Semantic Boost on Episodic Associations: An Empirically-based Computational Model

Yaron Silberman
Department of Psychology and Interdisciplinary Center for Neural Computation,
The Hebrew University of Jerusalem, Giv'at Ram, Jerusalem 91904 Israel

Shlomo Bentin
Department of Psychology and Interdisciplinary Center for Neural Computation,
The Hebrew University of Jerusalem, Giv'at Ram, Jerusalem 91904 Israel

Risto Miikkulainen
Department of Computer Science, The University of Texas at Austin
Austin, TX 78712-1188 USA

Address for editorial correspondence:

Risto Miikkulainen
Department of Computer Sciences
The University of Texas at Austin
1 University Station C0500
Austin, TX 78712
E:Mail: risto@cs.utexas.edu

Key Words: associations, episodic memory, semantic memory, neural networks, semantic maps

Abstract

Humans associate words remarkably easily and efficiently. Frequently the associations are formed incidentally and such associations provide a window into the cognitive structure of language networks and memory. How associations between words are formed, however, is not well understood. In a series of incidental associative learning experiments human participants associated semantically related but unassociated words easier than unrelated words. Furthermore, this advantage increased linearly as a function of the number of repeated co-occurrences of the word-pairs, up to a ceiling. In order to understand the mechanism of the semantic influence on word association processes a computational model, SEMANT, was developed. In this model, episodic associations are implemented by lateral connections that are formed between nodes in an already existing self-organized semantic map. These lateral connections are strengthened by each concomitant activation of two nodes in direct proportion with the amount of overlapping activity. Simulating associative learning with SEMANT replicated the empirical findings and led to two testable predictions. First, it should be easier to associate a word with few semantic neighbors to a word with many semantic neighbors than vice versa. Second, diffuse semantic activation, such as that characteristic to schizophrenic thought disorders, should reduce the advantage of associating semantically related pairs over unrelated pairs. The SEMANT model, therefore, suggests how the interaction of episodic and semantic memory can be understood in terms of specific neural computations.

1. Introduction

Word associations are basic building blocks of human cognition in general and human learning in particular. Understanding how such associations are formed can give us a fundamental insight into the cognitive structures underlying learning and memory.

A viable association between words is demonstrated phenomenologically when the presentation of one word brings the second word to the perceiver's awareness. By definition, associations are based primarily on episodic learning. That is, words are associated if they co-occur in time and space, and the strength of an association is determined by the cumulative effect of such episodes. An important aspect of this definition is that associative learning is not necessarily intentional. Indeed, associations are often established incidentally, that is, without intention and without allocating attention to the learning process.

Connections between words can also be based on a semantic relationship. For example, words can share common semantic features (e.g. when they belong to the same semantic category such as goat - lion), can maintain a part-whole relationship (e.g. wheel - car), a functional relationship (e.g. broom - floor) and more (Cruse, 1986). Note that if words co-occur frequently they may become associated whether they are semantically related or not, but semantically related words are not always associated (Prior & Geffet, 2003; Neely, 1991). Indeed, episodic associations and semantic relationships are probably based on different properties. Yet, many semantically related word pairs are also episodically associated, which might suggest that the two types of relations, indeed, interact.

This paper focuses on understanding this interaction in detail. Performance experiments with human participants were first performed to characterize the interaction between the categorical neighborhood organization of the semantic system and the episodic dynamics of associative learning. In these experiments the participants' ability to form associations between semantically related and unrelated words was compared. Semantic relations were found to facilitate forming new associations in a specific manner: They provide only a negligible initial advantage, but instead facilitate the process by which the episodic associations are strengthened over multiple repetitions, up to a ceiling.

These performance patterns were then used as a starting point for developing a computational theory of word associations, expressed in the SEMANT computational model. In SEMANT, semantic similarity is based on distributed representations (cf. HAL) and episodic associations are based on spreading activation, both represented together in a laterally connected semantic map. The model was validated in simulated associative learning experiments, and then used to draw predictions for future human experiments on asymmetry of associations and potential causes of mediated associations in schizophrenia. In this manner, SEMANT expresses a computational theory of semantic and episodic effects in word associations, and serves to drive future research.

2. Background

There is considerable evidence that semantic and episodic memory systems interact. Most of such studies (e.g. McKoon & Ratcliff, 1986; Neely & Durgunoglu, 1985; Shelton & Martin, 1992) were based on the semantic priming paradigm and focused on examining existing structures. In particular, several studies showed considerably larger semantic priming effects on performance if the prime and the

target are also episodically associated (Chiarello et al. 1990; McRae & Boisvert, 1998); to some extent, these results generalize into newly formed associations as well (Dagenbach, Horst, & Car, 1990; Schrijnemakers & Raaijmakers, 1997).

A converse influence of semantic relationship on associative learning has also been documented. For example, early studies reported more accurate cued recall for strongly related than for weakly related word pairs using the paired associate paradigm (e.g., McGuire, 1961; Underwood & Schultz, 1960; Wicklund, Palermo, & Jenkins, 1964). Supporting these earlier findings, Epstein, Phillips & Johnson (1975) suggested that semantically related words were more likely to be associated during an incidental learning procedure than unrelated pairs, even if the orientation task did not involve semantic processing. More recently, Guttentag (1995) reported similar findings in children. Moreover, using sentential context rather than relatedness between words, Prior & Bentin (2003) found that associations between unrelated words are formed more easily when they are embedded in a sentence than when they are presented as isolated pairs.

Although such studies suggest that semantic relationships between words affect how associations are formed between them, the data needs additional corroboration, primarily because the strength of pre-experimental associative links between semantically related words has not been sufficiently controlled. Hence, the reported advantage for associating semantically related over unrelated words could have been induced by associative connections that existed before the experiment rather than by an interaction between the two types of connections during associative learning. Therefore, human experiments investigating the semantic influence on the

formation of new associations between words were performed as a starting point in the current study, as will be described in the next section.

Given that empirical data exists, computational modeling is an ideal tool to make sense of it and to formulate specific theories about the process. The modeling can be done at several different levels of abstraction, depending on the level of explanation desired. At the highest level, the semantic and episodic interactions can be formalized in Bower's one-element model (Bower, 1961). For example, the semantic relationship advantage could be modeled so that in each trial there's a fixed chance that an association is learned between semantically related word pairs, and a smaller fixed chance that it is learned between semantically unrelated pairs. Such a model accounts for the interaction, but it does not explain how these learning patterns emerge from memory structures, and it is therefore limited in its predictive power.

At a more detailed level, a popular approach to explaining the organization of semantic memory is representing concepts by distinguishable patterns of activity over a large number of nodes (Farah & McClelland, 1991; Hinton, 1990; Hinton & Shallice, 1991; Masson, 1995; Moss et al., 1994; Plaut, 1995; Plaut & Shallice, 1993). Each node participating in a representation accounts for a specific semantic micro-feature, and semantic similarity is expressed as an overlap in activity patterns over the set of micro-features. Activity propagates through recurrent connections until the network settles to a stable state (an attractor), representing the semantically related items. Such a model matches much of the same data but allows more detailed explanations and predictions to be drawn. For example, Plaut's (1995) model suggests a distinction between the way by which semantic and episodic relations are formed: Whereas semantic relations are encoded in the micro-features, episodic relations

between word pairs are based on co-occurrence information throughout the learning history. On the other hand, the attractor models are rather difficult to interpret, and it is difficult to separate the different factors contributing to their behavior. In particular, Plaut's (1995) model does not specify what mechanisms underlie the process of forming episodic associations and how the strength of existing semantic connections among other factors affects this process.

The goal of computational modeling, however, is not just to model the data, but to be able to draw predictions that can be verified in future empirical experiments. The SEMANT model is a principled model specifically designed to suggest and test such a mechanism. SEMANT is not an attractor network, but is instead based on three well-known principles that make its processing and learning transparent: map-like memory organization, spreading activation, and Hebbian adaptation. Its core process is Spreading-Activation over Semantic Networks, which is a well-established theory of semantic memory organization (Collins & Loftus, 1975; Cottrell & Small, 1983; Quillian, 1966; Waltz & Pollack, 1985). According to this theory, each concept is represented by a node in a semantic network, and semantically related nodes are connected with weighted links. When a concept is processed, the appropriate node is activated and the activation spreads along the network connections in a manner proportional to the connection weights, gradually decreasing in strength over time. In order to bring a word to awareness, its activation must exceed a threshold. Thus, the spreading activation model provides a simple, transparent process that can be readily matched with experimental data.

Before describing the details of SEMANT and its predictions, in the following sections the empirical data on which it is based will be reviewed**.**

3. Human Performance Experiments

Two experiments were performed on human participants, in order to characterize the interactions between episodic and semantic memory systems in some detail. The first experiment showed that it is easier to form new episodic associations between semantically related words compared with unrelated words. During an incidental study phase, 10 randomly ordered Hebrew word pairs were repeated 20 times. In each trial, two words were displayed sequentially for 350 ms each and a Stimulus Onset Asynchrony (SOA) of 700 ms. After an additional SOA of 800 ms (relative to the onset of the second word), a single letter was presented and the participant was requested to determine (pressing an alternative button) whether this letter was or was not included in either of the preceding words (Kutas & Hillyard, 1988; Smith, Bentin & Spalek, 2001). Hence, episodic proximity was established by having the participants store the two words together in working memory for 800 milliseconds. The shallow processing orientation task was selected in order to avoid task-biased semantic activity at study and depth of processing effects for single words (cf. Craik and colleagues, e.g., Craik & Lockheart, 1972; Craik & Tulving, 1975).

Following the study session in which the participants performed with an accuracy of over 90%, the strength of the association between the words in each pair was unexpectedly tested using a cued recall test and a free association test. In the cued recall test, the first words of half of the studied pairs were presented as cues, and the participants were requested to respond with the cue's associate. In the free association test, the participants were presented with the first words of the other half of the pairs, and the participants were asked to respond with a first free-associate. Correct cued recall performance and responding with the episodically paired word in the free

association tests were considered evidence that an association between the two words was incidentally formed during study. The effect of semantic relationship was tested between subjects. One group of 32 participants was tested with semantically related but not associated pairs (e.g. **MILK-SOUP**)[1] and another equally sized group was tested with semantically unrelated words (e.g. **PAINT-FOREST**). The performance of the 64 subjects is 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 participants.

Both cued recall and free association revealed stronger associations between semantically related pairs than for unrelated pairs [$t(62)=2.8$, $p<0.01$ and $t(46)=2.36$, $p<0.025$ for cued recall and free associations, respectively). These results support earlier studies such as that of Epstein et al. (1975) reporting that semantically related words are easier to associate than unrelated words.

There are at least two possible explanations for this effect. One is that the probability of encountering two related words in real life episodes is higher than encountering two unrelated words. Hence, although not evident while testing word associations explicitly (in our pilot survey), weak episodic associations may have existed between semantically related words, and those could have facilitated associative learning. Another possible explanation is that activation spreads more efficiently among semantically related than unrelated words in the semantic system.

[1] Importantly, the absence of associations between the words in semantically related as well as unrelated words was tested in a pilot survey requesting 50 participants from the same pool to provide the first 3 associates to each of the words used. In no case a pair included words that were mentioned by a participant among these first associates.

This effect makes it easier to form associative connections during each new episode of concomitant activation. Although the two accounts are not mutually exclusive, they should affect cued recall performance in distinguishable ways. Pre-existent (weak) associations should enhance associative strength at the very beginning of the episodic associative learning process. Following this initial boost, however, pre-existent association should not influence the learning curve. In contrast, facilitated associative learning should be evident (at least up to a ceiling) at each recurrent learning episode. Consequently, the effect of efficient spread of activation should manifest as a steeper learning function for related than unrelated pairs.

In order to test the relative contribution of each of these two factors, in a subsequent experiment the semantic relationship between the words in each pair (within-subject factor) and the number of repetitions during the incidental learning phase (between-subjects factor) were manipulated. Twenty-four pairs of exemplars representing 24 semantic categories were selected. As in the previous experiment, the words in each pair did not elicit one another in free association questionnaire in which participants were requested to provide the first three associates in response to a cue word. From the list of 24 related pairs, two study lists were prepared. Each list comprised of 12 related pairs form the list above, and 12 unrelated pairs. The unrelated pairs were scrambled pairings of the other 12 pairs. The related pairs in list A were scrambled in list B and the unrelated pairs in list A were reorganized to form the related set in list B. Four different groups of 24 participants were assigned to one presentation (i.e., no repetition), 5, 10 or 20 presentations during an incidental learning phase, during which they performed the letter search task. Following study, the participants were unexpectedly presented with a cued recall test in which the first

word of all 24 pairs were presented as cues and the participants were requested to respond with the paired associated word from the study phase.

All participants performed the letter search task at study with accuracy above 85%, suggesting that they paid attention to the words. Cued recall was better for semantically related than unrelated words for all groups, and the semantic advantage increased with increased repetition (Figure 1). Mixed-model ANOVA showed that both main effects were significant [$F(1,92)=204$, $p<0.0001$, and $F(3,92)=25$ $p<0.0001$, for semantic relationship and number of repetition factors, respectively]. More revealing, however, was the significant interaction between the two factors [$F(3,92)=19$, $p<0.0001$], which showed that each repetition contributed more to related than to unrelated pairs. These results suggest that semantic information does not only provide an initial advantage, but enhances the process of forming episodic associations (at least if they are formed incidentally). The effect of the interaction, however, has a ceiling, above which additional repetition contributes equally to forming both related and unrelated associations. Any further learning is based on co-occurrence only.

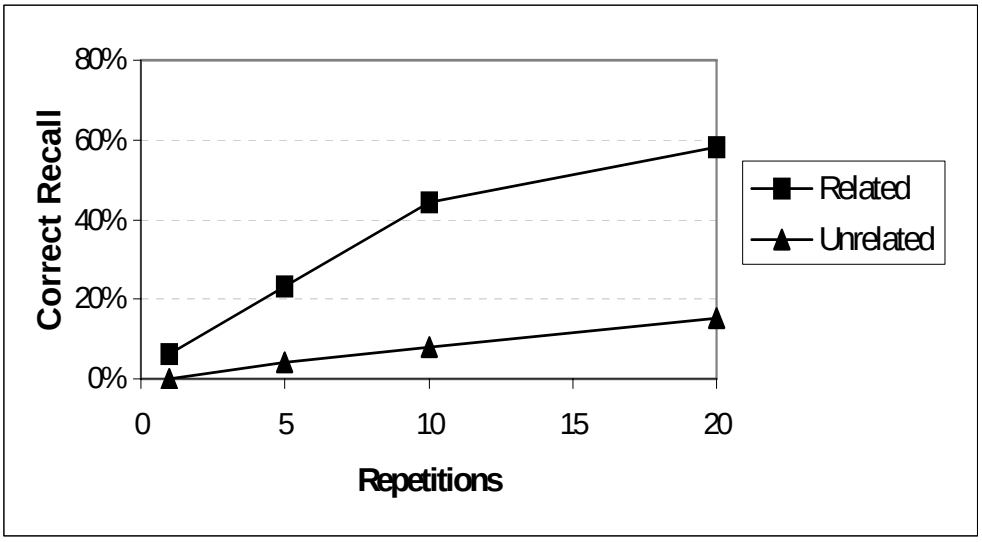

Figure 1: Average percentages of correct recall for the related and unrelated word pairs over increasing number of learning repetitions. Semantic information enhances the process of forming associations, up to a ceiling effect.

The second experiment, therefore, extended the results of the first experiment, distinguishing between semantic and episodic contributions to the associative process. Together, the results describe in detail how semantic relatedness facilitates forming incidental episodic associations between words. Next, a computational model for this process is presented, and predictions for further experiments derived.

4. Computational Model - SEMANT

The computational simulations were based on SEMANT (Semantic and Episodic Memory of Associations using Neural neTworks), a biologically-motivated and psychologically real neural network model of the semantic system. Following an overview of SEMANT, the semantic representations, the network architecture, and the model dynamics will be described in this section.

4.1 Overview

SEMANT is a two-dimensional self-organizing map network where semantically related concepts are represented by nodes that are closer to each other than semantically unrelated nodes. Episodic associations are represented on this map as direct connections among the nodes. Activation among nodes spreads along both semantic and associative connections. Inspired by the notion of Hebbian learning, the strength of an association between two conjointly activated nodes is enhanced when the "wave" of activation spreading from one node overlaps with the activation spreading from another node. Since spreading activation decays with distance from the source, the closer the two concepts are on the semantic map, the greater the overlap between their activations. Hence, when two semantically related words are conjointly activated, the associative connection strength between them increases by a large amount. On the other hand, because semantically unrelated words are further

apart on the semantic map, their concomitant activations overlap less and consequently their associative connection strength increases by a smaller amount.

4.2 Input Representations

Word semantics was based on lexical co-occurrence analysis according to the Hyperspace Analogue to Language (HAL) model of Burgess and Lund (1997). In this model, high-dimensional numeric representations for words are formed based on the context in which they occur in large text corpora. A moving window of several words is placed around each word in the text. A co-occurrence matrix is calculated according to how often each word occurs in the different positions in the window. The rows and columns of this matrix are high-dimensional vector representations for each word. To make them computationally tractable, the 100 dimensions with the highest variance are used for the final representations. As has been shown in a number of studies, the resulting HAL vectors capture the semantic of words fairly well: Semantically related words have similar, closer, representations (Burgess & Lund, 1997; Li, 1999; 2000; Li, Burgess & Lund, 2000; Lund, Burgess & Atchley, 1995; Lund, Burgess & Audet, 1996).

In the present study, HAL representations were 100 dimensions long and formed based on the 3.8-million-word CHILDES database (Li, 1999; MacWhinney, 2000). Forty-eight nouns were selected from this database to form 24 pairs of words. The words in each pair were exemplars from the same semantic category. The words were the English translation of the 48 Hebrew words used in the second of the human experiments described above. In some cases, if a direct translation did not exist or the translated word did not appear in the set of HAL representations, a similar English word was chosen. Another 202 nouns were selected randomly from the same set of

representations in order to create a richer semantic neighborhood for the 48 words of interest. The words are listed in the appendix.

4.3 Self-Organizing Maps

The SEMANT architecture is based on a Self-Organizing Semantic Map with lateral connections (Miikkulainen, 1992; Miikkulainen, Bednar, Choe, and Sirosh, 2005; Ritter & Kohonen, 1989). The Self- Organizing Feature Map (SOM) is an Artificial Neural Network (ANN) that classifies high-dimensional data and represents their topological structure in a two-dimensional map (Kohonen, 1990). The algorithm has two different versions of implementation. One version is biologically motivated and, therefore, includes computations that are biologically plausible. The other version is oriented towards practical applications, and the equations that govern it are abstractions of the computations in the biological version. SEMANT is based on a hybrid implementation that retains the computational efficiency of the abstract model but includes psychologically verified mechanisms when they are crucial for the performance of the model.

The SOM architecture is a feed-forward, two-layered neural network. The input is presented on the first layer, while the second layer, consisting of a 2-D array of nodes, self-organizes to represent the topology of the input space. In SEMANT, the nodes of the output layer are interconnected with all-to-all lateral connections. Inputs that are significantly different in their semantic features are mapped onto distinct locations in the output space; similarly, inputs with many overlapping semantic features are mapped onto nearby locations. The organization of the map, i.e. the assignment of weight vectors to the nodes in the second layer, is formed in an unsupervised, iterative process, driven by the statistics of the input's features.

The map's response to the input vector is determined by calculating the scalar product of the weight vectors of each node and the input vector:

$$s_{ij} = \sum_k w_{ij,k} x_k \quad , \tag{1}$$

where s_{ij} is the response of node *(i,j)* in a two-dimensional map, *x* is the input vector and $w_{ij,k}$ is the weight between component *k* of the input vector on node *(i,j)*. Each iterative adaptation step consists of two stages: First, the node in the output layer that responds most strongly to the current input vector is found. Second, the weight vectors of the most responsive node and neighboring nodes are modified to become more similar to the input vector. The neighborhood in which weight vectors will be modified is defined around the most responsive node, with a radius that decreases from nearly the size of the entire map down to zero as the self-organization progresses. The weight vectors of nodes in the neighborhood are changed to become closer to the input vector according to

$$w_{ij}(t+1) = w_{ij}(t) + \varepsilon(t)\left[x(t) - w_{ij}(t)\right] \quad , \tag{2}$$

where $wij(t)$ is the weight vector of node (i,j) at time t, $x(t)$ is the input vector at time t, and $\varepsilon(t)$ is the adaptation rate, which is usually decreased to zero with t. In this process, the weight vectors of the map nodes become approximations for the input vectors and the weight vectors of neighboring nodes become similar, which results in an ordered map of the input space.

4.4 Semantic Maps

The SOM algorithm is commonly used to visualize high-dimensional inputs with inherent metric properties (Kohonen 1990; 2000). Such inputs are typical to low-

level perception. However, when dealing with high-level processing such as semantics in language and memory, the inputs are often discrete and not necessarily metric. There are two major ways to represent semantics (Ritter and Kohonen, 1989). According to the first, the representation is a *feature vector*, where bits represent a set of fixed properties (big/small, has 2 legs/4 legs, etc.). According to the other, the context of the word, e.g. the random representation of the word and its predecessor and successor words, is used as the input. In both cases, *Semantic Maps* are created, that is, meaningful maps of word semantics that express grammatical and semantic relationships between words. Because the latter methodology can be automated, it can be applied to large corpora such as the entire text of the Grimm fairy-tales (Honkela, Pulkki and Kohonen, 1995). Therefore it was also adopted for SEMANT, using the HAL approach for creating the representations.

Semantic maps have been successfully used in various investigations of the semantic system addressing issues such as language acquisition, semantic priming, semantic and episodic memory, and were used to document representation and retrieval (Kohonen et al., 2000; Li, Farkas & MacWhinney, 2004; Lowe, 1997; Mayberry, 2004; Miikkulainen, 1993; Scholtes, 1991). Because self-organizing maps are based on 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 (Spitzer, 1997).

4.5 SEMANT architecture

The core of the SEMANT model is a semantic map consisting of 250 nouns organized in a 40 by 40 grid. The semantic map is extended to include all-to-all unidirectional lateral connections. These connections represent potential associations

between two words. The strength of each connection is composed of semantic and episodic components:

$$Lat_{ij,uv} = Sem_{ij,uv} + Epi_{ij,uv} \quad , \tag{3}$$

where $Lat_{ij,uv}$ is the connection weight from node (i,j) to node (u,v). The semantic component represents the distance on the map, given by

$$Sem_{ij,uv} = \frac{AMP}{\left(1 + e^{\left(\left|w_{ij} - w_{uv}\right|^2 - SFT\right)SLP}\right)} \quad , \tag{4}$$

where w_{ij} is the map's weight vector for node (i,j) and *AMP*, *SLP*, and *SFT* are free parameters. [2]The equation describes a reverse sigmoid with *AMP* defining its height, *SLP* its slope and *SFT* its offset. Therefore, the strength of each lateral connection is inversely proportional to the 100-dimensional distance between the nodes' weight vectors. Initially, the episodic part of all lateral connections is zero. Hence, prior to any learning of new associations, the lateral connections capture only the topographic organization of the map, that is, the words' semantic relations.

The semantic map was organized in two phases. In the first phase, aimed at producing a gross organization, the neighborhood size was decreased from 40 to 3 over 10,000 iterative adaptation step. In the second organization phase, aimed at fine tuning the map's representations, the neighborhood size was decreased from 3 to 0 over additional 20,000 iterations. The adaptation rate (ε) was 0.5 throughout the organization process but decreased linearly as a function of the Euclidian distance square between the updated node and the most responsive node at each iteration.

[2] There are a total of 19 free parameters in this model.

4.6 SEMANT Dynamics

When a word is presented to SEMANT, an activity "bubble" develops surrounding the node that represents the word in the network. The generated activity wave spreads from the activated node according to synchronized recurrent dynamics. At each time step, the input to each node is the sum of the activities of all nodes in the previous time step, weighted by the lateral connections. The node's activity is limited between the values 0 and 1 according to a sigmoid function

$$s_{ij}(t) = \sigma\left(\sum_{uv} Lat_{ij,uv} s_{uv}(t-1)\right), \quad (5)$$

where

$$\sigma(x) = \frac{1}{(1+e^{-x})}, \quad (6)$$

and $S_{ij}(t)$ is the activity of node (i,j) at time t.

When two words are presented to SEMANT during the same learning trial (τ), the episodic weights adapt to encode a lateral connection (i.e., an association) between them. Both activities spread independently over the map and the intersection of these activations is summed over all the map's nodes and over all time steps. This sum is added to the episodic component of the lateral connection between these words (only in the direction that corresponds to the order of presentation) according to

$$Epi_{i_{PRM}j_{PRM},i_{TGT}j_{TGT}}(\tau+1) = Epi_{i_{PRM}j_{PRM},i_{TGT}j_{TGT}}(\tau) + \sum_{mnt} \min\left(s_{mn}^{PRM}(t), s_{mn}^{TGT}(t)\right), \quad (7)$$

where $s_{mn}^{PRM}(t)$ is the activation of node (m,n) at time t resulting from the presentation of the prime word, which is represented in node (i_{PRM},j_{PRM}), and $s_{mn}^{PRM}(t)$ is the activation of node (m,n) at time t resulting from the presentation of the target word, which is represented in node (i_{TGT},j_{TGT}).

When the distance between the words is small (indicating that the words are strongly semantically related), the resulting activity waves overlap significantly and the connection between them is considerably amplified at each co-occurrence. Thus, it is easier for SEMANT to associate related words than unrelated words. Conceptually, this method is an abstraction of Hebbian learning of episodic links, since the resulting connection strength depends on the intersection of the activation waves of both words.

4.7 Computational Participants

In the simulation experiments, 12 different computational participants were generated by self-organizing the semantic map each time from different random initial starting points. The same sequence of input words was used for each participant (see appendix for detailed simulation parameters).

The resulting maps learned to represent the semantic similarities in the data, but differed in the details of how they were organized. For example, in the map of Figure 2, food-related concepts are clustered in the bottom and body parts on top. The food and body part clusters were prominent on other maps as well, but were shaped differently and located in different regions of the map. In this sense, the 12 maps can be seen as different individuals with roughly equivalent semantic systems. Simulated psychological

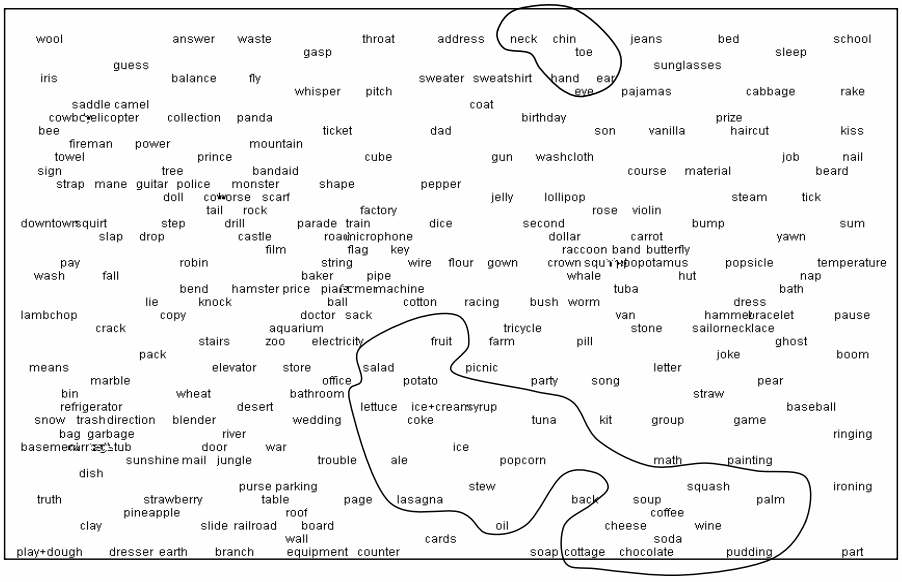

Figure 2: A self-organized semantic map. Semantically related words, such as those denoting body parts and foods, are mapped to adjacent nodes. Twelve different maps were organized from different random initial values, and were used to simulate different participants.

experiments were conducted in this group of computational participants in order to explain the performance observed in human studies.

5. Validating SEMANT as a Model of Human Learning

The first simulation was aimed at verifying the psychological validity of SEMANT, by replicating the human performance in the second experiment described above. Again, this experiment demonstrated that it is easier to form associations between semantically related than unrelated words, and that this advantage increases with repetitive presentation.

5.1 Simulation 1

5.1.1 Experimental Setup

Out of the 24 semantically related pairs on the semantic map 12 pairs were chosen with a Euclidean distance between HAL representations shorter than average (0.88) but larger then a pre-determined threshold (0.35). This selection insured that the words in each pair were semantically related, but not strongly enough to be always recalled regardless of any episodic associations. In addition, the other 12 pairs were randomly re-matched to form 12 pairs of semantically unrelated words. The simulation procedure was replicated 12 times using a different map each time and statistical analysis was performed on the data.

5.1.2 Procedure

In each trial during the simulated study phase, SEMANT was given two words one after the other at an appropriate SOA (i.e., the first word was given at time step 0 and the second word only at time step 3). The absolute time scale of the network is arbitrary and was adjusted to fit the data. Each of the 24 pairs (12 related and 12 unrelated) was presented once. Because episodic factors do not

affect how activation spreads in the modeled semantic network during the learning phase, the episodic associative strength resulting from multiple presentations was calculated by multiplying the result of a single presentation by the number of repetitions, which varied from 1 to 30.

In each trial during the test phase, the first word in each of the studied pairs was presented again to SEMANT. The resulting activation wave spread based on the same dynamics as during the study phase, except that both semantic and episodic components of the lateral connections now affected propagation. The activity continued to spread until one of the nodes representing a word reached a pre-determined activity threshold (0.98). This word was then emitted as output. If no word-node reached the threshold within eight time steps (at which time the activation wave had usually decayed to a negligible level), the word with the highest activation was emitted as an answer. The latter scenario simulated the situation in which the participant could not recall any word and would answer with the word that first came to mind.

5.1.3 Results and Discussion

As evident in Figure 3, SEMANT successfully replicated the pattern of results found in human subjects. At early stages, the semantically related pairs were associated faster than unrelated pairs. After about 10 repetitions, a ceiling effect reduced this pace, such that their advantage over the unrelated pairs stops growing. In contrast, the pace of learning unrelated pairs was relatively constant throughout the repetitions.

It is important to emphasize that due to the sigmoid transfer function (Equations 5 and 6) as well as the recurrent dynamics, SEMANT's response changes nonlinearly with increasing repetitions. Moreover, the performance in the simulated

cued-recall test depends nonlinearly on the lateral connection strengths. Hence, the nearly linear learning curves in SEMANT (until the ceiling effect) cannot stem merely

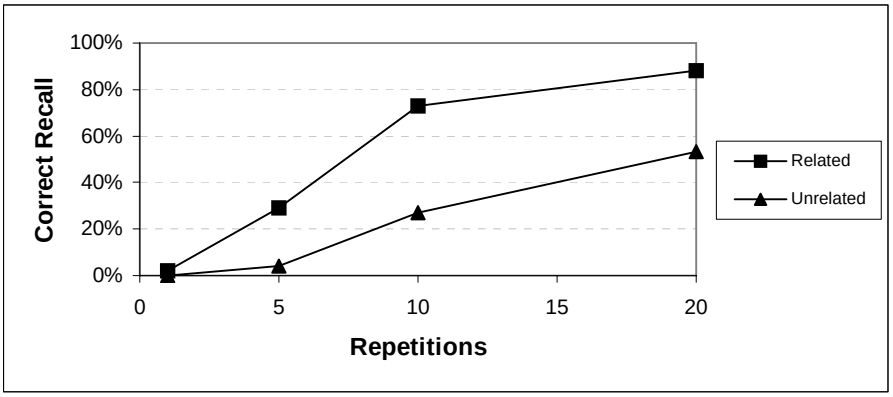

Figure 3: Average percentages of correct recall in the model over increasing repetitions. The performance of the model matches that of the behavioral experiments (Figure 1), suggesting that word recall could be due to computations in a laterally connected semantic map.

from the linear way in which multiple repetitions were modeled (Section 4.1.2). Instead, they accurately represent an equal contribution of each learning trial to the correct cued-recall performance.

Statistical analysis of the data revealed that both main effects, i.e. the effect of semantic relatedness and the effect of number of repetitions, as well as the interaction between them were reliable [$F(1,11)=104.4$ $P<0.001$, $F(3,33)=352.4$ $P<0.001$ and $F(3,33)=35.2$ $P<0.001$, respectively]. This result demonstrated that SEMANT is a psychologically valid model; it further suggests that the observed human associative learning and cued recall performance could arise from computations involving a laterally connected semantic map. In two subsequent simulations, SEMANT was used to derive predictions that characterize these processes in more detail.

6. Implicit Asymmetry in Forming Associations

Unlike semantic relationships, associations between words are directional. In free association questionnaires, for most pairs the participants would reply with word B after A with a different probability than the other way around (Koriat, 1981). In SEMANT, *explicit asymmetry* is achieved by the unidirectional lateral connections, which represent the association between two words in the map. However, SEMANT demonstrates an additional asymmetry that is *implicit*: 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. Simulation 2 was aimed at quantifying this phenomenon.

6.1 Simulation 2

6.1.1 Experimental Setup

The density of the semantic neighborhoods of the words used in Simulation 1 was determined by counting the number of words (out of the 250 total words in SEMANT's lexicon) that were within a fixed 100-dimensional distance (0.4) in their HAL representations. Based on this count 3 related and 3 unrelated pairs were selected in which the difference in the semantic densities of the two words in a pair was the greatest. These 6 pairs were used in this simulation. As before, the simulation procedure was replicated 12 times using a different map each time and statistical analysis was performed on the data.

6.1.2 Procedure

As in Simulation 1, the six pairs were presented one word at a time. First, the pairs were presented in the forward direction, from the word with the sparse semantic neighborhood to the word with the dense semantic neighborhood in each pair. Then,

the network was reset to its original state and the entire procedure was repeated with the pairs presented in the opposite order (from dense to sparse). The "cued-recall" performance of the two directions was compared statistically.

6.1.3 Results and Discussion

As shown in Figure 4, when the pairs were presented in the forward direction (sparse neighborhood → dense neighborhood) associations were learned faster than when the order of presentation was reversed [$F(1,11)=24.8$ $p<0.001$]. However, a significant interaction between direction of presentation and semantic relationship effects [$F(3,33)=3.9$ $p<0.05$] revealed that the order of presentation influenced only the semantically related words. Post hoc analysis of the order effect showed that whereas the related word pairs were learned faster in the forward than backward direction [$F(1,11)=46.6$ $p<0.001$], order had no effect on unrelated word pairs [$F(1,11) < 1.00$].

The reason for this implicit asymmetry is that activation spreads over a non-uniformly distributed high-dimensional space (as will be elaborated in the General Discussion). Consequently, each word results in a unique pattern of activation over the map over time. In particular, spreading the activation from a sparse neighborhood to a dense one takes longer than the other way around. Therefore, for semantically related pairs, the intersection of the two activation waves is larger and persists longer when the first word is in a sparse neighborhood than when it is in a dense neighborhood. For unrelated word pairs, the difference is inconsequential because the activation waves do not overlap significantly.

Although implicit asymmetry has not yet been observed in human performance, indirect evidence suggests that it might be psychologically real. For example, 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 existing in semantic memory. This result is consistent with the asymmetrical directionality prediction of SEMANT if newly learned words are assumed to be poorly embedded into their semantic neighborhood and, therefore, to have a sparser semantic neighborhood than familiar words do.

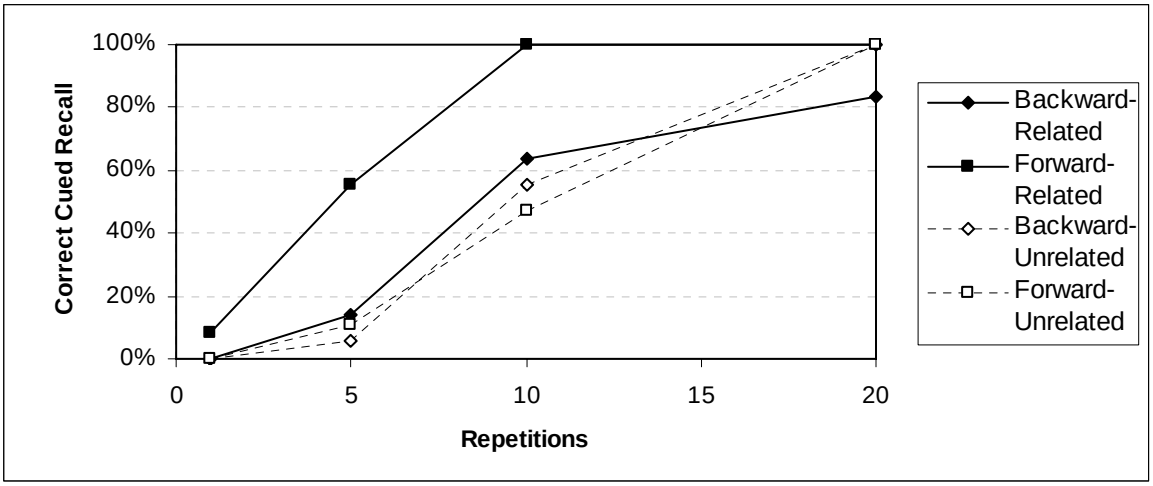

Figure 4: Average percentages of correct recall demonstrated by the model for related and unrelated word pairs in forward (sparse to dense) and backward (dense to sparse) directions. The model predicts that learning is easier from sparse to dense neighborhoods.

7. Modeling Associative Learning in Schizophrenia

Based on Collins and Loftus's (1975) theory of spreading activation, Spitzer (1997) suggested a model of schizophrenic thought disorder (STD) where the activation wave over the semantic network of STD patients spreads faster and farther away from origin than that of normal participants. Such atypical activation may explain experimental results in STD patients, who show stronger direct priming (e.g. between the words COW and MILK) as well as more indirect, mediated semantic priming (e.g. between the words BULL and MILK, mediated by the word COW) than normal subjects. It may also explain the clinical phenomenon of loose, oblique and derailed associations. There are at least two possible accounts for the aberrant spreading of activation in the STD semantic system: (1) the total amount of activation

may be higher in STD patients than in typical population, and (2) the activation may be shallower in STD patients but more widespread. In Simulations 3A and 3B these two possible accounts are tested in SEMANT. By comparing the performance in each of the simulations with those of human participants it is possible to obtain insight into what might be causing semantic disorders in STD patients.

7.1 Simulation 3A

7.1.1 Experimental Setup and Procedure

This simulation tested the effect of increasing the overall amount of activation on associative learning. The experimental setup and procedures of Simulation 1 were replicated, except that the amplitude of the spreading activation wave was relatively elevated. As described above (Equation 4), the distance between the word representations (nodes) determines how strongly they are linked according to a reverse sigmoid function (Equation 4). This sigmoid was elevated by multiplying the *SFT* parameter of Equation 4 by a factor of two. As a result, the area underneath the curve increased but the shape of the sigmoid remained similar to that used in Simulation 1 (Figure 5). The psychological interpretation of this change is that the stronger and unfocused activation wave results from a higher level of excitability in the system. As in Simulation 1, the simulation procedure was replicated 12 times using a different map each time and statistical analysis was performed on the data.

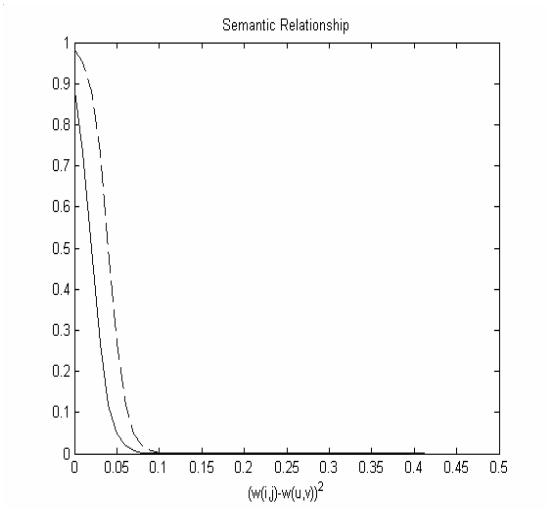

Figure 5: Semantic relatedness as a function of the distance between word representations in normal participants (continuous line) and in STD patients (dashed line) modeled by an elevated curve. Such a curve corresponds to a higher level of excitability in the system.

7.1.2 Results and Discussion

As shown in Figure 6, running the simulation, SEMANT predicted that the associative recall performance of STD patients in both related and unrelated conditions would improve, at least up to about 10 repetitions [$F(1,11)=58.9$ $P<0.001$]. Since both cued-recall and priming depend on memory organization, this finding is consistent with the increased semantic priming reported in schizophrenic patients. Computationally, the improvement in performance was caused by the elevated sigmoid, which generated a greater activity overall. Interestingly, unlike modeling typical performance, when STD constraints are imposed on the simulation the associations between unrelated words were formed as fast as between related words. This pattern was verified by a Relatedness × Number of repetitions × Model (typical, STD) ANOVA showing a significant three-way interaction $F(3,33)=13.4$ $P<0.001$. This result corresponds to the finding that STD patients show more mediated

associations, in the sense that semantically unrelated words are easier for STD patients to associate than for matched control participants.

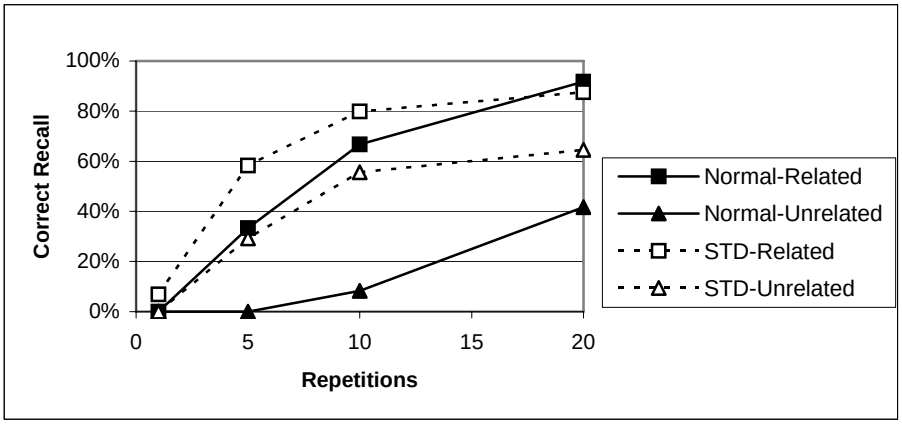

Figure 6: Averaged percentages of correct recall demonstrated by SEMANT for normal participants and for STD patients with elevated activity. Performance for STD patients is elevated for both related and unrelated pairs and the difference between the two conditions is diminished, suggesting that mediated associations in STD may result from enhanced semantic connections.

Nevertheless, the superior associative recall performance of STD patients predicted by simulation 3A for both related and unrelated pairs is intriguing. Notwithstanding the stronger semantic priming effect, STD patients do not usually succeed in memory tests better than normal participants, especially with incidentally learned associations. Note that the better cued-recall for STD patients resulting from Simulation 3A is a result of higher activation levels in general. To address this problem and explore the second account for the unusual priming effects observed in STD patients, in Simulation 3B a shallower but broader sigmoid was used. Such a sigmoid generates unfocused activation without increasing its overall amount. The goal was to see whether more loose associations would result without a significant increase in overall recall performance.

7.2 Simulation 3B

7.2.1 Experimental Setup and Procedure

The setup and procedures of this simulation were similar to that of Simulation 3A except that the sigmoid was normalized such that the sum of the semantic components of the lateral connections was identical to that used in Simulation 1 (Figure 7). Thus, the total strength of the lateral connections in the simulated STD was the same as in the normal case; keeping the overall activation at the same level in both groups. However, the connection profile was much wider, representing an unfocused association system. As in all our previous simulations, the simulation procedure was replicated on 12 different maps and analyzed statistically.

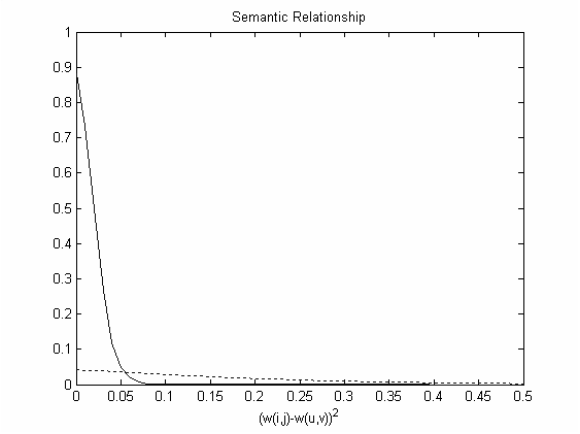

Figure 7: Semantic relatedness as a function of the distance between word representations in normal participants (solid line) and in STD patients (dashed line), modeled by a shallower and wider curve. Such a curve corresponds to higher unfocused associations.

7.2.2 Results and Discussion

Unfocused activation was implemented in SEMANT by making the sum of the semantic components of the lateral connections the same in STD and in the typical case. As can be seen in Figure 8, this simulation resulted in lower percentage of cued recall in related pairs while the performance of unrelated pairs was higher than in the

normal case. Statistical analysis (ANOVA) did not show a statistically reliable difference between normal and STD conditions [$F(1,11)=0.086$ $P>0.75$], but resulted in a significant relatedness effect [$F(1,11)=193.893$ $P<0.001$] and a significant interaction between the relatedness and STD conditions [$F(1,11)=26.620$ $P<0.001$]. This interaction indicates that the semantic relatedness had a larger effect in the normal than in the STD case. Moreover, as in Simulation 3A, the qualitative difference in the rate of learning related versus unrelated words was decreased in STD relative to normal [significant three-way interaction $F(3,33)=15.6$ $P<0.001$].

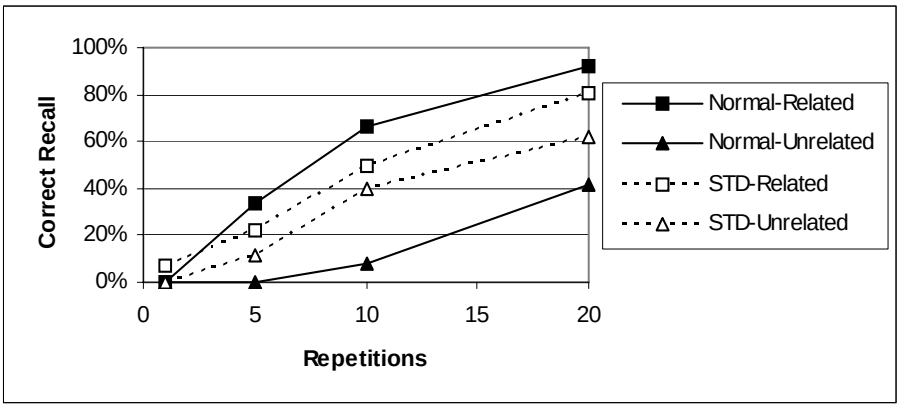

Figure 8: Average percentages of correct recall demonstrated by SEMANT for normal participants and STD patients with widespread activity. Performance for STD patients is degraded for related pairs, elevated for unrelated pairs, and the difference between the two conditions is diminished, suggesting that mediated associations may result from unfocused associations.

The conclusion from both simulations 3A and 3B is that the difference between associating related and unrelated words is diminished in STD relative to normal conditions. However, the pace of learning new associations and the final plateau was determined by the type of model: Assuming higher overall semantic activation in STD (simulated by higher amplitude sigmoid, Figure 5) resulted in more efficient learning of both related and unrelated pairs, even higher than in the control

condition. In contrast, assuming equal excitability in both groups, but spread more diffusely in the STD participants (simulated by a sigmoid with lower amplitude but further reaching influence, Figure 7) resulted in better associative learning of unrelated pairs but reduced associative learning of related pairs. This result can be useful in uncovering the source of thought disorders in schizophrenic patients. Indeed, preliminary data from an ongoing study in human STD patients and normal individuals suggests that the second model is better supported by real data (Bentin et al., in progress).

8. General Discussion

The goal of the present study was to determine how semantic and episodic factors interact in learning new incidental associations between words. Empirical studies with human participants showed that semantically related words are associated more easily than unrelated words, and that this difference is primarily due to higher associative strength added by each episode of co-occurrence. A computational model, SEMANT, was developed to account for this phenomenon and to suggest a more general mechanism for word association. An initial simulation demonstrated that the model is psychologically valid: The computational process of forming new associations between related and unrelated words accurately replicated human performance. Subsequent simulations generated predictions about associative learning between different types of words and modeled different accounts for the peculiar associative learning in schizophrenic patients with thought disorders. These model-derived predictions (as well as others) can be tested empirically in future experiments with human subjects. The model has also broader implications for the study of memory, as will be discussed in this section.

SEMANT suggests that both semantic relationship and episodic associations can be implemented in a single network structure using two types of representations. Semantic relationship is expressed as proximity in a high-dimensional features' space spanned by the numerical representations of the concepts. Episodic associations are represented by arbitrary "physical" connections between the nodes that represent the words. Both types of relations are implemented simultaneously in a self-organizing semantic map with lateral connections. Based on such structure, SEMANT showed that semantic relationship facilitates learning new associations. This facilitation emerges in a natural, mechanistic manner, without involving top-down intentional processes. It is achieved through a process similar to Hebbian learning, where connections are strengthened based on intersections of spreading activation waves over a semantic map.

In SEMANT, the episodic connections do not affect the organization of the semantic map. This modeling approach reflects the assumption that different rules govern the organization of the semantic and episodic memories: The semantic map is organized according to outside (possibly sensory) information and is not affected by the episodic connections. Indeed preliminary experiments show that newly formed episodic associations do not affect the semantic memory, i.e. they do not bring entire neighborhoods closer together (Silberman et al. 2005). Such effects may be possible at very long time scales, which are currently outside the scope of the model.

The basic principles underlying SEMANT (map-like memory organization, spreading activation, and Hebbian adaptation) are well established in the literature and have been used to model several other phenomena. The fact that these principles can also account for word associations suggests that they may, indeed, represent general mechanisms of cognition. The predictive power in the model was demonstrated in a

number of cases. One prediction was that it is easier to form associations for sparse to dense semantic neighborhoods than the other way around; another was that excessive mediated associations in STD could be due to elevated or diffuse spreading activation process, and these mechanisms can be distinguished in the data. A third prediction is that a strong association between semantically unrelated words (e.g. BEER - DAISY) facilitates forming new associations between other words from the same semantic neighborhoods (e.g. WINE - TULIP). This is because each of the new (to be associated) words activates its own semantic neighborhood, including the previously associated words. The pre-existent episodic association mediates the spread of activation between the new conjointly presented exemplars, thus enhancing the overlap of their activations This prediction has already been validated empirically in human (Silberman, et al., 2005).

The asymmetric nature of word associations is difficult to explain with computational models that rely on distances between high-dimensional numeric representations. One atypical example is Plaut's (1995) model, which can, in principle, capture asymmetric associations between word pairs based on the relative frequency of the two directions of presentation (although such behavior has not yet been demonstrated in that model). Similarly, SEMANT is based on perfectly symmetric foundations such as high-dimensional vectors and the self- organization algorithm that establishes the semantic map. Nonetheless, SEMANT demonstrates two kinds of asymmetries (as was discussed in section 6). The first asymmetry is explicit; it is achieved by 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 different episodic experiences in the two directions. The second asymmetry is implicit emerging from the non-uniform distribution of concepts in the high-dimensional

space. This distribution makes spreading activation asymmetric between two points that would otherwise be equidistant in the semantic space. By doing so, SEMANT suggests a mechanism based on which asymmetric behavioral phenomena can emerge from seemingly symmetric building blocks.

SEMANT demonstrates implicit asymmetry because the weight vectors in the semantic map are distributed non-homogeneously in the 100-dimensional space. However, factors other than semantics may influence this distribution and consequently influence the implicit asymmetry as well. As described in section 4, the density of the weight vectors in the semantic map is determined during the self-organization process and reflects how the numerical HAL representations of the 250 words are distributed. Lund et al. (1995, 1996) argued that HAL representations reflect primarily the semantic features of the words. Therefore the semantic map in SEMANT is likely to reflect primarily the semantic relationships between words prior to the episodic learning. However, other factors such as word frequency and word familiarity may be confounded in HAL representations as well, and therefore may affect the semantic density and implicit asymmetry in SEMANT. Additional studies are necessary to disentangle the putative contribution of such factors.

SEMANT also suggests that the nodes unassigned with words in the semantic map are important, since they serve as a mediating substrate for spreading activation. Any system organized in an unsupervised fashion to accommodate an unknown number of items (such as the human semantic system) should be expected to have a much larger capacity than is usually needed. In semantic maps, the number of nodes in the map is much larger than the number of represented words (e.g. in SEMANT, 250 nodes are embedded in 402 nodes, resulting in 6.4 nodes per word on average). However, due to the self-organization algorithm, these unassigned nodes are not

distributed uniformly in the high-dimensional space, but rather represent the densest areas of the input space. Consequently, they help magnify the effects that are due to the statistical properties of the input. An example of such an effect is the asymmetry in the learning efficiency between two directions of word pairs, as demonstrated in Simulation 2.

While the similarity between human performance and the computer simulations suggest that SEMANT is a valid psychological model, it is still a high-level abstraction of the underlying biological mechanisms. For example, recall that during the simulation of the learning phase, two words are presented to SEMANT with a certain SOA and the two activation waves spread independently. The lack of interaction between the two spreading waves is a computational abstraction that could be elaborated in a more biologically plausible model. More specifically, in order for two (or more) waves to spread without interaction within the same system, the waves can be labeled, or marked, to carry the information to identify their source throughout their propagation. Such a mechanism is central in the marker-passing extensions to spreading activation (Charniak, 1983; Hendler, 1988). One biologically plausible marker passing mechanism is the synchronized spiking model (Choe & Miikkulainen, 1998; Miikkulainen et al., 2005; Shastri and Ajjanagadde, 1993; von der Malsburg, 1986). If two activations spike at different phases, the information regarding their source is preserved. On the other hand, if the enhancement of the association strength between the two words is done in real-time during the propagation of the waves, and if the sensory input that created the activation is accessible throughout the propagation, a marking scheme may not be needed: The nodes where the waves originated can be identified because they respond maximally to the sensory input. However, it may not be realistic to assume that two (or more) activation waves may

propagate through the same system (semantic memory) without interaction. Hence, a more plausible assumption would be that waves that originate from different sources interact in some manner and that the propagation of one affects the propagation of the others. Changing SEMANT to support unlabeled interacting activation and still replicating human behavior is an interesting future research line.

9. Conclusion

Based on detailed studies with human participants, a computational model of forming incidental associations between words was developed. The model suggests that this process could be based on spreading activation on a laterally connected self-organizing map that combines episodic and semantic factors. It also demonstrates how such associations could be asymmetric and how the associative process can be impaired in STD patients. These results constitute a promising first step towards a computational theory of associative thinking in human.

Acknowledgements

We are grateful to Ping Li and Curt Burgess for providing the HAL vectors. This research was supported in part by the National Science Foundation grant IIS-981147 and National Institute of Mental Health Human Brain Program grant R01 to Risto Miikkulainen, and by a German-Israeli Science Foundation grant I-771-250.4/2002 to Shlomo Bentin.

References

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, 26, 255-280

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

Charniak, E. (1983). Passing markers: {a} theory of contextual influence in language comprehension. *Cognitive Science*, 7, 171-190.

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't… sometimes, some places. *Brain and Language*, 38, 75-104.

Choe, Y. & Miikkulainen, R. (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing*, 21, 139-157.

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

Cottrell, G. W. & Small, S. L. (1983). A connectionist scheme for modeling word sense disambiguation. Cognition and Brain Theory, 6, 89-120.

Craik, F.I.M., & Lockheart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.

Dagenbach, D., Horst, S., and Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 581-591.

Epstein, M. L., Phillips, W. D., & Johnson, S. J. (1975). Recall of related and unrelated word pairs as a function of processing level. *Journal of Experimental Psychology: Human Learning and Memory*, 104(2), 149-152.

Farah, M., & McClelland, J. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357

Guttentag, R. (1995). Children's associative learning: automatic and deliberate encoding of meaningful associations. *American Journal of Psychology*, 108(1), 99-114.

Hendler, J. A. (1988). *Integrating Marker Passing and Problem Solving: {A} Spreading Activation Approach to Improved Choice in Planning*. Hillsdale, NJ: Erlbaum.

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

Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74–95.

Honkela, T., Pulkki, V., & Kohonen, T. (1995). Contextual relation of words in Grimm tales, analyzed by self-organizing map. *Proceedings of international conference on artificial neural networks* (pp. 3-7).

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

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE* (pp. 1464-1478).

Kohonen, T., Kaski, S., Lagus, K., Salojrvi, J., Paatero, V., & Saarela A. (2000). Self-organization of a massive document collection. *IEEE Transactions on Neural Networks*, 11, 574-585.

Kohonen, T., Kaski, S., Lagus, K., Salojrvi, J., Paatero, V., & Saarela A. (2000). Self-organization of a massive document collection. *IEEE Transactions on Neural Networks*, 11, 574-585.

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

Kutas M and Hillyard SA (1988): An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, 1: 38-49.

Li, P. (1999). Generalization, representation, and recovery in a self-organizing feature-map model of language acquisition. *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp.308-313). Mahwah, NJ: Lawrence Erlbaum.

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). Mahwah, NJ: Lawrence Erlbaum.

Li, P., Burgess, C., & Lund, K. (2000). The acquisition of word meaning through global lexical co-occurrences. *Proceedings of the 30th Child Language Research Forum* (pp. 167-178). Stanford, CA: Center for the Study of Language and Information.

Li, P., Farkas, I., & MacWhinney, B. (2004) Early lexical development in a self-organizing neural network. *Neural Networks* 17, 1345-1362.

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.

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 660-665). Mahwah, NJ: Lawrence Erlbaum.

Lund, K., Burgess, C., & Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 603-608). Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Volume 1: Transcription format and programs. Volume 2: The Database.* Mahwah, NJ: Lawrence Erlbaum Associates.

Masson, M. E. J. (1995). A distributed memory model for semantic priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 21, 3-23.

Mayberry, M. R. (2004). Incremental Nonmonotonic Parsing Through Semantic Self-Organization, Doctoral Dissertation, Department of Computer Sciences, The University of Texas at Austin. Technical Report AI-TR-04-310.

McGuire, W. J. (1961). A multiprocess model for paired-as-associate learning. *Journal of Experimental Psychology*, 62, 335-347.

McKoon, G., & Ratcliff, R. (1986). Automatic activation of episodic information in a semantic memory task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 108-115.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* , 24(3), 558-572.

Miikkulainen, R. (1992). Trace feature map: a model of episodic associative memory. *Biological Cybernetics*, 66, 273-282.

Miikkulainen, R. (1993). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon and Memory*. Cambridge, MA: MIT press.

Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. New York: Springer.

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.

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.

Neely, J.H. & Durgunoglu, A. (1985). Dissociative episodic and semantic priming effects in episodic recognition and lexical decision tasks. *Journal of Memory and Language*, 24, 466-489.

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.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377–500.

Prior, A., & Bentin, S. (2003). Incidental formation of episodic associations: the importance of senential context. *Memory & Cognition*, 31(2), 306-316.

Prior, A., & Geffet, M. (2003). Mutual information and semantic similarity as predictors of word association strength: Modulation by association type and semantic relation. *Proceedings of the European Cognitive Science Conference 2003* Mahwah, NJ: Lawrence Erlbaum Associates.

Quillian, M. R. (1966). Semantic Theory. Doctoral Dissertation, Carnegie Institute of Technology.

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

Scholtes, J. C. (1991). Unsupervised learning and the information retrieval problem. *Proceedings of the IEEE International Joint Conference on Neural Networks* (pp. 95-100). Piscataway, NJ: IEEE Service center.

Schrijnemakers, J.M.C. & Raaijmakers, J.G.W. (1997). Adding new word associations to semantic memory: Evidence for two interactive learning components. *Acta Psychologica*, 96, 103-132.

Silberman, Y., Miikkulainen, R., & Bentin S. (2005). Associating unseen events: Semantically mediated formation of episodic associations. *Psychological Science*, 16, 161-166.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review,* 96, 523-568

Shastri, L. & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: {a} connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16, 417-494.

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

Smith, M.C., Bentin, S. & Spalek, T.M. (2001). Is spreading of semantic activation truly automatic? The influence of task and response delay. *Journal of Experimental Psychology: Learning Memory and Cognition*, 27, 1289-1298.

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

Underwood, B. J., & Schultz, R. W. (1960). Response dominance and rate of learning paired associates. *Journal of General Psychology*, 62, 153-158.

von der Malsburg, C. (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54, 29-40

Waltz, D. L. & Pollack, J. B. (1985). Massively parallel parsing: a strongly interactive model of natural language interpretation. Cognitive Science, 9, 51-74.

Wicklund, D. A., Palermo, D. S., & Jenkins, J. J. (1964). The effects of associative strength and response hierarchy on paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 3, 413-420.

Appendix

Entire lexicon:

1. address
2. ale
3. answer
4. aquarium
5. back
6. bag
7. baker
8. balance
9. ball
10. band
11. band-aid
12. baseball
13. basement
14. bath
15. bathroom
16. bathtub
17. beard
18. bed
19. bee
20. bend
21. bin
22. birthday
23. blender
24. board
25. boom
26. bracelet
27. branch
28. bump
29. bush
30. butterfly
31. cabbage
32. camel
33. cards
34. carriage
35. carrot
36. castle
37. cheese
38. chin
39. chocolate
40. clay
41. coat
42. coffee
43. coke
44. collection
45. copy
46. cottage
47. cotton
48. counter
49. course
50. cow
51. cowboy
52. crack
53. crown
54. cube
55. dad
56. desert
57. dice
58. direction
59. dish
60. doctor
61. doll
62. dollar
63. door
64. downtown
65. dress
66. dresser
67. drill
68. drop
69. ear
70. earth
71. electricity
72. elevator
73. equipment
74. eye
75. factory
76. fall
77. farm
78. farmer
79. film
80. fireman
81. flag
82. flour
83. fly
84. fruit
85. game
86. garbage
87. gasp

88. ghost
89. gown
90. group
91. guess
92. guitar
93. gun
94. haircut
95. hammer
96. hamster
97. hand
98. helicopter
99. hippopotamus
100. horse
101. hut
102. ice
103. ice-cream
104. iris
105. ironing
106. jeans
107. jelly
108. job
109. joke
110. jungle
111. key
112. kiss
113. kit
114. knock
115. lamb-chop
116. lasagna
117. letter
118. lettuce
119. lie
120. lollipop
121. machine
122. mail
123. mane
124. marble
125. material
126. math
127. means
128. microphone
129. monster
130. mountain
131. nail
132. nap
133. neck
134. necklace
135. office
136. oil
137. pack
138. page
139. painting
140. pajamas
141. palm
142. panda
143. parade
144. parking
145. part
146. party
147. pause
148. pay
149. pear
150. pepper
151. piano
152. picnic
153. pill
154. pineapple
155. pipe
156. pitch
157. play-dough
158. police
159. popcorn
160. popsicle
161. potato
162. power
163. price
164. prince
165. prize
166. pudding
167. purse
168. raccoon
169. racing
170. railroad
171. rake
172. refrigerator
173. ringing
174. river
175. road
176. robin
177. rock
178. roof
179. rose
180. sack
181. saddle
182. sailor
183. salad

184. scarf
185. school
186. second
187. shape
188. sign
189. slap
190. sleep
191. slide
192. snow
193. soap
194. soda
195. son
196. song
197. soup
198. squash
199. squirrel
200. squirt
201. stairs
202. steam
203. step
204. stew
205. stone
206. store
207. strap
208. straw
209. strawberry
210. string
211. sum
212. sunglasses
213. sunshine
214. sweater
215. sweatshirt
216. syrup
217. table
218. tail
219. temperature
220. throat
221. tick
222. ticket
223. toe
224. towel
225. train
226. trash
227. tree
228. tricycle
229. trouble
230. truth
231. tuba
232. tuna
233. van
234. vanilla
235. violin
236. wall
237. war
238. wash
239. washcloth
240. waste
241. wedding
242. whale
243. wheat
244. whisper
245. wine
246. wire
247. wool
248. worm
249. yawn
250. zoo

Word pairs used in the simulations:

Semantically unrelated:

1. necklace - vanilla
2. painting - train
3. van - store
4. guitar - cabbage
5. pear - electricity
6. pepper - song
7. coat - violin
8. office - dress
9. camel - roof
10. oil - cow
11. wall - strawberry
12. carrot - bracelet

Semantically related:

1. bag - purse
2. eye - chin
3. butterfly - bee
4. gun - stone
5. lettuce - chocolate
6. snow - sunshine
7. doctor - baker
8. dice - cards
9. coffee - soda
10. bandage - pill
11. ball - doll
12. desert - river