To appear in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*
Hillsdale, NJ: Erlbaum, 2001.

Semantic Effect on Episodic Associations

Yaron Silberman (yarons@alice.nc.huji.ac.il)
Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem
Giv'at Ram, Jerusalem 91904 Israel

Risto Miikkulainen (risto@cs.utexas.edu)
Department of Computer Science, The University of Texas at Austin
Austin, TX 78712-1188 USA

Shlomo Bentin (msbentin@mscc.huji.ac.il)
Department of Psychology, The Hebrew University of Jerusalem
Mount Scopus; Jerusalem 91905 Israel

Abstract

We examined the influence of the pre-existing organization of the semantic memory on forming new episodic associations between words. Testing human subjects' performance we found that a semantic relationship between words facilitates forming episodic associations between them. Furthermore, the amount of facilitation increases linearly as a function of the number of co-occurrence of the words, up to a ceiling. Constrained by these empirical findings we developed a computational model, based on the theory of spreading activation over semantic networks. The model uses self-organizing maps to represent semantic relatedness, and lateral connections to represent the episodic associations. When two words are presented to the model, the interaction of the two activation waves is summed and added to the direct lateral connection between them. The main result is that the model is capable of replicating the empirical results. The model also makes several testable predictions: First, it should be easier to form an association from a word with few semantic neighbors to a word with many semantic neighbors than vice-versa. Second, after associating an unrelated word pair it should be easier to associate another two words each related to one of the words in the first pair. Third, a less focused activation wave, which may be the cause of schizophrenic thought disorder, should decrease the advantage in learning rate of related over unrelated pairs.

Introduction

The principles of forming associations between concepts in memory have been studied since the early days of psychological research. For example, the British empiricist school of philosophers (e.g., Hume, 1738), proposed three main principles of association: Contiguity (i.e., proximity in time and space), Similarity (or Contrast), and Cause and Effect. Fulfilling any of these conditions should be sufficient to form an association between concepts. The strength of the association is determined by the frequency at which any of the above conditions is fulfilled. An important aspect of this theory is that intentionality is not a necessary condition for the associative process to occur. Indeed, associations are frequently established without intention and without allocating attention to the learning process. We will refer to these associations as *incidental associations*. This paper presents a computational component of a larger study in which we examine characteristics of forming incidental associations between words.

Phenomenologically defined, two words are associated if the presentation of one brings the second to the perceiver's awareness. Associations between words can be formed in at least two different ways. First, *episodic associations* are formed when two words co-occur in time and space. An episodic association is therefore a subjective experience. Second, *semantic associations* are based on semantic relatedness between the words. Words are considered semantically related if they share common semantic features, for example, if they belong to the same semantic category. Although the two classes of associations are based on different properties, many associated word pairs are also semantically related, which raises the possibility of an interaction between the two types of associations.

A well-known interaction of this type is the semantic priming effect (Meyer and Schvaneveldt, 1971). The presentation of a related prime word prior to performing a lexical task, such as naming and/or lexical decision, results in faster and more accurate performance (see Neely, 1991, for a review). Ample research was aimed at isolating the types of word relations that mediate this phenomenon (e.g. Fischler, 1977; McKoon & Ratcliff, 1979; Moss et al., 1995; Shelton & Martin, 1992). More specifically, a frequently asked question was whether words that are related only in one way, either semantically or episodically, would induce effective priming and how an interaction between these to types of relations would affect priming. Although a debate still exists, it is safe to say that both types of relations prime effectively and that their combined effect is additive.

A common theory of the organization principles of the semantic system and the mechanisms underlying semantic priming is the theory of spreading activation over semantic networks (Collins & Loftus, 1975). In a semantic network, a concept is represented as a node. Semantically related nodes are connected with unidirectional weighted links. When a concept is processed, the appropriate node is activated and the activation spreads along the connections of the network, gradually decreasing in strength. The decrease is proportional to the weights of the links in the path. In addition, the activation decays over time. Awareness of word occurs when its activation exceeds a threshold. According to this theory semantic priming occurs when activation from the prime spreads and partially pre-activates related targets so that a smaller amount of additional processing is required to bring its node activation above threshold.

An alternative and computationally more explicit modeling approach was recently proposed to explain semantic priming. Such models represent concepts in the semantic system by distributed (rather than local) representations. Concepts are not represented by single units, but rather by distinguishable patterns of activity over a large number of units (Hinton, 1990; Masson, 1995; Moss et al., 1994; Plaut, 1995). Each participating unit accounts for a specific semantic microfeature, and semantic similarity is thus expressed as overlap in activity patterns over the set of micro-features. In these models, recurrent dynamics is employed until the net settles to a stable state (an attractor). Semantic priming is explained by the fact that after settling to a prime, fewer modifications in the nodes' state are necessary for settling to a related target, making the latter settling process faster.

All the currently computational models of semantic priming have focused on processes based on existing associations. The process of acquiring new associations was abstracted in the training of the network. In the current study, we propose a computational model of forming new episodic associations between words on the basis of an already existing semantic network and show how this process is influenced by the organization of the semantic system.

Behavioral Experiments

A series of human performance experiments was conducted to supply constraints on the associating process. The following is a brief description of the relevant experiments and results (see Silberman & Bentin, submitted, for an elaborated report).

In one experiment, 10 randomly ordered Hebrew word pairs were repeated 20 times. In each trial, the two words were displayed one after the other with a Stimulus Onset Asynchronicity (SOA) of 700 ms. The subjects searched whether a letter presented 800 ms after the onset of the second word was included in the preceding word pair. Hence, proximity was achieved by having the subjects store the two words together in working memory for 800 ms. Following this "study session", the strength of the association between the words in each pair were unexpectedly tested using cued recall and a free association tests. In the cued recall test, the subjects were presented with the first words that occurred in half of the pairs, and asked to remember each word's associate. In the free association test, they were presented with the first words the other pairs, and asked to respond with their first free associate. We compared the strength of incidentally formed associations between semantically related (e.g. *milk-soup*) and semantically unrelated words (e.g. *paint-forest*). The results of this experiment, based on 64 subjects, are presented in Table 1.

Table 1: Percentage of cued recall and free association for pairs of semantically related and unrelated words.

Relatedness	Cued Recall	Free Associations
Related	38.8%	7.5%
Unrelated	19.4%	1.3%[1]

[1] Based on 16 subjects only.

As is evident in Table 1, semantic relatedness between words doubled the probability that an association would be incidentally formed between them. A between-subjects ANOVA of the cued recall performance showed that the difference between the two groups was statistically reliable [$F(1,62)=7.84$, $p<0.01$].

If semantic relationship facilitates the formation of associations by providing a higher initial linkage baseline or a smaller pool of candidates in the test phase, its effect should not interact with the number of episodic repetitions. Hence the difference between recall performance for related and unrelated word pairs should be the same, regardless the number of repetitions in the incidental learning phase (obviously, the absolute performance for both groups should positively correlate with the number of repetitions, up to a ceiling effect). To test this hypothesis, we manipulated the number of times each pair of the semantically related and unrelated pairs was repeated during the study phase.

Twenty-four Hebrew word pairs were selected for this experiment. The words in each pair were semantically related (belonged to the same semantic category) but not strongly associated (verified using free association questionnaires, in which we tested that none of the words was elicited by its pair among the first three free associates). Two study lists were prepared. Each consisted of 12 originally related pairs and 12 unrelated pairs formed by randomly pairing the other words. Pairs presented in the related condition in one list were used to form the pairs of the unrelated condition in the other list. Four groups of 24 subjects each were assigned to either 1 presentation (i.e., no repetition), 5, 10 or 20 presentations during the incidental study phase, in which subjects performed in the letter search task. The results of this experiment are presented in Figure 1.

An ANOVA showed that semantically related pairs were associated better than semantically unrelated pairs [$F(1,92)=204$, $p<0.0001$], and that the main effect of the number of repetitions was significant [$F(3,92)=25$ $p<0.0001$]. More revealing, however, was the significant interaction between the two factors [$F(3,92)=19$, $p<0.0001$], suggesting that each repetition contributed more to related than to unrelated pairs. These results suggest that semantic information reinforces the formation of associations (at least if these are formed incidentally). The semantic effect has a ceiling at which additional repetition contributes equally to forming both related and unrelated associations.

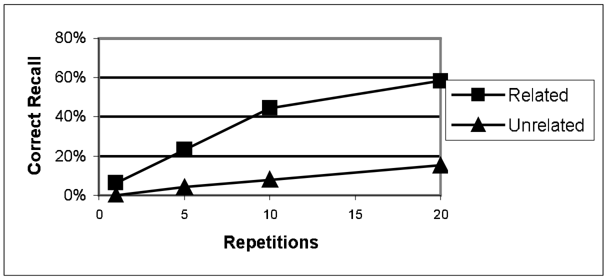

Figure 1: Percentages of correct recall for related and unrelated pairs in several learning repetition conditions. It is easier to form associations between related words.

These data demonstrated that semantic information is involved in forming episodic associations. For words that are semantically related, each learning repetition is more efficient. This facilitation is seen even if the subject's attention is not directed to the semantic level.

The aim of the present study was to develop a computational understanding of how associations are formed, including the influence of semantic factors on that process, as suggested by the above experiments.

Computational Model

Network Architecture

Our model is based on a Self-Organizing Semantic Map with lateral connections (Kohonen, 1995; Miikkulainen, 1992; Ritter & Kohonen, 1989). Semantic maps are 2-D networks that represent words by their nodes. The maps are formed by an unsupervised learning algorithm, such that words that are close in their meaning are represented by nearby nodes in the map. Hence, semantic relatedness is modeled by distance over the map. Semantic maps have been successfully used in various studies in which aspects of the semantic system were modeled, such as language acquisition (Li, 2000), semantic priming (Lowe, 1997), and semantic and episodic memory (Miikkulainen, 1992). Because self-organizing maps are based on biologically plausible Hebbian learning, and maps in general are common in many parts of the cortex (Knudsen, Lac & Esterly, 1987), self-organizing maps are most appealing as a biologically plausible analogue of classic semantic networks.

Based on a semantic map we added all-to-all unidirectional lateral connections to represent the potential associations between two words. The strength of each such connection is composed of semantic and episodic components:

$$Lat\left((i,j),(u,v)\right) = Sem\left((i,j),(u,v)\right) + Epis\left((i,j),(u,v)\right) \ , \quad (1)$$

where $Lat\left((i,j),(u,v)\right)$ is the connection weight from node (u,v) to node (i,j). The semantic component represents the distance on the map and is given by the equation:

$$Sem\ \left((i,j),(u,v)\right) = 1 \Big/ \left(1 + e^{\left|w(i,j)-w(u,v)\right|}\right) \ , \quad (2)$$

where $w(i,j)$ is the map's weights vector for neuron(i,j). Initially, the episodic part of all the lateral connections was set to zero. Hence, prior to any learning of associations, the lateral links only capture the topographic organization of the map, i.e. the semantic relatedness of words.

When a word is presented to the model, an activity bubble is generated surrounding the node that represents it. The activity wave then spreads according to synchronized recurrent dynamics. At each time step, the input to each neuron is the sum of the activities of all neurons in the previous time step, weighted by the lateral connections. Then, the neuron's activity is set according to a sigmoid function

$$S^t_{(i,j)} = \sigma\left(\sum_{(u,v)} Lat\left((i,j),(u,v)\right)\ S^{t-1}_{(u,v)}\right) \ , \quad (3)$$

where

$$\sigma(x) = 1 \Big/ \left(1 + e^{-x}\right) \ , \quad (4)$$

and $S^t_{(i,j)}$ is the activity of the neuron *(i, j)* at time *t*.

When two words are presented to the model (such as in the learning phase of Experiment 1 below) both activities spread independently over the map. The sum of the intersection of activation (the MIN of the two values) over all the map's neurons and over all time steps is calculated and added to the episodic component of the lateral connection between these two words. When the geometric distance between the two words is smaller (indicating stronger semantic relatedness), the resulting activity waves overlap more extensively, causing a greater amplification of the direct link between them. Thus, it is easier for the model to associate related words than unrelated words. Conceptually, this method is an abstraction of Hebbian learning of the associative links since the resulting connection strength depends on the intersection of both words' activation waves.

Input Representations

In order to organize the semantic map, we used numeric representations based on the lexical co-occurrence analysis in the Hyperspace Analogue to Language (HAL) model of Burgess and Lund (1997). These vectors have been shown to capture the semantics of words quite well (Burgess & Lund, 1997) and have been found successful in creating sensible Self-Organized Semantic Maps (Li, 2000). In the current simulations, HAL representations were based on the 3.8 million word CHILDES database, a corpus with particularly clearly defined word semantics.

The semantic map in our model consisted of 250 nouns organized on a 40 by 40 grid. We selected 48 nouns that formed 24 pairs of words, with the criterion that words in each pair belong to the same semantic category and thus are semantically related. The words were English translations of the 48 Hebrew words used in the behavioral experiment described above. In some cases, where a direct translation did not exist or the translation word did not appear in our set of HAL representations, a similar English word was selected. Another 202 nouns were selected randomly from the set of representations as "fillers" in the map, to create a richer semantic neighborhood in which the 48 words of interest could organize. See Figure 2 for the final semantic map that was used in the current simulations.

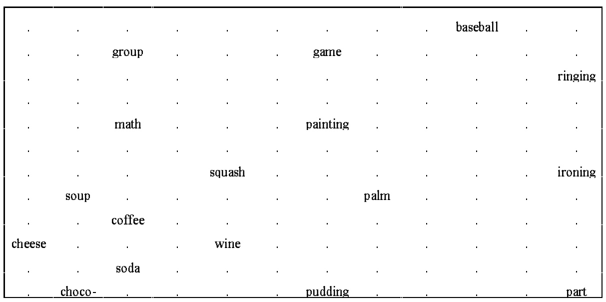

Figure 2: A section of the organized semantic map. Similar words are mapped to adjacent nodes.

Semantic Facilitation Simulation

The first simulation was aimed at replicating the empirical results of the second behavioral experiment that demonstrated semantic facilitation of associations' formation.

Experimental Setup

Out of the 24 semantically related pairs that were embedded in the semantic map, we selected 12 pairs that had numeric representations with a Euclidean distance shorter than the theoretical average (0.707) but larger then a threshold (0.5). Thus, these pairs were semantically related but not associated to each other prior to the experiment. In addition, we randomly re-matched the other 12 pairs such that 12 semantically unrelated pairs were formed as well.

Procedure

During the simulation of the learning phase of the experiment, in each trial the model was presented with two words with a certain time delay (i.e. SOA). Note that the absolute time scale of the network is arbitrary and can be adjusted to fit the data. Each of the 24 pairs (12 related and 12 unrelated) was presented once. Since during the learning phase, the episodic information does not affect the spreading activation process, the resulting association from multiple presentations was calculated simply by multiplying the result of a single presentation by the number of repetitions. The number of learning repetitions was varied from 1 to 30. During the simulation of the test phase, only one word was presented to the model. The resulting activation wave spread based on the same dynamics, except that in this phase, both the semantic and the episodic components of the lateral connections were taken into account. The activity continued to spread until the first neuron reached an activity threshold (0.98). The word represented by this node was then output as the result.

Results and Discussion

In Figure 3 the results that corresponds to the number of repetitions used in the behavioral experiment (1, 5, 10, 20) are shown. The percentages of correct recall demonstrated by the model for the related and unrelated pairs are shown for each repetition condition. As shown by the Figure, the model successfully replicates the results from the behavioral experiment. In the early stages, the learning rate of the related pairs is higher than the learning rate of the unrelated pairs. At about 10 repetitions, a ceiling effect reduces the learning rate of the related pairs, such that the advantage of these pairs over the unrelated pairs is abolished. In addition, the learning rate of the unrelated pairs is relatively constant. It is important to emphasize that non-linearity is introduced to the testing phase of the simulated experiment by the recurrent dynamics of the model. Hence, the linear way in which multiple repetitions were modeled does not dictate linearity in the output learning rate.

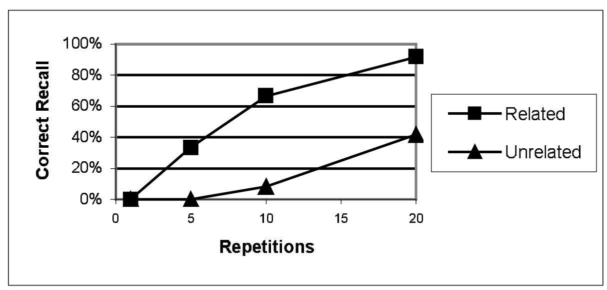

Figure 3: Percentages of correct recall demonstrated by the model, matching these obtained in the behavioral experiments (Figure 1).

Implicit Asymmetry Simulation

Associations between word pairs are directional. In free association questionnaires, for most pairs the subject would reply with word B after A with a different probability than the other way around (Koriat, 1981). In our model, this *explicit asymmetry* is achieved by the unidirectional lateral connections, which represent the association between two words in the map. However, our model demonstrates an additional asymmetry, which we call *implicit asymmetry*: it is sometimes easier to form an association between two words in one direction than in the opposite direction even before any episodic information is taken into account. The second simulation was aimed at quantifying this phenomenon.

Experimental Setup

First, we examined the density of the semantic neighborhoods of the words that were used in experiment 1. For each of the 48 words of interest we counted how many of the 250 total words in the model's semantic system were within a fixed 100-dimensional distance (0.4) according to their HAL representations. For each pair we then calculated the difference in the densities of the semantic neighborhoods of the two words and selected 3 related and 3 unrelated pairs with the greatest difference (in absolute values).

Procedure

We replicated the procedure of experiment 1 twice with the 6 pairs selected. First, the pairs were presented in the forward direction, from sparse to dense. Then, we repeated the entire procedure with the pairs presented in the opposite order (backward).

Results and Discussion

Figure 4 shows the percentages of correct recall as demonstrated by the model for pairs in the forward and backward direction in each repetition condition. Word pairs in the forward direction (sparser neighborhood → denser neighborhood) show advantage in correct recall as well as in learning speed throughout the first ten repetitions. The reason for this implicit asymmetry is the spreading of activation over a non-uniformly distributed high-dimensional space (elaborated in the General Discussion below). Although implicit asymmetry has not yet been observed experimentally, there is indirect evidence that suggests that such a process might indeed exist in the brain. Dagenbach, Horst and Carr (1990) found that it is easier to add a new word to semantic memory than to establish a link between two formerly unconnected words already in semantic memory. This result may apply to our prediction since we may assume that newly learned words were not yet well embedded into their semantic neighborhood and thus have a sparser semantic neighborhood than familiar words. In future work, we intend to test this prediction of the model with behavioral experiments.

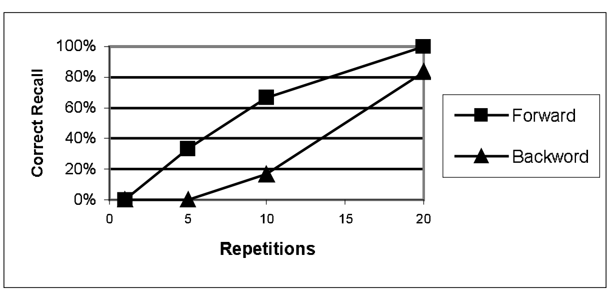

Figure 4: Percentages of correct recall demonstrated by the model for forward and backward pairs in several learning repetition conditions. The results show that learning is easier from sparse to dense neighborhoods.

General Discussion

As described in the "Behavioral Experiments" section, in a series of experiments we investigated semantic factors that affect the process of forming associations between words. The goal of this study was to present and evaluate a computational model that could account for these results and to produce further predictions regarding the process of associating words. Our model suggests that semantic relatedness between words, as well as episodic associations, could be implemented in a single structure using two distinct types of representations. On one hand, semantic relatedness is expressed as geometrical proximity in a high-dimensional (100D) feature space. On the other, episodic associations are represented by arbitrary "physical" connections between the units that represent the concepts. Both types of relations are implemented simultaneously in a semantic map with lateral connections artificial neural network.

As was demonstrated in human subjects, the model shows the facilitation semantic information has on learning new associations. This facilitation emerges in a natural, mechanistic manner, without involvement of top-down, intentional processes. It is achieved by implementing Hebbian link strengthening based on intersections of activation waves over a semantic map.

The asymmetric nature of relationships between words and more specifically of associations imposes difficulties for computational models that rely on geometric distances between high-dimensional numeric representations of words. Our model is also based on such high-dimensional vectors and the self-organization algorithm that establishes the semantic map is symmetric. Nonetheless, the model demonstrates two kinds of asymmetries. The first, explicit asymmetry is achieved by the unidirectional lateral connections that are implemented on top of the symmetric organization of the semantic map. These connections make it possible to have asymmetric associations between two words, based on the episodic experience of the two possible directions of the word pair. The second, implicit asymmetry, emerges from the non-uniform distribution of concepts in the high-dimensional space. This non-uniform distribution induces asymmetric distances in terms of spreading activation between two points in the semantic space that otherwise would have equal distance from one another in both directions.

Further empirical studies can be derived from this computational research. The model suggests that when an association is formed between two semantically distinct words, it can serve as a "pipeline" that enhances the spreading of activity from the semantic neighborhood of the first word to the semantic neighborhood of the second. Since this activity is, in turn, used to form other associations, we infer that the existence of an association between words from distinct semantic neighborhoods (e.g. different categories) would facilitate forming other associations between unrelated pairs that belong to those semantic neighborhoods.

Another possible implication of this model is in testing one of the theories concerning *Schizophrenic Thought Disorder* (hereinafter STD) as suggested by Spitzer (1997). According to Spitzer's theory, the activation over the semantic network of STD patients spreads faster and further than that of normal subjects. This unfocused activation can explain experimental results in STD patients that show stronger semantic priming and indirect semantic priming. It may also explain the clinical STD phenomena of loose, oblique and derailed associations. By manipulating our model we can computationally test this theory. It is possible to vary the parameters of the functions that govern the spreading activation (equations 1-4) to make it less focused. By examining the resulting changes in the model's behavior, we may be able to gain insight regarding the processes that lead to this pathology.

Conclusion

We set out to study the process of creating new associations between words in human memory during incidental learning. Empirical results suggest that semantic information enhances the process of forming episodic associations. A model based on spreading activation on a laterally connected self-organizing map matches these results and leads to further insights into why such associations tend to be asymmetric. In future work, we plan to test some of the model's predictions, including implicit associations and processes of abnormal behavior.

Acknowledgements

We are grateful to Ping Li and Curt Burgess for providing the HAL vectors. This research has been supported in part by the National Science Foundation grant IIS-981147 to Dr. R. Miikkulainen, and by a German-Israeli Science Foundation grant #567 to Dr. S. Bentin.

References

Burgess, C. & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 1-34.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.

Dagenbach, D., Horst, S., and Carr, T. H. (1990). Adding new information to semantic memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 581-591.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition*, 5, 335-339.

Hinton, G. E. (1990). Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, 46, 47-75.

Hume, D. (1738/1962). A Treatise of Human Nature. London: J.M. Dent & Sons, Ltd.

Knudsen, E. I., Lac, S., & Esterly, S. D. (1987). Computational maps in the brain. *Annual Review of Neuroscience*, 10, 41-65.

Kohonen, T. (1995). *Self-Organizing Maps*. Springer.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition*, 9, 587-598.

Li, P. (2000). The acquisition of tense-aspect morphology in a self-organizing feature map model. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp.304-309). Lawrence Erlbaum.

Lowe, W. (1997). Semantic representation and priming in a self-organizing lexicon. *Proceedings of the 4th Neural Computation and Psychology Workshop* (pp. 227-239). Springer-Verlag.

Masson, M. E. J. (1995). A Distributed Memory Model for Semantic Priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 21, 3-23.

McKoon, G. & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of verbal learning and verbal behavior*, 18, 463-480.

Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words. *Journal of Experimental Psychology*, 90, 227-234.

Miikkulainen, R. (1992). Trace feature map. *Biological Cybernetics*, 66, 273-282.

Moss, H. E., Hare, M. L., Day, P., & Tyler, L. K. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science*, 6, 413-427.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from Priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 863-883.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Lawrence Erlbaum.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37-42). Mahwah, NJ: Lawrence Erlbaum.

Ritter, H. & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics*, 61, 241-254.

Shelton, J. R., & Martin, R. C. (1992). How semantic is semantic priming? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 18, 1191-1210.

Silberman, Y. & Bentin, S. (submitted). *Semantic boost of forming episodic associations.*

Spitzer, M. (1997). A cognitive neuroscience view of schizophrenic thought disorder. *Schizophrenia Bulletin*, 23, 29-50.