Forgetting Curves Emerge from Dynamics of Integrated Memory

Eva Cadez (ecadez@ucmerced.edu) and Evan Heit (eheit@ucmerced.edu)
Cognitive and Information Science, University of California, Merced, 5200 North Lake Road
Merced, CA 95343 USA

Abstract

We present a Dynamical Integrated Memory Hypothesis (DIMH) and illustrate its use by arguing that forgetting curves are emergent properties of dynamical memory that includes decay and influences of complex context on memory traces. Because forgetting curves are emergent, it is not likely that a single simple function will be able to model them. Forgetting at different time scales is similar because similar dynamics occur at each scale and not because there is a single underlying mechanism that produces them. We argue that the dynamical systems approach is particularly suited for investigating systems that evolve, such as memory, at a very abstract level.

Keywords: dynamical systems; forgetting curves; memory.

Introduction

Forgetting trends, the initial fast loss of learned material and its slower decrease later, have been found in almost any memory task on any time scale (see Wixted & Ebbesen, 1991). Several mathematical functions can accommodate this general trend and therefore there is no consensus on whether an exponential, power, or some other non-linear function or combination of functions fits the data better than any other (e.g., Wixted & Ebbesen, 1991; Brown & Lewandowsky, 2010). We return to this issue later in the text. Based on these shape similarities, Brown, Neath, & Chater (2007) suggested that the same kind of memory mechanisms may underlie episodic memory over various time scales and tasks: probed serial recall, free recall, immediate recognition, forward serial recall, absolute identification task, etc.

After reviewing decades of research devoted to finding the exact shape of the forgetting curve, Brown and Lewandowsky (2010) reported varying parameters in the Scale Invariant Memory, Perception, and Learning (SIMPLE; Brown, Neath, Chater, 2007) and obtaining several generally similar but not identical forgetting functions. They showed that parts of this function can be well described with one kind of curve while other parts of the same forgetting curve are better described by some other function. The authors concluded that finding the shape of The Forgetting Function may be an unobtainable goal.

In what follows, we introduce a dynamical systems perspective on this issue which mostly agrees with Brown and Lewandowsky's (2010) conclusions about forgetting. We show how the dynamical system perspective may shed a new light on memory processes. We ask: Why would there be only one forgetting curve? The memory system is complex and dynamic; decay and other influences affect the evolution of every memory trace and the result is forgetting. The shape of the forgetting curve is an epiphenomenon of this process; there is no one static mechanism that always, necessarily underlies the power law or exponential law of forgetting per se.

Dynamical Systems

Dynamical systems are systems that evolve in time, whose future states are determined by initial states. Compare a dynamical system to a static model of changes in the intensity of some quantity over time. In such systems, history does not matter in any way. We can determine the value of one or more dependent variables of the system, based only on the values of one or more independent variables at a current point in time.

Hotton and Yoshimi (2011) define a dynamical system in the following way:

> Abstractly, a dynamical system is a function of the form $\varphi: S \times T \to S$. This function takes a state $s_0 \to S$ (which we think of as an initial condition) and a time $t \to T$ and returns the state the system will be in at time t starting from state s_0. This state can be written as $\varphi(s_0, t)$. To be a dynamical system the function $\varphi(s_0, t)$ must satisfy the two properties:
>
> - There is a time $t_0 \in T$ such that for all states $s_0 \in S$ $\varphi(s_0, t_0) = s_0$
> - For all states $s_0 \in S$ and all times $t_1, t_2 \in T$ $\varphi(s_0, t_1 + t_2) = \varphi(\varphi(s_0, t_1), t_2)$

Dynamical systems in this sense are deterministic, meaning that there is only one possible future state that follows from a specific current (initial) state. Here, we consider a specific class of dynamical systems, systems with memory, in which the evolution of the system explicitly depends on the past history of the system. Depending on the shape of a path leading the system to its current state, it will behave in different ways in the future.

Memory in general as a dynamical system

Dynamic systems typically are systems of a large number of elements. In case of memory, these elements are memory traces. The many elements of a dynamical system interact with each other in numerous ways, too. These interactions are very complex because the value of any trace's intensity in any point in time influences in a more or less direct way any other trace's intensity.

We do not make more specific assumptions about memory traces beyond the fact that they are the neural substrates of remembering (Brown, 1958). We present a formal mathematical model of evolution of an intensity of a memory trace. This model is abstracted from biological networks in the brain and is even more abstract than connectionist network models as described by Smolensky (1988). Though the model could shed light on how memory processes are implemented in the brain, it is strictly neutral on these issues.

In general, dynamical systems may be described by one or more differential or difference equations. This system of equations only rarely has an exact analytical solution which allows for detailed mathematical analyses of the system's behavior. Most physical dynamical systems are also very hard to implement in connectionist networks models due to their complexity. However, when these kinds of analyses are not possible, the dynamical systems approach offers other, qualitative methods that allow researchers to gain some insight into the system's behavior over time. These methods were pioneered by Henri Poincare in late 19th century (Farlow, Hall, McDill, & West, 2007) and they include analysis of state spaces, direction fields, attractors, etc.

At this time, we do not use these methods and we rely only on numerical solutions for simulations of the system's behavior based on the mathematical model described next in the text. We refer to all of the above mentioned approaches to investigation of systems as the dynamical systems approach. We treat the cognitive operations in episodic memory (learning and recalling a list of items, for example), as a dynamic process that can be mathematically modeled by suitable differential equations. The model we use in research is based on a similar procedure that is used in physics to describe transport processes of radiation.

The Dynamic Integrated Memory Hypothesis (DIMH) in episodic memory

The DIMH assumes that the change of the intensity of a memory trace depends on the intensity itself, on newly formed memories (such as those evoked by words presented in a list), and on the influence of other active memories that make a context for the evolution. Memory traces are parts of a dynamical system and are highly integrated. We think of this dynamical system in the same way as Bechtel (1997) does: The system "can have a form of weak modularity in which components make different contributions even while sharing information."

The DIMH considers time dynamics of memory in multidimensional parameter space. In what follows, we proceed with an intuitive description of the model while indicating the equivalent mathematical relations. To do this, we introduce relevant variables and parameters that will describe abstract objects of the considered dynamic process, and then we choose a mapping procedure to relate numerical values to them.

The first notion in the model is that of time, representing a parameter in the system: The system evolves in time. This can either be a real physical time t or it may be taken as an individual's psychological time. In the latter case their relation has to be additionally specified.

In the model, the parameter x represents the position of a concept in memory. For example, one can look at the x as a result of the following hypothetical experiment:

A person is given a stimulus, say a color red, and is asked to make a list of somehow associated concepts. The person is asked to place the associates in a sequence according to their sense of closeness or distance from the presented concept. In this way it is possible to place some objects to the left and some to the right of the original concept, depending on the individual's sense of mutual relationships between the associates. For example, colors green and orange may be at the opposite sides of red.

Thus, we can introduce the parameter x which defines the position of a concept relative to the initial associate located at some arbitrarily chosen position $x = x_0$. The quantity we can now call the distance between two concepts is therefore given by $x_0 - x$. Similarities between concepts are represented by similarities in their positions. It is obvious that the concept of distance here only weakly resembles distance in physics but some assumptions about relative positions and similarity may be reasonably put forward. The use of a one dimensional continuous parameter space to represent a multidimensional memory space is obviously a simplification and only a model of that space. Strictly speaking, some information is necessarily lost in doing this but the model is mathematically more convenient and computationally simplified this way, while there are justifications for the claim that it is still functional enough to model data and give insights into memory processes.

First, one may conceive that even though there are multiple relations between objects in multidimensional space, in reality we do not use all the information and relations pertaining to one object, for example, when we are engaged in serial recall of words. Therefore, the one dimension that may be used for a specific modeling task may be thought of as rotation of a multidimensional space to align the dimension of the model with a chosen single dimension along which the items differ in the most relevant way for the current task only. For a different task, the appropriate dimension used may be different. Second, for the purpose of this modeling, it is sufficient to determine whether the items in the task are facilitating recall of each other or not. The exact distances are, as mentioned, very intuitive in this kind of research unlike in, say, physics, but this is not too much of a simplification for modeling memory.

Hence, it seems reasonable that the described ordering of inter-item distances along the parameter dimension x in DIMH is acceptable for modeling, along with careful consideration of choices of elements of the model. As in any modeling work, one has to be aware of the implications of assumptions and approximations that are made.

The aim now is to model the time evolution of a memory trace's intensity. As the first approximation, we may accept the initial assumption that the memory intensity $I(t;x)$ spontaneously decays in time if no additional influences are present. As is common in numerous natural decay processes, we can assume the following rate of change:

$$\frac{d}{dt} I(t;x) = -\alpha(t;x) I(t;x) \tag{1}$$

where $\alpha(t;x)$ is known as the decay or attenuation constant for the considered time evolution whose reciprocal is the e-folding time. In general, we consider it as both time and parameter dependent quantity. Related to different objects, i.e. different positions x, the memory fades by different amounts and possibly at different rates. The memory attenuation amount or rate may vary in time, too. The specific form of functional dependence for the $\alpha(t;x)$ could be chosen according to the specific task under consideration.

This process is now generalized by introducing an external local source of excitation which affects the time evolution of the trace intensity I. This is done by adding an extra term, the source function $S(t;x)$ to the right hand side of Eq.(1):

$$\frac{d}{dt}I(t;x) = -\alpha(t;x)I(t;x) + S(t;x) \qquad (2)$$

This external input may be described by a Gaussian function to allow for the idea that one specific memory may also intensify very closely related memories to some extent. Finally, the influences of the context on the intensity at a specific position are added and the complete model equation is written:

$$\frac{d}{dt}I(t;x) = -\alpha(t;x)I(t;x) + S(t;x) + h + C(I;t,x) \qquad (3)$$

where:

$$C(I;t,x) \equiv \int_{-\infty}^{+\infty} dx' \int_0^t dt' W(t,t';x,x')I(t';x') \qquad (4)$$

is the spatial correlation represented by the integration over x'. The functional dependence W is the weight function determining influences of the temporal and spatial context at $(t';x')$ on the selected item at $(t;x)$. Its analytical expression has to be specified in the model either following some experimental results or by logical expectations about how the processes should run. For example, in the reported simulation we assume lateral inhibition at greater distances and excitation at smaller distances between items. In addition, we assume that small lateral activations x' and t' contribute less to the activation in x than greater activations (both negative and positive) while at the same time there is also a limit to the possible contribution of the large activations from other positions. We achieve this by using a Gaussian weighting function for distance contributions and a log function for intensity contributions from a site x' to the site x and from a point at t' to the point t. The third term in Eq.(3), h, is some constant level of global activation or resting level. We here refer readers to Erlhagen and Schoener(2002), for details on mathematics and parameter choices for simulations.

One issue with this model should be noted. The second integral in Eq.(4) indicates that the system has some explicit memory of its past history. The weighting function along the time axis, $W(t,t\prime)$ covers a finite time interval of an episode modeled and is of an appropriate shape so that the system is potentially implementable in the brain. However, as noted above, this is an abstract model and we are neutral about these implementation issues.

In sum, the change of the intensity of a memory depends on the intensity itself, on the added external memories such as the other words presented in a list, and on the influence of internal memories that make a context for the evolution. This mathematical model has been implemented in MATLAB to simulate various experiments in memory research. External inputs can be presented at different rates in time and at different positions corresponding to different stimuli and memories. This allows for simulations of many tasks used in memory research. Additional constraints on model parameters may be posited this way and the model itself may be a unifying account of many findings.

Evolution of integrated memory-forgetting

What phenomena may cause lowering of intensity of a memory trace? Most researchers mention at least one of three processes-interference, spontaneous decay, and reduced distinctiveness of older memories but there is no consensus on which of these are useful for modeling or should be investigated as natural phenomena. We mention them here very briefly due to space constraints but we use simulations to argue that simple decay needs to operate in order to account for effects of interference and distinctiveness.

It is obvious that spontaneous decay may be responsible for forgetting. In addition, changes may be due to influences of other stimuli like proactive interference and retroactive interference. Both refer to an effect that one learned material has on another one. In retroactive interference what is learned later influences the memory of previously learned material. In proactive interference current learning influences how the next material will be learned (e.g., Underwood, 1945; Postman 1961, Waugh & Norman, 1965). Further, currently activated context of a memory trace may interfere with it and thus impede the recall (Anderson, Bjork, & Bjork, 1994). The influence of the past events also changes based on present events. Finally, older memories may be harder to distinguish from one another so their recall is less accurate (Crowder, 1976; Brown et al. 2007).

In DIMH, one source of influence to any point x comes from the entire memory system. Consider the point at, say, $x = 170$, in Figure 1. Intensities of all other points along the x (at one point in time) are integrated with the one at $x = 170$. In addition to this, at the same point in time there may exist activations from external sources (S). This point also has a history, so its current value depends on the previous one, too (decay term, and the integration over time in the main equation). In other words, if a person has some initial distribution $I(t = 0)$ of memories and the memory evolves, what guides this evolution is the initial distribution (initial values for the main differential equation, old knowledge)), temporal and spatial context (inner dynamical input) and any new input (new knowledge). Different initial distributions influence the evolution in different ways and this is most closely related to the idea of interference in mentioned memory research.

Simulations

Using the following simulations we are going to argue that forgetting only happens if complicated dynamics is operating. The initial distribution is not going to change in time if there is no influence of any integration from context or if there is no decay. On the other hand if there is change, it must be the result of an interaction of the initial distribution and at least a decay or new inputs (which change the entire inner input to each x point), if not both. Hence, the forgetting emerges from the dynamics of interactions in dynamical memory.

Let us consider an evolution of two memory entities. Fig. 1 shows a simple case of evolution of two different and separate Gaussians. Each Gaussian is the activation of an entity in memory, which may be thought of in different ways, depending on applications of the model: a set of features, a set of words, or some other complex knowledge. One may think about these sets as representing the results of processing at different time scales. We use a Gaussian function and continuous memory space for mathematical convenience, not from a cognitive theory of how memories are distributed. Spontaneous decay of memory traces, interference from other memories and other possible sources of activation all integrated together guide the evolution of intensity of any activation, regardless of how a single Gaussian is interpreted, within this model. In following simulation (Fig. 1), all terms of the main equation are non-zero.

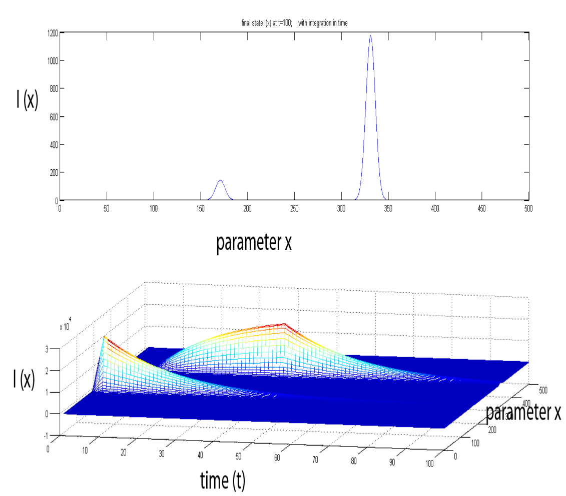

Figure 1: Top panel shows current state after evolution, the distribution over the parameter space x at the end of the evolution ($t = 100$). Bottom panel shows the evolution from $t = 0$ to $t = 100$.

The activations reach their peaks and if there is no new activation in their positions, they gradually lose intensity. Again, the intensity of a trace is calculated so that it includes all the influences of all other points along the x dimension, in accord with the main general equation. Units for intensity, the position dimension, and time are arbitrary and do not necessary map directly into familiar units used in experiments. However, they are on an at least interval scales which makes the appropriate trends visible. The accuracy of recall performance is directly proportional to the intensities.

Greater forgetting means lower magnitudes of remaining intensities, $I(x)$, after evolution. In Fig. 1, at the time $t = 60$ more x points have the activation above some level (which may be interpreted as a threshold for recall) than at the time $t = 90$. We here use relatively vague terms to allow for variations in interpretation in different forgetting research paradigms while still pointing out important similarities in dynamics of integrated memory. The amount of forgotten items, therefore, increases in time as a result of complex dynamics.

A role of spontaneous decay

We believe that spontaneous decay is necessary for the dynamics presented and perhaps is a necessary phenomenon in the brain. The following two figures illustrate the role of decay by showing the difference of evolution with and without the first term of general equation, the spontaneous decay term. The smaller Gaussian that was previously far away from the larger one is now very close, they overlap. In addition, another Gaussian is added- it is same as the larger one in Fig. 1 but begins at the same early time as the smaller one in Fig. 1.

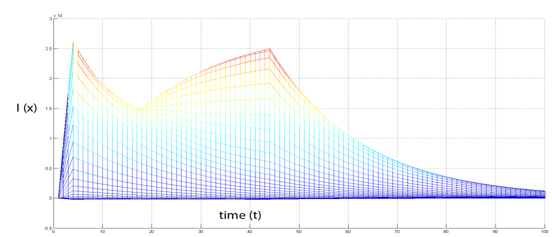

Figure 2: Evolution of three Gaussians ($x = 320$, $x = 330$, second $x = 330$) with decay.

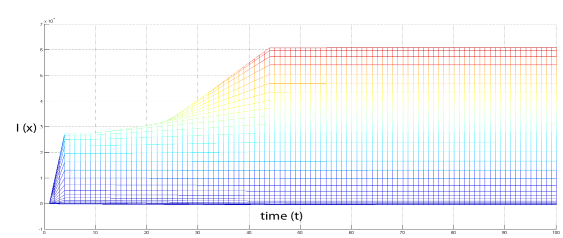

Figure 3: The same three Gaussians but without the decay term.

Three points evident from these simulations need to be stressed here. First, if there is no decay operating, the intensity of the larger Gaussian increases quickly over time, potentially reaching extremely large magnitudes. This problem is not necessarily present when decay operates. The decay is largely preventing extreme activations because it reduces higher intensities faster and lowers overall activations (and much more than lateral inhibition) so that new inputs do not increase the total activation of the system too much. Altmann & Schunn (2002) made the same points after using their Functional Decay Model to fit experimental data.

Second, consider Jost's second law- the rate of forgetting slows over time (Wixted, 2004). As Brown & Lewandowsky (2010) illustrated, intuitively we want that if two people remember 10 French words today but one of them has learned them three months earlier and the other one learned them three days ago, tomorrow they should remember significantly different amounts because the rate of change for them is different due to the different time passed from initial learning. This is not possible in the case presented in Fig. 3. If two people have the same configuration of their memory today, they will have the same configuration tomorrow regardless of when they learned the material. This means that under the same conditions for those two people the activation will change the same way for both.

In sum, on the one hand, decay in an integrated system, without new inputs is enough to produce changes in the distribution of memories because different intensities change differently. Therefore, the effect of decay, even if it is the only mechanism of forgetting, is non-trivial. The shape of decay, say an exponential function, is not necessarily the same as the forgetting curve.

On the other hand, it may be possible to argue that the interference from other events may mimic decay both in this and the previous case. We believe that this would be very hard to do at least in the first case presented above. However, even without decay processes, the evolution is still guided by changes in integrated influences of the total context, along with new inputs, on the initial distribution. In both cases, therefore, the forgetting curve is an emergent property of a dynamical system.

Third, temporal distinctiveness, as mentioned above, is not reduced in this model if decay is not present. For now it is hard to see how any other model might argue that it is possible to obtain the reduction unless some other mechanism that *decays* this distinctiveness is assumed.

These three points, we argue, mean that spontaneous decay is a necessary component of this model. Without decay, the model does not fit human data. Based on the preceding simulations using the DIMH, we further argue that the entire process of forgetting some initial distribution of active memories always follows complex dynamics where memory traces are integrated and most likely both spontaneous decay and interference from the context play significant roles.

Discussion and conclusions

There are several reasons why we believe DIMH as a novel approach may be a valuable addition in memory research.

First, it is a new kind of model that considers time evolution of a memory trace in episodic memory. It seems obvious that memory processes and their integration are complex and dynamical and this is why the dynamical system approach might be particularly useful, especially in combination with other approaches. Second, the use of differential equations to describe the change itself, rather than states, points out that each initial distribution, or old knowledge, develops slightly differently but that there may be common trends in the behavior. We believe that this stresses the value of simulations that are not necessary curve fitting and parameter estimation tasks. These kinds of analyses have their own value and should be combined with parameter estimation. Third, we show how a general equation produces behavior that is seen on many time scales. The parameters of equations on different time scales may still vary significantly but they produce similar behaviors. Most importantly, the emergent forgetting curves are similar to each other because the dynamics of many mechanisms on different time scales are similar and not because there is one mechanism producing the self similar forgetting curve such as the power law (Brown and Lewandowsky, 2010).

The DIMH uses a different level of explanation than most connectionist networks. Even though the equations of the mathematical model are very similar to the connectionist equations, they for the most part do not have the same interpretation. To clarify the way we interpret this model, it is helpful to consider Smolensky's (1988) distinction between three levels of cognitive explanation: neural, sub-symbolic, and symbolic.

At each of these levels dynamical models can be developed. A dynamical system at the neural level would describe a group of interconnected neurons in a region of a brain. A dynamical system at the sub-symbolic level is something like a connectionist network- a more abstract approximation (Smolensky, 1988) of the brain both in terms of structure and dynamics. It is based on general neural principles but abstracts from the details. A dynamical system at the conceptual level describes processing of symbols and relations between them, for example words and concepts in a natural language. Memory traces described by DIMH should probably be treated as falling somewhere in between Smolensky's symbolic and sub-symbolic levels.

Smolensky (1988) described the sub-symbolic level in relation to the symbolic level as follows:

> The interactions between individual units are simple, but these units do not have conceptual semantics: they are subconceptual. The interactions between entities with conceptual semantics-interactions between complex patterns of activity- are not at all simple. [Importantly,] Interactions at the level of activity patterns are not directly described by the formal definition of a subsymbolic model; they must be computed by the analyst.

In a sense, this is what we do. If the model proposed here was implemented in a connectionist network it would probably be most closely related to the Adaptive Resonance Theory

(Grossberg, 1976) where the DIMH is more general in some aspects.

The DIMH is not the same as Dynamical Field Theory (DFT). The DIMH and the DFT share some equations and ideas but are not the same. The DFT was developed by Erlhagen and Schoener (see Erlhagen & Schoener, 2002) and was based on Amari's mathematical generalization describing a behavior of a field of interconnected neurons (Amari, 1977). Mathematically similar models have been used in other areas of research to describe dynamics of neuronal activities (Wilson & Cowan, 1972; Grossberg, 1980), saccadic eye movement (Kopesz & Schoener, 1995), visual cognition (Johnson, Spencer & Schoener, 2008), and infant motor learning (Thelen, Schoener, Scheier, & Smith, 2000). However, earlier than work on DFT, Grossberg (1969) had generalized dynamics of memory and obtained similar equations to our own. In a sense, we continue this work, making an even more general model. In addition, while the DFT aims to describe cognition on connectionist networks levels, the DIMH operates at the symbolic level, as mentioned. Finally, in Shankar and Howard (TILT, 2010), the mathematics and modeling of decay of a memory trace is almost identical as ours but the dynamics of interactions are more generally described in the DIMH. The fact that the same mathematical model may be used to model physical phenomena such as decay, behavior of neurons, and behaviors of memory traces on various time scales might be a coincidence but perhaps looking into these similarities may help illuminate how these phenomena are connected.

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics, 27,* 77-87.

Altmann, E. M. & Schunn, C. D. (2002). Integrating decay and interference: A new look at an old interaction. *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp.65-70). Hillsdale, NJ: Erlbaum.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063-1087.

Bechtel, W. (1997). Dynamics and Decomposition: Are They Compatible? In *Proceedings of the Australian Cognitive Science Society*.

Brown, G. D. A., & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. In S. Della Sala (Ed.), *Forgetting* (pp. 49-75). Psychology Press.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539-576.

Brown, J. (1958). Some tests of the Decay Theory of immediate memory. *Quarterly Journal of Experimental Psychology 10*, pp. 12-21).

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Erlhagen W., & Schoener, G. (2002). Dynamic field theory of movement preparation. *Psychological Review, 109, 3,* 545-572.

Grossberg, S. (1969). On the serial learning of lists. *Mathematical Biosciences, 4,* 201-253.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, 1: Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 187-202.

Grossberg, S. (1980). Biological competition: Decision rules, pattern formation, and oscillations. In *Proceedings of the National Academy of Sciences (USA), 77,* 2338-2342.

Farlow, Hall, McDill, & West (2007). *Differential Equations and Linear algebra.*(2nd ed). New Saddle River, NJ: Pearson.

Hotton, S., & Yoshimi, J. (2011). Extending Dynamical Systems Theory to Model Embodied Cognition. *Cognitive Science 35, 3* 444-479.

Johnson, J. S., Spencer, J. P., & Schoener, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. In F. Garzon, A. Laakso, & T. Gomila (Eds.) *Dynamics and Psychology [special issue]. New Ideas in Psychology, 26,* 227-251.

Kopecz, K., & Schoener, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural, dynamic fields. *Biological Cybernetics, 73,* 49-60.

Postman, L. (1961). In: Cofer, C. N. (Ed.) *Verbal learning and verbal behavior: Proceedings of a conference sponsored by the Office of Naval Research and New York University,* pp. 152, McGraw-Hill, London.

Shankar, K. H., & Howard, M. W. (2010). Timing using temporal context. *Brain Research 1365,* 3-17.

Smolensky, P. (1988). On the proper treatment of connectionism. *The Behavioral and Brain Sciences, 11,* 1-23.

Thelen, E., Schoener, G., Scheier, C., & Smith, L. (2000). The dynamics of embodiment: A field theory of infant perseverative reaching. *Brain and Behavioral Sciences, 24,* 1-33.

Underwood, B. J. (1945). The Effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs, 59, 3.*

Wilson, H. R. & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysics Journal 12* 1-24.

Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review 111,* 864-879.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89-104.