *anterior*, *dorsal* and *distal* (Carroll 2005; Shubin, Tabin et al. 2009), while others represent complex organs, such as eyes. Non-biologists are sometimes startled to learn that there is such a thing as a gene for an eye; turn it on and you get an eye (Halder, Callaerts et al. 1995; Gehring 1998). What is more amazing, we now understand how this is possible. The old saw that genes cannot represent complex organic structures and abstract properties of a structure is simply false; they can, and they do.

The symbols strung out along the double helix carry *inherited* information, information about what worked in the ancestors of the current carrier of that information. These symbols are organized into data structures, just as are the symbols in the memory of a computer. The computational principle that makes this organization possible —the indirect addressing of stored information—is the same in the nucleus of a cell as it is in a computer (Gallistel and King 2010). Moreover, the molecular machinery that reads the information and uses it to guide organ construction and govern cell function implements the logic gates that are the building blocks of computational machinery. In short, the DNA symbols carry information in a form that makes it accessible to computation, and there is molecular machinery that performs computational operations in the course of reading this information. Could it be that the process of evolution has found a way to

[1] A codon is a reading-frame triplet of nucleotides. There are four nucleotides. During the transcription of a gene, the double helix is read in the sense direction along the sense strand (as opposed to the antisense strand) in nucleotide triplets (3-letter words, written in the 4-letter alphabet of nucleotides). The reading frame is determined by the nucleotide from which transcription starts. In addition to the codons (words) that code for amino acids, there are punctuation codons that indicate the beginning and end of a codon sequence that constitutes one gene.

make use of this machinery—or the closely allied RNA machinery—to store *acquired* information and to carry out computations with it?

Stored Information in Computers and Genes

In computer memory, the symbols, that is, the words in memory locations, have a bipartite structure. One part digitally encodes some information; the other part, the address part, makes that information accessible to information processing operations. Genetic symbols have this same structure: every gene has a coding portion, in which the codon sequence encodes the amino acid sequence of a protein. Every gene also has one or more promoter and repressor components. They give the rest of the cellular machinery controlled (programmed) access to the information in the coding portions of genes.

In computer memory, the coding portion and the address portion of a symbol are both bit patterns. Thus, a copy of the bit pattern that constitutes the address portion of one symbol may be stored in the coding portion of another. The storing of addresses makes possible indirect addressing. Indirect addressing makes variable binding possible. Variable binding makes data structures possible. Data structures are the soul of a computing machine. They embody the computer's knowledge.

In genetic memory, the coding portion of a gene and the promoter/repressor portions are both nucleotide sequences. The proteins called transcription factors contain segments that bind to the promoter and repressor regions of specific other genes. The bipartite structure of the gene and the selective binding of transcription factors to particular promoter and repressor sequences together implement indirect addressing in genetic memory. Indirect addressing in genetic memory is what makes the eye gene possible. It sits atop a genetic data structure in the cellular nucleus, just as the symbol for a document file (the name of the file) sits atop a data structure in the memory of a computer.

Variable Binding, Indirect Addressing and Data Structures

Among the first operations that the beginning student of computer programming learns is the operation of assigning a value to a variable. In most computer languages, it goes like this:

"W = 135," which translates as "set the value of a variable W to 135"

Conceptually, this operation creates two symbols in the memory of the computer, that is, two bit patterns stored at different locations in memory. One is the bit pattern for the variable, W. The other is the bit pattern for the current value of this variable, namely, 135. This latter bit pattern is the physical realization of the number that specifies, say, someone's weight. Each location in memory has a unique address, its own zip code, so to speak. The problem of variable binding is the problem of getting from the symbol for the variable to the symbol for its value.

Given the bipartite structure of computer symbols, it is fairly obvious how to solve this problem: Make the bit pattern for the address of the value the bit pattern that represents the variable. Then, when the machine goes to the address where the symbol for the variable is to be found, the bit pattern it finds at that address is the address of the variable's value. This bit pattern is called a pointer. To get to the value of a variable, the machine rarely goes directly to the address where the value itself is to be found; rather, it goes to an address where it finds a symbol that points to the address of the value. Or, in the more complex reality, it goes to an address where it finds a number, which, after some, possibly rather complex computations involving other numbers, yields the address of a variable's value. These computations are called pointer arithmetic. One consequence of this principle is that the contents of many words in computer memory do not refer to things outside the machine; rather they are the addresses of other locations in memory. They have a purely internal reference. That is how data structures are built up in the memory of a computer.

The genetic machinery works in the same way. The coding portion of many genes does not encode the structure of a protein that forms an element of cellular or tissue structure; rather, it encodes a protein that is a transcription factor, a genetic pointer. Transcription factors bind to the promoter and repressor regions of genes, activating or suppressing the transcription (reading) of their coding portions. Thus, for example, the gene for an eye does not encode the structure of a protein found in the realized eye; rather, it encodes the structure of a transcription factor. And, the genes to whose promoters that factor binds also encode transcription factors. One has to go down a fair ways in the genetic data structure to get to genes that encode proteins that form structural components of the realized eye.

The genes that encode transcription factors are symbols for variables. Indirect addressing gives the cellular machinery structured access to the data that specifies how to build an actual eye, that is, how to realize the value of the eye variable. The distinction between the symbol for the variable (the genetic symbol for an eye) and the symbols for the value of that variable in a particular case (the genetic data structure that, when appropriately read, yields a realized eye) is dramatically illustrated by the fact that the genetic symbol for an eye is homologous in the human and the fruit fly. The homology is so close that one can put the human gene into the cells of a developing fruit fly, turn it on at some location, and generate an eye at that location—the faceted dome fruit fly eye, not a human eye, with its lens and pupil (Quiring, Walldorf et al. 1994). Thus, the physical realization of the symbol for an eye is (almost) the same in the fruit fly and the human genome, but the realized eye—the value of the variable—is radically different. The physical symbol for the genetic program that makes an eye has remained the same through hundreds of millions of years, while the nature of the eye-constructing program itself, hence, the structure of the realized variable, has diverged greatly.

The Building Blocks of Computation

The building blocks of physically realized computations are logic gates, simple structures that realize the logical operations AND, OR, NOT, NAND and XOR (exclusive or). These operations are implemented at the molecular level in the reading of genetic data structures. Transcription factors often form dimers, that is, they transiently bind to one another, forming a molecular compound with functional properties its constituents lack. Transcription Factor A and Transcription Factor B may neither of them bind to the promoter of Gene X, but their dimer may do so. In that case, when either factor is present alone, Gene X is not transcribed, but when both are present, it is. This is a molecular AND gate. If, on the other hand, the AB dimer binds to a repressor of Gene X, then the gene may be transcribed when neither factor is present or when either is present alone but not when both are present. This is a molecular level NAND gate. (There is a proof in theoretical computer science that to build a computer all you need are NAND gates.) Or, it may be that either A or B will bind to the promoter of Gene X, thereby activating its transcription, but when both are present, they dimerize, and the dimer no longer binds to that promoter. This implements XOR. Finally, of course, the binding of a transcription factor to a repressor implements NOT.

In short, processes operating within cells at the level of individual molecules implement the basic building blocks of computation, and they do so in close connection with the reading of stored information. The information in question is hereditary information, not experientially acquired information. Nonetheless, it is tempting to think that evolution long ago found a way to use this machinery, *or closely related machinery, or, at the very least, functionally similar molecular machinery*, to do the same with experientially acquired information.

Size Matters

The logical gates that are the building blocks of computational machinery can also be implemented by neural circuits. However, in pondering the relative plausibility of intracellular molecular implementation versus neural circuit implementation of basic computational operations, one should keep in mind the vast difference in the size of the posited machinery. One turn of the DNA helix, which contains 11 nucleotides and can encode 22 bits of information (2 bits per nucleotide), has a volume of about 1.1x10^{-26} meters (11 cubic nanometers), whereas one neuron has a volume on the order of 2x10^{-14} meters (20,000 cubic microns). Thus, machinery built at the level of molecules occupies 12 to 15 orders of magnitude less volume than machinery built at the level of neurons. It is hard to grasp how great a difference in size this is—roughly the difference in size between a neuron and the original Univac computer. It is also roughly the difference in size between the original Univac and a contemporary computing state-of-the art CPU. The contemporary CPU is a very much better computing machine than the original Univac, largely because it is so much smaller, faster and more energy efficient. For the same reasons, molecule-sized computing machinery inside neurons would be many orders of magnitude smaller faster and more energy efficient than the same machinery implemented at the level of neuronal circuits, using synapses as memory elements.

Evidence

These are the thoughts that have slowly taken form in my mind. But where is the evidence? Other than plausibility arguments for why memory and computation in nervous tissue ought to be an intracellular molecular-level process rather than an intercellular circuit-level process, is there any experimental evidence? Until very recently, I had to admit the answer was, no. However, the Hesslow laboratory in Lund have recently described work showing that the acquired information that informs the appropriately timed conditioned eyeblink response in the ferret resides within individual Purkinje cells in the cerebellar cortex. This same experiment shows that the cell possesses machinery capable of reading out this information into complexly structured spike trains in response to synaptic inputs, which inputs indicate simply and only the onset of a conditioned stimulus. This minimally informative input, which contains no information about the temporal relation between the conditioned stimulus and the unconditioned stimulus, produces a complex spike-train output that is informed by acquired temporal information stored within the cell.

Behavioral experiments long ago showed that a critical component of the information acquired during Pavlovian conditioning was the duration of the inter-stimulus interval (ISI), the interval between the onset of a predictive stimulus (the CS, which is short for conditioned stimulus) and the onset of the event it predicts (the US, short for unconditioned stimulus). Experiment showed that the timing of the acquired response to the CS varies in a systematic, functionally appropriate way with the inter-stimulus interval. The animal does not simply blink in response to the CS; it blinks at the right time. The latency of blink onset varies in proportion to the variation in the duration of the ISI in such a way that the eye reaches maximum closure at the moment when the CS predicts that the US will occur. The ISI-dependent timing of the conditioned response is observed in all of the simple learning preparations that are used to investigate the neurobiology of associative memory (Gallistel and Gibbon 2000; Balsam and Gallistel 2009; Balsam, Drew et al. 2010), so finding the mechanism that stores this temporal information is critical to a neurobiological understanding of learning and memory.

It has always been assumed that the structural change mediating an appropriately-timed acquired response (a CR for conditioned response) must lurk within the mechanism of synaptic transmission. It is taken for granted in the literature on the neurobiology of learning, that learning of any kind must alter either the release of transmitter from pre-synaptic terminals or the mechanisms that mediate the binding of the transmitter to receptor molecules in the post-synaptic membrane, or perhaps both. However, how alterations in those synaptic processes could store the duration of an interval has always been a mystery. The neurobiologists have taken the primary challenge to be explaining the fact that the conditioned response occurs, not to explain the fact that it occurs at the right time. The latter fact has been treated as one of those mysteries that we would tackle later, once we had solved the mystery of why a conditioned response develops.

Neural net modelers (connectionists) have suggested that the information about the duration of the ISI resided in some unspecified way in alterations in the pattern of synaptic connections (synaptic weights) within a complex neural network (Martin and Morris 2002), but these suggestions have been vague. There has never been a specification of what the synaptic code might be, nor how the stored information could be made accessible to computation. Specifying the synaptic code would enable us to understand why one pattern of synaptic connections encoded the fact that the ISI lasted 150 ms while another encoded the fact that it lasted 300 ms. Specifying how the encoded information entered into computations would enable us to understand how the first encoding produced a blink that culminated at 150 ms while the second produced a blink that culminated at 300 ms. Other theorists have assumed that the answer lay in selective associations between neurons—or even whole networks of neurons (Matell, Meck et al. 2003; Meck, Penney et al. 2008)—with intrinsically different temporal dynamics. It has been suggested that neurons that intrinsically (in the absence of any informative experience) respond to the CS with a firing rate that rises to a peak at 150 ms and then declines rapidly become preferentially associated with the blink response when the ISI is 150 ms, whereas neurons with a slower dynamic, peaking at 300 ms, become selectively associated with the response when the ISI is 300 ms (Grossberg and Schmajuk 1989; Yamazaki and Tanaka 2009). None of these models was plausible on its face and none has received empirical confirmation (Hesslow, Jirenhed et al. 2013). So, the mystery of the structural change that mediates the appropriate timing of the conditioned response has remained throughout more than half a century of intensive research on the neurobiology of learning and memory.

The just-appearing experiment from the Hesslow laboratory builds on the progressive refinement of eye-blink preparation over the last half century. In this preparation, the CS is usually a tone whose onset warns of a perioccular shock or an air puff to the eyeball (the US). The US causes the eye to blink. If the onset of the tone CS reliably predicts the US at some fixed latency, the animal learns to blink in response to the onset of the tone. As already emphasized, the latency at which it blinks varies in proportion to the latency between tone onset and shock (the ISI).

Earlier work, most notably from the laboratory of Richard Thompson, showed that the critical circuitry was in the cerebellum (Krupa, Thompson et al. 1993; Bao, Chen et al. 2002; Christian and Thompson 2003). This was itself a surprise. It was also exciting, because the neuroanatomy of the cerebellum is relatively simple and extremely repetitive. For which reason, there has been a huge amount of work on the neurophysiology and neuropharmacology of cerebellar circuitry. When focus shifted to the cerebellum with its (relatively!) simple circuits, it was shown that the blink response was gated by a CS-produced pause in the endogenously generated

(Cerminara and Rawson 2004) basal firing of the large Purkinje cells (Jirenhed, Bengtsson et al. 2007). The axons of the Purkinje cells carry the output signals from the cerebellar cortex. The duration of this learned pause in Purkinje cell firing covaried with the ISI used in training (Jirenhed and Hesslow 2011; Jirenhed and Hesslow 2011). Thus, the pause in the firing of the Purkinje cells is an electrophysiological proxy for the conditioned response.

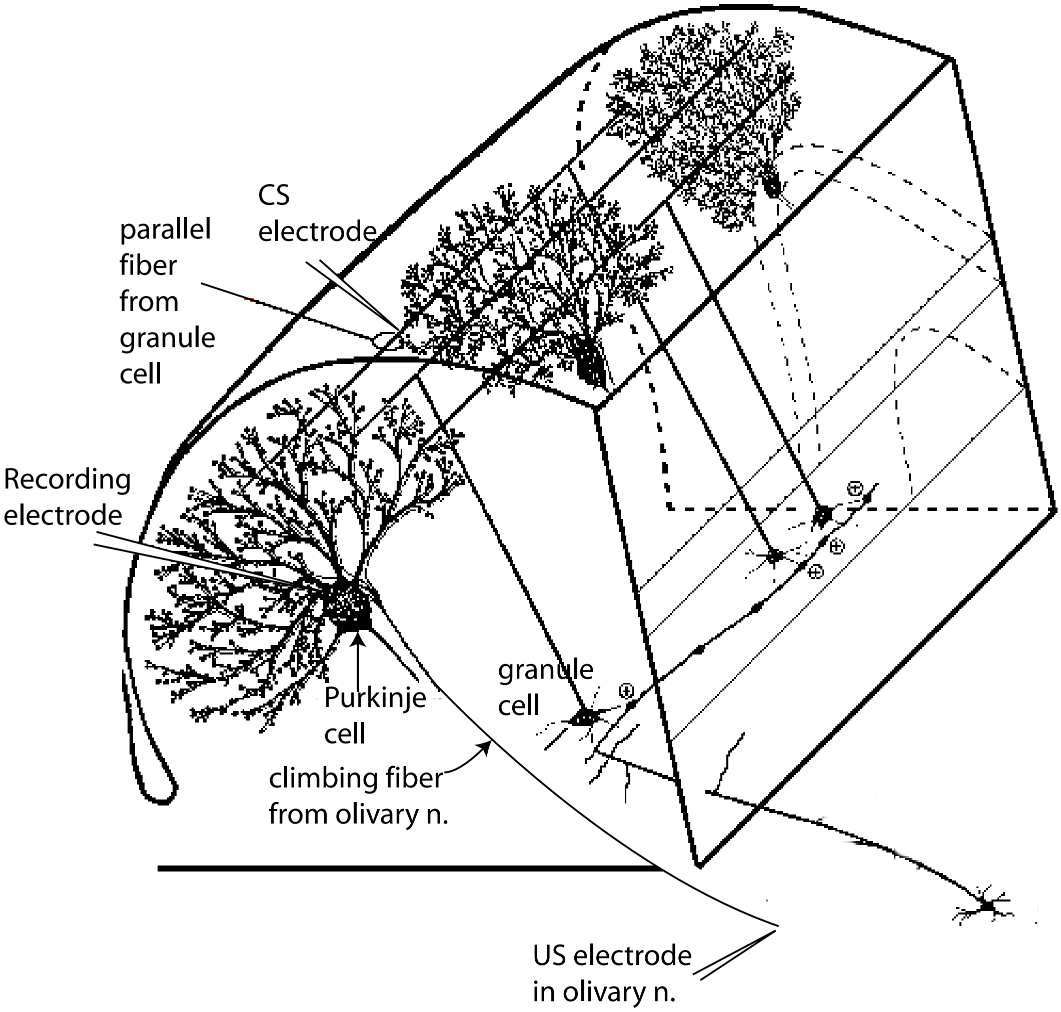
CS electrode
parallel fiber from granule cell
Recording electrode
Purkinje cell
granule cell
climbing fiber from olivary n.
US electrode in olivary n.

Figure 1. *The experimental preparation in the recent experiment from the Hesslow laboratory. The artificial CS was a spike train generated in the parallel fibers by a train of stimulating pulses delivered through the CS electrode. The artificial US was a spike train in the climbing fibers generated by direct electrical stimulation of the olivary nucleus. The conditioned response of the Purkinje cell was monitored via a recording electrode.*

The massive dendritic tree of a Purkinje cell spreads athwart the parallel fiber system in the cerebellum (Figure 1). The parallel fibers coursing along the folds in the cerebellar cortex are analogous to the signal bus in a computer. A Purkinje cell reads the signal pattern across a portion of this signal bus. The schematic in Figure 1 is grossly misleading as regards the density of the parallel fibers. The dendritic tree of a single Purkinje cell is synaptically contacted by as many as 200,000 parallel fibers (Harvey and Napper 1991). The neural signals generated by behaviorally effective CSs reach the Purkinje cells by way of the mossy fibers that synapse on granule cells in deep cerebellar nuclei. The granule cells give rise to fibers that ascend almost to the cortical surface of the cerebellum, where they send branches that run along the folds in the cerebellar cortex. These branches are the parallel fiber system, the cerebellar signal bus. Signals generated by USs (predicted stimuli) reach the cerebellar Purkinje cells by way of the so-called climbing fibers, which originate from cells in the olivary nucleus of the cerebellum. Thus, the Purkinje cell is one of several sites of convergence of CS and US signals.

A second exciting advance was the demonstration that the learned pause in the firing of the Purkinje cells was seem even when direct electrical stimulation of the parallel fibers themselves was used in place of a natural conditioned stimulus and stimulation of the olivary nucleus was used in place of a natural unconditioned stimulus (Jirenhed, Bengtsson et al. 2007; Hesslow, Jirenhed et al. 2013; Johansson, Jirenhed et al. 2013 MS). This discovery radically reduces the neural circuitry that is the focus of attention (Figure 1) and gives experimenters unprecedented control of the inputs to the Purkinje cell.

In the most recent experiments from the Hesslow laboratory, they electrically stimulate some portion of parallel fibers while recording from a Purkinje cell that is reading some portion of the stimulated fibers. In place of a periorbital shock for the US, they use direct stimulation of the olivary nucleus. Thus, they have direct control of the pre-synaptic spike train that carries the CS signal; each parallel fiber within the stimulating field follows the stimulating pulses one spike per one stimulus pulse. In this way, the experimenters directly determine the parallel fiber signal arriving at the dendrites of the Purkinje cell from which they record.

Before training, the stimulation-elicited spike train in the parallel fibers—the artificial CS—elicits an immediate, strong increase in the firing of the Purkinje cell. The increase rate of firing lasts as long as the CS spike train, and ceases abruptly when that input terminates, to be followed by a profound and prolonged reduction in the basal firing rate.

In their experimental protocol, they then “condition” (teach) the Purkinje cell by pairing stimulation of the parallel fibers (the artificial CS) with stimulation of the olivary nucleus (the artificial US). In different repetitions of the experiment, they use different ISIs, different intervals between the onset of parallel fiber stimulation and the onset of olivary stimulation.

The training profoundly alters the Purkinje cell’s response to the spike train in the parallel fibers. After training, the onset of the pre-synaptic spike train in the parallel fibers no longer elicits an increase Purkinje cell firing; rather, it elicits an almost complete pause in the cell’s basal firing. In other words, the training appears to convert an excitatory synapse into an inhibitory synapse, although I think that this will prove to be a misleading way of thinking about the phenomenon. This conversion in the apparent properties of the parallel-fiber-to-Purkinje cell synapses is not subtle; the training produces a huge sign-reversing change in the input-output characteristics of these synapses. (There is also a dramatic change in the duration of the pause in Purkinje cell that follows the offset of the CS.) Yhe Purkinje cell's post-training response to the artificial CS signal is both complex and radically different from its pre-training response.

More importantly still, the duration of the pause in Purkinje cell firing varies in proportion to the training ISI. And most importantly, the duration of this learned pause depends only on the training ISI, not on the duration of the presynaptic spike train that causes it. The duration of the Purkinje cell's firing pause does not vary in response to large (post-training) variations in the inter-spike intervals within the presynaptic spike train, nor to large changes in the overall duration of this spike train. The learned, well-timed post-training response of the Purkinje cell is the same when the pre-synaptic spike train is produced by a stimulus train lasting only 17.5 ms and containing 8 pulses (hence, with an inter-pulse interval of slightly more than 2 ms) as when the it lasts 800 ms and contains 81 pulses (hence, with an interpulse interval of 10 ms). In short, radically different synaptic inputs produce the same learned output from the Purkinje cell.

These results show that the temporal information acquired during the training experience—the remembered duration of the ISI—is expressed in the time course of the Purkinje cell's response to the onset of a pre-synaptic spike train, under circumstances where it is almost inconceivable that this temporal information is in the activating input (the pre-synaptic spike train) or in the synaptic conductances between the parallel fiber input and the postsynaptic Purkinje cell. All that the input appears to do is trigger the output; the temporal characteristics of the output are quite unrelated to the temporal characteristics of the input. The results would seem to imply that the acquired temporal information is not encoded in the synaptic connections between the parallel fibers and the Purkinje cell, much less in some complex pattern of synaptic weights, spread throughout some neural net. Rather, it is encoded by a change in some (molecular?) structure within the Purkinje cell itself.

In the post-training Purkinje cell, the onset of a pre-synaptic spike train causes the information in this intracellular structure to be read out into a spike train whose temporal complexity depends not at all on the temporal structure of the pre-synaptic spike train that activates the read out, but rather on: 1) an intracellular mechanism that has stored the temporal information acquired from experience; and 2) on intracellular machinery capable of converting the stored information into a complex output signal when activated by a simple input signal. The output conveys the intracellularly stored acquired information to the neurons in the deep nuclei on which the output axon of the Purkinje cell synapses.

This result is incomprehensible on the basis of the simple properties that neural net theorists imagine neurons to possess, which are those of a leaky integrator with a threshold on its output. On the other hand, this result is perfectly intelligible if one imagines that the physical basis of memory is not in the synapse but rather in information-storing changes in molecules inside neurons and if one further imagines that the neuron also contains the molecular level machinery necessary to read that stored information out into a complexly patterned spike train. This spike train is informed almost entirely by acquired information that has been stored inside the neuron rather than by the information conveyed to the neuron through its synaptic inputs or by the intrinsic dynamics of the neuron itself.

Given results this revolutionary in their implications, it is natural and appropriate to ask whether some other interpretation is possible. How sure can we be that it is the parallel fiber input that it critical to both the pre-training response and the radically different post-training responses of the Purkinje cell? In the top layer of the cerebellar cortex, one finds not only the dense parallel fiber system but also two other kinds of neurons, stellate cells and basket cells. Both of these make inhibitory synapses on the Purkinje cell. It is natural to wonder whether these inputs might somehow explain the appropriately timed pause in the post-training firing of the Purkinje cell, because it is possible, perhaps even likely, that the electrical stimulation of the parallel fibers stimulates some of these neurons as well.

To address this question, the Hesslow lab turned to another phenomenon observable in the same preparation: When one stimulates parallel fibers that are "off beam", that is, that do not synapse on the Purkinje cell from which one is recording, one observes a profound inhibition of the basal firing in the cell from which one is recording. There is reason to believe that this inhibition is mediated by either the stellate cells or the basket cells, both of which are known to make inhibitory synapses on the near dendrites and cell body of the Purkinje cell. When Hesslow and his colleagues inject a drug that blocks the action of the inhibitory transmitter, the inhibitory effect of off-beam stimulation on the basal firing of the Purkinje cell is eliminated, but this drug injection has no effect on the cell's learned, well-timed response to the artificial CS. This is strong evidence against a role for these inhibitory cells in explaining the timing of the learned pause in the Purkinje cell's response.

More Evidence: Abrupt Changes in Hippocampal Frames of Reference

The evidence from the Hesslow lab is the most direct evidence that machinery for storing acquired information resides inside neurons rather than in the synaptic connections, and so does the machinery for reading out that information into a spike train. Less direct evidence comes from at least two other sources: 1) the learned signaling characteristics of the neurons in the hippocampus and associated structures and 2) learned alterations in presynaptic transmitter release from olfactory neurons.

The firing of neurons in the hippocampus and in other closely connected regions of the medial temporal lobe is dramatically dependent on previously acquired spatial and temporal information (see for recent review Gallistel and Matzel 2013). The firing of these neurons is not determined by what if anything the rat currently sees or hears or smells or feels. Rather, it is determined by the animal's location and orientation on its cognitive map, as computed by its brain from a variety of past sensory inputs (Gallistel and Matzel 2013)[2].

A location and an orientation are represented by systems of coordinates. But coordinates represent a location or orientation in an experienced environment only when they have been anchored to an experienced frame of reference. This anchoring is what endows a system of coordinates with a semantics, that is, with a specific spatial reference. The location and orientation specified by a set of coordinates depends on the learned frame of reference to which they refer. What is innate in the brain's system for representing the experienced geometry of its environment are systems of coordinates, the machinery that implements vector spaces. A vector space is a symbolic system that can in principle represent the geometry of an environment. For a vector space to represent an actually experienced space, it must be anchored to a frame of reference with the experienced environment. In the course of constructing its cognitive maps, the brain anchors systems of coordinates to many different frames of reference. In one frame, location may be specified by reference to a prominent white card on an otherwise black wall. In another frame, the same location may be signaled by reference to the geometry of the enclosure or by the geometry of the large space that contains the enclosure.

The firing of the head-direction cells, place cells, and grid cells that signal the animal's current location and orientation is anchored to different frames of reference, even for one and the same neuron. There are object-based frames of reference, enclosure-based frames of reference, and large scale (extra-enclosure) based frames of reference (Gallistel and Matzel 2013). The same grid or place cell or the same head-direction cell may signal location or direction within one of these frames of reference at one moment and a small fraction of a second later signal location or direction within a different frame of reference (Gothard, Skaggs et al. 1996; Gothard, Skaggs et al. 1996; Frank, Brown et al. 2000; Redish, Rosenzweig et al. 2000; Rivard and al. 2004; Diba and Buzsáki 2008; Derdikman, Whitlock et al. 2009). The astonishingly abrupt changes in the frame of reference within which the cell's firing specifies a location or direction is difficult to explain if one assumes that the acquired information about the geometry of the experienced environment is

[2] Recent work from Eichenbaum's laboratory (MacDonald, C. J., K. Q. Lepage, et al. (2011). "Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events." Neuron **71**(4): 737-749, Eichenbaum, H. (2013). "Memory on time." Trends in Cognitive Science **17**(2): 81-88.) shows that these cells also signal temporal location, that is, the current temporal distance from recent events that function as temporal landmarks in that they occur at a fixed (temporal) distance from other events of interest. Thus, these cells appear to signal the animal's spatio-temporal location in a spatio-temporal cognitive map.

encoded in complex patterns of synaptic strength spread throughout an extensive neuronal network (a so-called distributed representation). Just how difficult it is to explain these abrupt transitions on such a basis is hard to say, because the antecedent question of how patterns of synaptic strengths might encode environmental geometry has not been addressed. There are no theoretical proposals about how to embed a vector space in a set of synapses, only hand waves. An extremely abrupt (<80ms) change in a frame of reference is much easier to explain if one assumes that the acquired information about the geometry of the animal's environment is stored within the cell itself; in other words, if one assumes that the spatio-temporal map is in the neuron itself rather than in a neural circuit.

I digress here to emphasize the following point: whereas there are no theories about how geometric information might be stored in a pattern of synaptic conductances, we know very well how information of any kind might be stored in DNA. The structure of DNA permits the storage of information at 2 bits per nucleotide, because any of the 4 nucleotides may follow any other in the sequence of nucleotides in a DNA or RNA molecule. A single nucleotide is approximately 1/3 of a nanometer in length. Therefore, DNA stores information at a linear density of 6 bits per nanometer. (To return for a moment to the consideration of size, the width of a synaptic cleft is about 20 nm; the diameter of the presynaptic vessicles that package neurotransmitters for release from presynaptic terminals is 35 nm.) A basic truth of computer science is that a medium suited to the storage of one kind of information is suited to the storage of any kind of information. When it comes to information storage and transmission, information is information; it's all just bits. That is why even poems can be stored in bacterial DNA (Gardiner 2010). That is why there are laboratories actively exploring the use of DNA as the memory component in a future computing machine (Team 2010; Goldman, Bertone et al. 2013). If and when DNA becomes the memory component of a computing machine, geometric information will be stored in it in essentially the same way it is now stored in the memories of the servers that you access when you use Google maps. In short, there is no mystery about how to store geometric information in the structure of DNA-, or RNA-like molecules, whereas there is a profound mystery about how to store information in synapses.

Still More Evidence: Learned, Selective Immediate Enhancement of Neurotransmitter Release from First Order Olfactory Neurons

When we grasp the fact that acquired information may be stored in a complex molecular computing machine inside each neuron, there is no reason not to assume that this occurs in every kind of neuron, including sensory neurons and motor neurons. Sensory neurons may use acquired information to help them interpret the information picked up by their transducer elements. That is, the process of interpreting current sensory input in the light of previous experience may begin within the first order sensory neurons themselves. Recent quite astonishing findings from the laboratory of my colleague, John McGann, suggest just that.

The McGann laboratory brings state-of-the-art neurobiological visualization methodology to bear on the question of how the brain represents olfactory input. We do not experience smells as a meaningless sensations; rather, they are freighted with learned significance: We smell bacon or the sea or manure or eucalyptus or the odor of a loved one. (One is reminded of Napolean's famous epistolary admonition to Josephine: "Coming home in three days; don't bathe.") It seems likely that the same is true for non-human animals, perhaps even more so than for us, as odor plays a larger role in the sensory/perceptual life of many animals than it does in ours.

However, until recently, the study of olfactory perception was a neurobiological and psychophysical backwater. This changed with the advances in the understanding of olfactory neuroanatomy consequent upon the discovery of the molecular biology of olfactory transduction (Mombaerts, Wang et al. 1996; Su, Menuz et al. 2009).

From a functional/computational standpoint, a basic property of any sensory system is the number of distinct channels that are operative. Each functionally distinct channel filters the stimulus in a different way and adds a degree of freedom (a dimension in a vector space) to the brain's representation of that stimulus. The scotopic visual system, which operates in dim light, has only one channel; the photopic system, which operates in brighter light, has three; the auditory system has thousands. It turns out that the olfactory system has hundreds. Each different functional olfactory channel is composed of neurons that express one and only one of the hundreds of different olfactory receptor molecules in the receptor end of the sensory neuron in the olfactory mucosa. Remarkably, all of the neurons that express the same receptor in their mucosal transducer portion project their signal-carrying axons to one or two glomeruli in the brain's olfactory bulb. Glomeruli are small spherical synapse-rich structures in the olfactory bulb. Each glomerulus receives projections from only one odor channel. Thus, the functional unit—the olfactory channel—maps to an anatomical unit—the glomerulus. Every different odorant creates a different pattern of activation of the olfactory glomeruli. For any given odorant, most glomeruli are inactive, but a few show a pattern of activation in which there is odorant-specific variation in the relative strengths of the activation.

McGann's laboratory visualizes these activation patterns in mice both before and after they have been trained with different odorants as discriminative stimuli. One of the odorants the mouse sniffs during training predicts shock; the other odors do not. McGann and his students find that this training selectively increases neurotransmitter release from the presynaptic endings of the first-order olfactory neurons synapsing on the glomeruli encoding the shock-predicting odor. In other words, information gained from an experienced predictive relationship between that odor and a fear-inducing shock finds its way to the presynaptic endings of the first stage olfactory neurons. This acquired information selectively alters their signaling at the point where they pass on to the rest of the brain the information they have gleaned from the odorant molecules currently binding to their receptors in the olfactory mucosa.

Remarkably, when the spectrum of glomeruli activated by the predictive odorant overlaps to some extent with the spectrum activated by a non-predictive odorant, the enhanced neurotransmitter release in the glomeruli in the intersect is specific to the predictive odorant. The release of transmitter caused by the non-predictive odors in those glomeruli is not enhanced; only the release produced there by the predictive odor is enhanced.

As in most sensory systems, there are extensive efferent projections from higher levels of the brain to the synaptic endings of these sensory neurons. Thus, there is no neuroanatomical mystery as to how the information acquired from the experience of the predictive relation between a given odor and shock may reach these pre-synaptic endings. There are, however, two quite different stories that one may imagine about how information conveyed by these efferents comes to inform the release of neurotransmitter from those endings. On one hypothesis, the information about the predictive relation between the one odorant and shock is not stored in the presynaptic endings themselves; on the other hypothesis, it is.

On the one hand, one may imagine that the acquired information about the predictive relation between odor and shock is stored more centrally in the brain. A connectionist would imagine that the acquired information is stored in some distributed pattern of synaptic conductances in some complex circuit, perhaps located in the amygdala, which is known to play an important, but ill-defined role in fear conditioning, or perhaps in the neocortex. On this story, each time an odorant evokes a spike train in the first order sensory neurons, the first few spikes in this train cause postsynaptic activity that propagates to the complex central circuit in which the information acquired from the training experience is stored. These initial afferent signals activate the complex central circuit is such a way as to cause it to generate an efferent signal that propagates back to the endings of the first-order sensory neurons. This efferent recognition signal enhances the release of neurotransmitter by later portions of the sensory spike train. On this story, the predictive significance of the sensory signal is recognized centrally—as has always been assumed.

A different possibility —until recently, almost unthinkable—is that when computations on the temporal map of past experience (Balsam and Gallistel 2009; Balsam, Drew et al. 2010) reveal the predictive relation between a specific odor and shock, this information is relayed to the presynaptic endings of the first-order neurons to be stored there. Then, as in the cerebellar circuit studied in the Hesslow lab, this intracellularly stored acquired information alters the release of neurotransmitter by the odorant-induced spike train. In this way, the signal passed on from the first-order sensory neurons to the postsynaptic circuit is already partially interpreted in the light of the predictive relation revealed by previous experience.

On the first hypothesis, which almost any neuroscientist would judge to be far more plausible, the selective enhancement of neurotransmitter release from the presynaptic endings of the first-order olfactory neurons can occur only some while after the onset of the odorant-evoked spike train in the first-order sensory neurons, because it depends on real-time feedback from the central circuits where the recognition of the signal's predictive significance occurs. On the second hypothesis, by contrast, the enhancement of neurotransmitter release can occur at signal onset, because it does not depend on real-time feedback. It depends instead on locally stored information conveyed to the presynaptic endings by earlier "off-line" feedback. This earlier off-line feedback came from the more central structures that computed the predictive relation from a time-stamped record of past events (Balsam and Gallistel 2009).

In fact, the enhancement that McGann's lab observes is present throughout the signal. As best they can determine, it is already there at signal onset. If the evidence for the immediate enhancement of transmitter release holds up, it strongly favors the second hypothesis, the local, intracellular storage of acquired information. It will be interesting to see just how much information is stored locally at that earliest possible stage of sensory signal processing, and at what level of abstraction.

In short there is now evidence that acquired information relevant to the interpretation of sensory signals may be stored within the sensory neurons themselves. One wonders whether the evidence for learning at the spinal level (Windhorst 2007; Wolpaw 2007) will lead to the discovery that acquired information relevant to the regulation of muscle activation and joint control is stored within the motor neurons themselves.

Back to Jerry

What I lay at Jerry's door is an insight that—in the fullness of time—may transform neuroscientists conceptual framework in ways as profound as the transformation in biochemists' conceptual framework wrought by the identification of the molecular structure of the gene. Jerry realized that there must be symbols in the brain, just as Mendel realized that there must be physically mysterious "particles" in seeds, particles that carried heritable information from generation to generation, quite independently of whether the information they carried was expressed in the observable structure of the organisms produced in any one generation. The physical realization of the symbols that carry acquired information is at this time as mysterious as was the physical realization of Mendels' particles. Like Mendel's particles, the information carried by these symbols is often not expressed in behavior. Jerry also realized that there must be computational machinery that operates on those symbols, the machinery that embodies the syntax. He realized, in other words, that the brain must have a language in exactly the sense in which a computing machine has a language. This was a truly profound insight, which is, of course, why it has also generated so much debate. The old ways of thinking die hard, very hard. To paraphrase Planck, science progresses one funeral at a time.

If, as I expect, Jerry's insight comes to inform the foundations of neuroscientific thinking, there will be great ironies. Jerry is conspicuous among cognitive scientists for his indifference to the question how the language of thought might be implemented in the brain. He commented on neurobiologically-inspired theories of cognition only so far as to point out that they lacked the productivity, systematicity and compositionality that are seen in a machine that has a language. He and Zenon Pylyshyn rightly argued that these properties were such salient properties of thought that any model of thought that denied these properties, either explicitly or implicitly, as neurobiologically inspired cognitive theories generally did, was clearly untenable in the face of the behavioral evidence (Fodor and Pylyshyn 1988). Jerry was sublimely indifferent to the protests from some philosophers and many psychologists and cognitive psychologists that there was no neurobiological foundation for the language of thought. Like the classical geneticists who were unperturbed by the biochemists claims that the gene was biochemically incomprehensible, Jerry believed in the implications of the data he knew. He was unperturbed by the neuroscientists who took absence of neurobiological evidence to be evidence of neurobiological absence. What an irony it will be if the language of thought hypothesis becomes the key to understanding the neurobiology of cognition.

I am immensely excited by the prospect that Jerry's insight may finally begin to influence neuroscientific thinking. Until the recent discoveries that I have described, I thought there was no prospect that we would know the physical identity of the brain's symbols in my lifetime. I thought there was even less prospect that we would know the machinery that implemented its computational operations. I suspected that the answers were to be found at the molecular level within neurons, rather than at the circuit level, where neuroscientists have assumed they must lie and where, therefore, they have looked for them to little avail throughout my career. Until these recent discoveries, there was no neurobiological evidence in favor of the hypothesis that acquired information is stored intracellularly at the molecular level, where it is operated on by molecular level computational machinery. Now that there is at least some neurobiological evidence pointing in that direction, my hope is that the molecular biologists will jump in and begin a serious quest for the intracellular molecular biology of neural memory and computation.

References

Aydede, M. (1977). "Language of thought: The connecitonist contribution." Minds & Machines **7**: 57-101.

Balsam, P. and C. R. Gallistel (2009). "Temporal maps and informativeness in associative learning." Trends in Neurosciences **32**(2): 73-78.

Balsam, P. D., M. R. Drew, et al. (2010). " Time and Associative Learning." Comparative Cognition & Behavior Reviews **5**: 1-22.

Bao, S., L. Chen, et al. (2002). "Cerebellar cortical inhibition and classical eyeblink conditioning." Proceedings of the National Academy of Sciences **99**: 1592-1597.

Carroll, S. B. (2005). Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom. New York, Norton.

Cerminara, N. L. and J. A. Rawson (2004). "Evidence that climbing fibers control an intrinsic spike generator in cerebellar Purkinje cells." Journal of Neuroscience **24**: 4510–4517.

Christian, K. and R. Thompson (2003). "Neural substrates of eyeblink conditioning: acquisition and retention." Learning and Memory **10**(6): 427–455.

Derdikman, D., J. R. Whitlock, et al. (2009). "Fragmentation of grid cell maps in a multicompartment environment." Nature Neuroscience **12**: 1325–1332.

Diba, K. and G. Buzsáki (2008). "Hippocampal Network Dynamics Constrain the Time Lag between Pyramidal Cells across Modified Environments." Journal of Neuroscience **28**(50): 13448 –13456.

Eichenbaum, H. (2013). "Memory on time." Trends in Cognitive Science **17**(2): 81-88.

Fodor, J. A. (1975). The Language of Thought. Trowbridge, Wiltshire, UK, Crowell Press.

Fodor, J. A. and Z. Pylyshyn (1988). "Connectionism and cognitive architecture: A critical analysis." Cognition **28**: 3-71.

Frank, L. M., E. N. Brown , et al. (2000). "Trajectory encoding in the hippocampus and entorhinal cortex , ." Neuron **27**: 169–178.

Gallistel, C. R. and J. Gibbon (2000). "Time, rate, and conditioning." Psychological Review **107**(2): 289-344.

Gallistel, C. R. and A. P. King (2010). Memory and the computaitonal brain: Why cognitive science will transform neuroscience. New York, Wiley/Blackwell.

Gallistel, C. R. and A. P. King (2010). Memory and the computational brain: Why cognitive science will transform neuroscience. New York, Wiley/Blackwell.

Gallistel, C. R. and L. D. Matzel (2013). "The neuroscience of learning: Beyond the Hebbian Synapse." Annual Review of Psychology **64**: 169-200.

Gardiner, B. (2010). "Recombinant rhymer encodes poetry in DNA." Wired(April).

Gehring, W. J. (1998). Master control genes in development and evolution : the homeobox story. New Haven, Yale University Press.

Goldman, N., P. Bertone, et al. (2013). "Towards practical, high-capacity, how maintenance information storage in synthesized DNA." Nature.

Gothard, K. M., W. E. Skaggs, et al. (1996). "Dynamics of mismatch correction in the hippocamapal ensemble code for space: Interaction between path integration and environmental cues." Journal of Neuroscience **16**(24): 8027-8040.

Gothard, K. M., W. E. Skaggs, et al. (1996). "Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task." Journal of Neuroscience **16**(2): 823-835.

Grossberg, S. and N. A. Schmajuk (1989). " Neural dynamics of adaptive timing and temporal discrimination during associative learning. ." Neural Networks **2**: 79-102.

Halder, G., P. Callaerts, et al. (1995). "Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*." Science **267**: 1788-1792.

Harvey, R. J. and R. M. A. Napper (1991). "Quantitative studies on the mammalian cerebellum." Progress in Neurobiology **36**: 437–463.

Hesslow, G., D.-A. Jirenhed, et al. (2013). "Classical conditioning of motor responses: What is the learning mechanism?" Neural Networks **47**: 81-87.

Jirenhed, D. A., F. Bengtsson, et al. (2007). "Acquisition, extinction, and reacquisition of a cerebellar cortical memory trace." Journal of Neuroscience **27**(10): 2493–2502.

Jirenhed, D. A. and G. Hesslow (2011). "Learning stimulus intervals—adaptive timing of conditioned Purkinje cell responses." The Cerebellum **10**: 523–535.

Jirenhed, D. A. and G. Hesslow (2011). "Time course of classically conditioned Purkinje cell response is determined by initial part of conditioned stimulus." Journal of Neuroscience **31**: 9070–9074.

Johansson, F., D.-A. Jirenhed, et al. (2013 MS). "Memory trace and timing mechanism localized to cerebellar Purkinje cells."??[got it from Denis Peli.

Krupa, D. J., J. K. Thompson, et al. (1993). "Localization of a memory trace in the mammalian brain." **260**: 989-991.

Laurence, S. and E. Margolis (1997). "Regress argeuments against the language of thought." Analysis **57**(1).

MacDonald, C. J., K. Q. Lepage, et al. (2011). "Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events." Neuron **71**(4): 737-749.

Martin, S. J. and R. G. M. Morris (2002). "New life in an old idea: The synaptic plasticity and memory hypothesis revisited." Hippocampus **12**: 609-636.

Matell, M. S., W. H. Meck, et al. (2003). "Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons." Behavioral Neuroscience **117**(4): 760-773.

Meck, W. H., T. B. Penney, et al. (2008). "Cortico-striatal representation of time in animals and humans." Current Opinion in Neurobiology **18**: 145-152.

Mombaerts, P., F. Wang, et al. (1996). "The molecular biology of olfactory perception." Cold Spring Harbor Symposia on Quantitative Biology **61**: 135-145.

Quiring, R., U. Walldorf, et al. (1994). "Homology of the *eyeless* gene of the *Drosophila* to the *small eye* gene in mice and *aniridia* in humans." **265**: 785-789.

Redish, A. D., E. S. Rosenzweig, et al. (2000). "Dynamics of hippocampal ensemble activity realignment: time versus space." Journal of Neuroscience **20**: 9298–9309.

Rivard, B. and e. al. (2004). "Representation of objects in space by two classes of hippocampal pyramidal cells." Journal of General Physiology **124**: 9–25.

Schneider, S. (2009). "LOT, CTM, and the elephant in the room." Synthese **170**(2): 235-250.

Shubin, N., C. Tabin, et al. (2009). "Deep homology and the origins of evolutionary novelty." Nature **457**(7231): 818-823.

Su, C.-Y., K. Menuz, et al. (2009). "Olfactory perception: Receptors, cells, and circuits." Cell **139**(1): 45-59.

Team, H. K.-C. (2010). Bacterial based storage and encryption device, Chinese University of Hong Kong iGEM.

Windhorst, U. (2007). "Muscle proprioceptive feedback and spinal networks." Brain Research Bulletion **73**.

Wolpaw, J. R. (2007). "Spinal cord plasticity in acquisition and maintenance of motor skills." Acta Physiolologica **189**: 155–169.

Yamazaki, T. and S. Tanaka (2009). " Computational models of timing mechanisms in the cerebellar granular layer." The Cerebellum **8**: 423–432.