Efficiency, Plasticity, and Error:

A Critical Analysis of Memory

Jared Jolton

University of Colorado at Boulder

Word Count: 2644

Author Note

Jared Jolton, Department of Psychology, Department of Computer Science, Institute of Cognitive Science, University of Colorado at Boulder.

This research was completed as part of the requirements for the Fall 2015 PSYC 3005 Cognitive Science course at the University of Colorado at Boulder.

Memory encoding and recall is essential for all cognitive functioning. Without memory, we would be incapable of learning, our social interactions would be meaningless, and our ability to carry out both complex and trivial tasks alike would be dramatically hindered. Yet, this cardinal process is largely out of our control. The fact - somber as it may be - is that not all memories endure, and the ones that do are often fragile and fallacious. This paper will discuss several aspects of memory that make it specifically prone to error, falsification, and deletion.

For the brain to function properly, the patterns of spreading neural activation that travel from region to region must be precisely defined. When any given presynaptic neuron emits an action potential, there are several factors that determine whether or not a downstream postsynaptic neuron will consequently reach its own electrical threshold - causing it too to fire an action potential. These factors specify the precise patterns of spreading neural activation that facilitate complex functioning. Connectionist models provide a simple way of conceptualizing this phenomenon. In neural network models, weights are used to represent the connections between two neurons. As such, when any upstream neuron is activated, the downstream neurons connected to it will be activated to an extent determined by the strength of the weights that connect them. Stronger weights for any connection will correspond to a higher likelihood that the activation of the presynaptic neuron will consequently activate the postsynaptic neuron. Critically, the weights in these models are dynamically updated, allowing the model to "memorize" the correct patterns of activation.

When examining the brain, this abstraction becomes much more complex, as there are many factors that contribute to updates of the actual "weights". When a presynaptic neuron fires, it releases the excitatory neurotransmitter glutamate into the synaptic cleft. This in turn causes

the membrane potential of the postsynaptic neuron to become elevated. The positive charge of the neuron forces positively charged magnesium ions out of the cell, unblocking the NMDA receptors, which allow the cell to translate chemical signals into electrical potentials. Upon receipt of the glutamate molecules, the NMDA receptors facilitate a critical process - the influx of calcium ions into the neuron. These calcium ions play an essential role in the strengthening of the connection between two neurons, as they facilitate the reorganization of the AMPA channels on the postsynaptic neuron. (O'Reilly, 2014) Poorly structured AMPA channels make it unlikely that the synapse will be conductive, whereas efficiently structured channels likely facilitate action potential transmission each time the presynaptic neuron fires. The location, number, size, and efficacy of these channels correspond directly to the weight associated with the connection between two neurons in a connectionist network. Importantly, this entire process is reliant on *both* the presynaptic neuron and the postsynaptic neuron, confirming the computational and theoretical basis of Hebbian learning: Neurons that fire together, wire together (Hebb, 1949). The restructuring and reformation of the synapse is known as synaptic consolidation.

There are several caveats to the biological learning process discussed above, which will prove useful in later analysis. Depending on the total concentration of calcium ions that enter the cell, the AMPA channels will be altered in different ways. With low levels of calcium ions, the restructuring of the postsynaptic AMPA channels actually weakens the connection between the two neurons, causing the postsynaptic neuron to experience Long Term Depression (LTD). This means that the next time the presynaptic neuron fires an action potential, the postsynaptic neuron will be less likely to also fire. Conversely, high levels of calcium ions always strengthen the connection, as they cause the neuron to undergo Long Term Potentiation (LTP) (O'Reilly, 2014).

The fact that the synapse can be both strengthened and weakened is an important facet of synaptic consolidation, as it entails a relative impermanence in the memory trace.

While these neuronal changes are undoubtedly responsible for facilitating memory formation, higher order processes are required for efficient storage and recall. As detailed earlier, memories are represented in the brain as complex patterns of neural activation. Initially, these patterns are rapidly formed in the hippocampus, using synaptic consolidation (Winocer, 2011). Certain kinds of memories, such as autobiographical or episodic memories are often retained in the hippocampus. These memories encode simple relationships, because they never get the chance to distribute their information to upstream neocortical areas. A clear example of these simple relationships is found in the "place cells" identified in the rat hippocampus that directly encode the location of the animal in its environment (O'Keefe, 1979). While these memory cells encode nothing more than simple spatial information, they can be used to facilitate complex cognitive tasks, such as the formation of cognitive maps and consequent localization within the environment.

Not all memories can be maintained in the hippocampus, however. In order for abstract and complex semantic memories to be derived and retained, they must be consolidated and transferred to other brain areas. Hippocampal damage does not seem to affect semantic memories, providing evidence of their storage in other neocortical areas (Winocer, 2011). The process by which these memories are systematically consolidated can be modelled computationally using deep neural networks. These networks utilize *hidden* layers, which are completely cut off from interaction with the environment. They do not receive inputs from the environment, and cannot drive outputs - they are connected only to the other layers of the

network. Because of this relationship, they form highly abstracted, distributed representations of upstream layers, which proves to be incredibly useful in storing large amounts of information with relatively few neurons, and deriving incredibly specific statistical information using self-organizing learning (O'Reilly, 2014). This is the same mechanism employed by the brain to transfer memories out of the hippocampus. As the memories are consolidated into the neocortex, the hidden layer representations become more and more abstract, allowing for the memories to encode complex semantic information. This process is known as systems consolidation, and the memories it creates are robust, well-formed and resilient. However, encoding such complex memory traces is no easy task, and not all information ever makes it to this final stage.

In fact, most information the brain perceives never even makes it to the initial step of synaptic consolidation. Any being in a complex environment *must* identify and attend to salient information. With a constant stream of information being processed at all times, it is highly inefficient to encode all incoming information. Rather, the brain must carefully select what is to be remembered. The biology of the synapse ensures that this is always the case. Only the most salient or attended to information will trigger enough downstream activity to encourage LTP, while the low levels of activation associated with irrelevant inputs will produce LTD, preventing synaptic - and thus systematic - consolidation from ever occurring. In fact, the LTD will make it harder to remember the irrelevant information, because it decreases the chance that weakly attended to stimuli will be encoded at all. This could prove very problematic in many situations, as essential information might accidentally be glossed over, making it harder to remember. The memory system is thus indirectly reliant on the attentional system. While this relationship is efficient, it practically ensures that important information will never make it into memory.

At the neural level, the restructuring of AMPA channels, the corresponding LTP and LTD, and the consolidation of a memory trace are largely capable of explaining memory formation. However, at the molecular level, these abstracted representations no longer provide sufficient explanations of the underlying mechanisms at play. Studies (Drier, 2002; Frankland and Josselyn, 2013; Hardt, 2009; Kwapis, 2014; Li, 2011; Ling, 2002; Pauli, 2012; Shema, 2011; Stebbins, 2002) have identified brain-specific enzymes which seem to be of paramount importance to the maintenance of synaptic memory. Tinkering with these enzymes in many different ways produces significant challenges for short-term memory in particular - though their effects on the synapse make them of unparalleled importance at all stages of memory formation. One enzyme specifically, protein kinase M-Zeta (hereafter referred to as PKM-ζ), appears to have particularly dramatic effects on the process. Even without observing its effects, the importance of PKM-ζ is confirmed by noting that it is synthesized in response to synaptic activity, and that the mRNA that encodes it is usually present in the synapse (Pauli, 2012). Conveniently, this enzyme can be inhibited with an amino-acid sequence known as zeta inhibitory peptide (ZIP), allowing researchers to study PKM-ζ's effects by observing performance on memory related tasks in its absence. When introduced to the hippocampus, ZIP actually reverses the synaptic effects of LTP, and has been proven to eradicate many different kinds of memories in animals (Frankland & Josselyn, 2013). Conversely, introducing PKM-ζ actually seems to enhance memory maintenance, as documented in odor memory studies using fruit flies (Drier, 2002) and taste aversion studies using rats (Shema, 2011). While the precise effects of the enzyme are not yet completely understood, it is apparent that PKM-ζ directly affects the number of AMPA channels created during LTP (Ling, 2002). Through a complex

mechanism, PKM-ζ supports the trafficking of AMPA channels to their precisely specified location on the synapse, facilitating synaptic efficacy (Kwapis, 2014). It appears that without PKM-ζ, we would be incapable of performing the neuronal adaptations required for memory formation and learning as a whole.

Knowledge of this enzyme and the mechanisms by which it supports memory is of the utmost importance for modern memory research. Being able to carefully control this enzyme and its inhibitors would afford us an unprecedented level of control over our own minds. Certain memories can interfere with our ability to function properly in daily life. For example, the long lasting memories associated with drug use make addicts incredibly prone to relapse, as these associations are very well-formed and can persist for months or years, long after they were encoded. By injecting ZIP into the nucleus accumbens - an area that receives input directly from the hippocampus - Li et al. were able to effectively eliminate this drug reward memory in rats (Li, 2011). Eliminating these memories in humans would reduce the chance of relapse, and would make abstinence significantly easier. Many of the memories and associations we form affect us negatively, such as traumatic memories of violent events, preferences for high-fat foods, or unhealthy false beliefs about oneself. Being able to target and remove these memories with ZIP would allow us to shape our minds in healthy ways, making our lives easier and our memories more useful and beneficial.

However, identifying a particular memory and the neurons that represent it would be incredibly difficult. There are more than quite a few neurons in the brain, and while guessing at the total number is trivial, it is largely agreed upon that the brain is made up of billions of neurons, and that these neurons exist as parts of networks consisting of trillions of synaptic

connections. As these connections are the basis of memory, their vast number represents a seemingly infinite capacity for information storage. Complex memory tasks, such as the memorization of the entire Japanese *hiragana*, can be simulated with simple neural network models utilizing very few artificial neurons (Jolton, 2015). The fact that only a few units are required for tasks like this hints at the endless capacity of the brain.

While the infinite nature of memory might seem beneficial, its immensity makes it intrinsically a difficult thing to maintain. Not only is memory highly influenced by unconscious, uncontrollable factors like PKM-ζ, it is also utilized by many different brain areas to complete various tasks. Speech sounds, for example, require many different memory representations. The memory of a phoneme (or in the case of *hiragana*, a *mora*), must, at the very least, encode the individual frequencies that make up the sound, the distribution of these frequencies across the length of the sound, the motor movements required to produce the sound, and the syntactic information required to properly detect and use the sound in speech. As speech sounds are perceived and utilized, different brain areas require access to these memories - the memory trace is activated many times from many areas. As such, the memory trace often requires reconsolidation (Stickgold, 2007). This process is prone to error, as the stimuli being processed at the time of the reconsolidation could form falsely generated associations with the memory trace - a result of the groups of implicated neurons firing at the same time. This phenomenon provides a simple explanation for the importance of sleep in consolidation, as during sleep there is little risk of intruding stimuli. During consolidation and reconsolidation, the number of neurons required to actively maintain the trace grows, the synaptic connections that support

recall become more and more specific, and as recall performance from all brain areas increases, the ability of the network to generalize to novel stimuli diminishes (McLelland, 2002).

The massive size of our neural networks and the necessity for activation to flow throughout forces us to shape our minds based on our memories. The more information we encode, the better we get at recognizing and remembering similar information. Unfortunately, this network specialization process makes it significantly harder to recognize and remember new information. For example, face recognition studies have shown that as we become familiar with faces of our own race, we develop difficulty differentiating between the faces of people from other races (Walker, 2003), and phoneme discrimination studies have shown that we have trouble discriminating between foreign phonemes that are not present in our native language (McLelland, 2002). Computational neural networks struggle with similar generalization problems (Jolton, 2015).

Considering the incredibly complex process required to encode, maintain, and recall a simple memory makes it difficult to acknowledge the many shortcomings of the memory system. Yet, the aforementioned cases of accidental inattention, generalization issues, and false associations made during reconsolidation are but a few examples of problems that arise due to the very nature of the memory process. Perhaps it is most concerning that false beliefs about autobiographical events can be directly implanted into memory with surprising ease. Studies have shown that participants can easily be made to report false beliefs about their own childhoods, simply by making them think about manufactured events (Loftus, 1997). In situations like this, the original memory trace is activated, immediately implicating reconsolidation. At the same time, neurons that represent the new, false event are activated,

strengthening the connection between the memory trace and the false belief. False beliefs, especially those relating to autobiographical memories, are very common, and depict a fundamental problem within the memory process - contiguous activations may elicit unwarranted memory formation, regardless of whether or not the contents of the encoded memory are valid.

Though memory may be invariably prone to error, it is relatively robust solution to a difficult problem. Ironically, its intelligent design produces its shortcomings. Even though the brain has enough space to encode a seemingly endless number of distinct memories, doing so would be highly inefficient, and would require implausible amounts of consolidation and reconsolidation. As a way of avoiding this unnecessary complexity, the brain forms fewer distinct representations that are molded and modulated over time and experience. While this allows for complex things to be learned, it means that memories are vulnerable, and can easily be altered. As memories evolve and change in the brain, their representations come to encode highly abstract information, but complete representations are difficult to maintain. Each time a memory is retrieved, the synapses that facilitate its storage undergo LTP and LTD, and the memory itself is undoubtedly changed in some way. Beyond the potentially negative effects of consolidation and reconsolidation, the actual patterns of neuronal activation that represent memory are inherently fragile. The maintenance mechanisms used to maintain these patterns, such as those performed by PKM-ζ, are unsustainable. Memory may be both efficient and plastic, but it is also inaccurate, prone to error, and highly malleable - these downfalls and their causes must be acknowledged.

References:

Drier, E.A., Tello, M.K., Cowan, M., Wu, P., Blace, N., Sacktor, T.C., Yin, J.C.P. (2002). Memory enhancement and formation by atypical PKM activity in Drosophila melanogaster. Nat. Neurosci. 5(4): 316--324.

Franklin, P., & Josselyn, S. (2013). Memory and the Single Molecule. *Nature*.

Hardt, O., Migues, P., Hastings, M., Wong, J., & Nader, K. (2009). PKMζ maintains 1-day- and 6-day-old long-term object location but not object identity memory in dorsal hippocampus. *Hippocampus*.

Hebb, D. (1949). *The organization of behavior; a neuropsychological theory.* New York: Wiley.

Jolton, J. (2015). A Neural Network Simulation of the Non-Native Phoneme Generalization Problem. https://github.com/2PacIsAlive/phoneme_classification_network

Kwapis, J., & Helmstetter, F. (2014). Does PKM(zeta) maintain memory? *Brain Research Bulletin,* 36-45.

Li, Y., Xue, Y., He, Y., Li, F., Xue, L., Xu, C., . . . Lu, L. (2011). Inhibition of PKM in Nucleus Accumbens Core Abolishes Long-Term Drug Reward Memory. *Journal of Neuroscience,* 5436-5446.

Ling, D., Benardo, L., Serrano, P., Blace, N., Kelly, M., Crary, J., & Sacktor, T. (2002). Protein kinase Mζ is necessary and sufficient for LTP maintenance. *Nat. Neurosci. Nature Neuroscience,* 295-296.

Loftus, E. (1979). Eyewitness Reliability. *Science,* 386-387.

Mcclelland, J., Fiez, J., & Mccandliss, B. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: Behavioral and neural aspects. Physiology & Behavior, 657-662.

O'Keefe, J. (1979). A Review of Hippocampal Place Cells. *Progress in Neurobiology, 13*, 419-439.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors (2014). *Computational Cognitive Neuroscience*. Wiki Book, 2nd Edition. URL: http://ccnbook.colorado.edu

Pauli, W., Clark, A., Guenther, H., O'reilly, R., & Rudy, J. (2012). Inhibiting PKM reveals dorsal lateral and dorsal medial striatum store the different memories needed to support adaptive behavior. *Learning & Memory,* 307-314.

Rudy, J. (2014). Variation in the persistence of memory: An interplay between actin dynamics and AMPA receptors. *Brain Research,* 29-37.

Shema, R., Haramati, S., Ron, S., Hazvi, S., Chen, A., Sacktor, T., & Dudai, Y. (2011). Enhancement of Consolidated Long-Term Memory by Overexpression of Protein Kinase M in the Neocortex. *Science,* 1207-1210.

Stebbins, M. (2002). SYNAPTIC PLASTICITY: Remember PKC? *Nature Reviews Neuroscience Nat Rev Neurosci,* 336-336.

Stickgold, R., & Walker, M. (2007). Sleep-dependent Memory Consolidation And Reconsolidation.*Sleep Medicine,* 331-343.

Walker, P., & Tanaka, J. (2003). An encoding advantage for own-race versus other-race faces. *Perception,* 1117-1125.

Winocer, G., & Moscovitch, M. (2011). Memory Transformation and Systems Consolidation. *Journal of the International Neuropsychological Society, 17*, 766-780.