Neuromodulation of Word Meaning Selection

Gabriele Scheler*
ICS
1947 Center Street
Berkeley, Ca. 94704
e-mail: scheler@icsi.berkeley.edu

Abstract
Processes of word meaning generation, word association and understanding are known to be impaired in schizophrenia and related diseases. Word meaning selection requires the involvement of prefrontal cortex and processes of working memory and selective attention. Under the dopaminergic hypothesis of schizophrenia, the normal neuromodulatory activation of prefrontal cortex for the performance of working memory-related tasks is disturbed. We present a model of selective attention and its modulation by dopamine and show how abnormal levels of dopamine availability may lead to some of the observed impairments in word meaning selection, namely (a) failure to construct contextually appropriate meanings and (b) intrusions of phonological and episodic associative links within semantic processing.

Introduction
Studies on brain imaging have revealed that the task of word generation, such as in a constrained association task ("name a typical verb for a noun"), reliably produces activity in two distinct brain regions: in parietal cortex, the putative storage area for long-term semantic knowledge, and in areas 44 and 45 in prefrontal cortex (Broca's area) as sites for attentional selection (Thompson-Schill et al. 1997).

Attentional selection is a process that is known to underlie neuromodulation, i.e. activation of primarily dopaminergic and noradrenergic receptors on neurons in prefrontal cortex. There are a number of behavioral tests of selective attention and distractability designed both for animals and humans and a specific enhancement of reaction times and accuracy under conditions of raised dopamine levels or raised D1-receptor activity could be shown.

The processes of active maintenance of representations in working memory and selection of behaviorally relevant information have also been investigated using recordings of neuronal activity of behaving monkeys. During a delayed alternation task, which requires to store the location of an object on a screen during a brief (1–2 s) delay period, a number of neurons show raised activity (i.e. an increased firing rate). This has been interpreted as showing that these neurons provide a representation of the inaccessible stimulus, which requires some internal active maintenance as for instance a resonant state in an attractor network (Zipser et al. 1993).

Recently, a further experiment has looked at the selectivity of representations in prefrontal cortex (Rainer, Asaad, & Miller 1998). Monkeys performed a delayed-matching-to-sample task with an array of three objects. After a cue object (the target) had been shown, monkeys had to remember the location of the target object for a delay of 1.5 s. Approximately 50% of the recorded neurons showed a different activity depending on the location of the target, only about 5% showed activity that varied with the location of non-target objects. This effect was present about 140 ms after stimulus onset, and a parallel effect concerning the target object rather than the location, which was unchanged during trail blocks showed persistent activity preferably for target objects.

Similar attentional effects in IT as an area of visual processing (Chelazzi et al. 1993) take somewhat longer, 175 ms, and this may suggest that selective representation in prefrontal cortex provides an attentional bias for processing in other parts of the cortex (Miller 1999).

Looking at language processing in the brain, we may note that processes of word meaning selection require the short-term memory storage of several contextual items as well as a selection of relevant information from long-term storage of semantic meanings or features. Selection involves the resolution of competition between retrieved representations.

Neuroimaging studies of verbal working memory (Smith & Jonides 1999) indicate an anatomical separation of maintenance and executive processes such as selection and inhibition.

To provide a model of word meaning selection we therefore need to address three issues:

- a model of selective attention in prefrontal cortex
- a model of long-term storage of semantic information

---

*On leave from Department of Computer Science, Technical University of Munich, D-80290 München
- a model of retrieval and integration of long-term information with short-term storage

In the following we will provide a model of semantic memory that is based on the notion of semantic features integrated in a self-organizing map-like structure, and that is biologically plausible but not tightly constrained by experimental data of the kind that are available for animal models. We will then propose a model of selective attention and its enhancement and degradation by levels of neuromodulators. Finally we will show how the impairment of processes of selection in prefrontal cortex lead to characteristic disturbances in word meaning processes, such as verb generation, association and semantic priming which can be observed under changed levels of neuromodulators in humans. The problem of retrieval and integration of long-term and short-term information will not be addressed here - we will assume a simple mirroring or pointer-like address system that links these systems together (Collette et al. 1999), (Fuster 1998), (Chafee & Goldman-Rakic 1998).

**Semantic Feature Maps**

We start from the assumption that semantic meanings are created out of a set of atomic (unanalyzable) features and that these features are represented by individual neurons or small sets of them. As a matter of fact there are reasons to assume that a distributed representation of atomic features ("microfeatures") by a set of neurons has computational advantages and may be the preferred method of implementation in the brain. In our model, however, we will only deal with the level of features and therefore there is no great difference between the two alternatives.

Feature-representing neurons are linked with other feature neurons by synaptic weights of differing strength (cf. Fig. 1). This means that features are usually activated in clusters, or "structures" with stronger and weaker activated features. Activation within a network means that certain feature neurons receive synaptic input from perceptual (phonological) layers and in turn send action potentials to their projection neurons. Within an interconnected network this leads to reverberations and a general increase of firing rate for the activated neurons, which is proportional to their connectivity pattern (cf. Fig. 2).

If we assume that a certain structure of this kind is represented in prefrontal cortex, i.e. is present in verbal working memory rather than in long-term storage, there is a further process that contributes to activation. In addition to dedicated feature-representing neurons there are a large number of unassigned neurons that maintain weak links to most other neurons and that become "recruited" by strongly activated neurons. There is a competition for the activation of these neurons in that they integrate synaptic input from different sources but will fire in synchrony only with synchronized strong input. This means that a pool of neurons that is tightly

![Figure 1: An example of a semantic map with features and links of differing strength. The structure of the maps can be estimated from psychological data. Adapted from a real example in (Spitzer 1997).](image)

![Figure 2: Spiking pattern in a network of integrate-and-fire neurons. Synaptic strengths are taken from Fig. 1. An input is presented to features C and G as an example of a phonological word stimulus. Firing rate of features corresponds to their strength of activation.](image)
or less semi-active and inactive neurons, but a categorical structure seemed to be easier to understand and we could have implemented a graded structure of more neurons. These neurons are more weakly active (have lower firing rates) than the representational neurons, we call these neurons "semi-active". Alternatively, there may be a cascade-like activation of neuronal groups which receive some input and which have stronger connections to the active neurons than the background neurons. There will be a number of neurons receiving strong input, which represent a certain item that is active in working memory, and which are connected in the manner of a reverberating attractor (Amit, Brunel, & Tsodyks 1994). There is a pool of inhibitory neurons which will be activated by the current attractor (representation) and suppress activity of the other neuronal groups that receive input activation. By choosing a certain connectivity structure of inhibitory and excitatory groups, a cascade-like activation of neuronal groups can be achieved (Fig. 3, cf. appendix B for a specification of the architecture).

Finally, we assume that there is a set of neurons, which receive some input and which have stronger connections to the active neurons than the background neurons. These neurons are more weakly active (have lower firing rates) than the representational neurons, and they do not follow an oscillatory pattern. We call these neurons "semi-active". In a sense, it is clear that we could have implemented a graded structure of more or less semi-active and inactive neurons, but a categorical structure seemed to be easier to understand and to analyze. Fig. 4 shows a raster-plot of neuronal activity in such a working memory model.

Representations in working memory are not stable but subject to attentional switches. Items may be rehearsed and accessible for several seconds but normally there are constant switches of content, both voluntary and involuntary. We propose to enhance the basic oscillatory model by a certain amount of background activity, which may spontaneously cluster to small semi-active groups of neurons and which may become entrained to the major oscillations by increased synaptic input, in particular coincident synaptic input. Selective attention may be localized in a different network and can increase or inhibit certain neuronal activations in primary working memory. Selective attention can thus stabilize (focus) representations but also induce switching of representations.

Figure 3: Architecture of oscillatory working memory. Representational neurons form strongly interconnected clusters and activated inhibitory clusters with a k-1 inhibitory pattern. There is a pool of semi-active neurons. Excitatory connections=straight lines, inhibitory connections=dotted lines.
pair tasks involving delayed decisions (working memory) and inhibition of alternatives (attention). It seems that a specific level of dopamine receptor activity is optimal in maintaining and stabilizing representations and in directing attention.

Dopamine acts on single neurons in prefrontal cortex primarily via D1-receptors on excitatory (pyramidal) cells and D2-receptors on inhibitory cells (Goldman-Rakic, Lidow, & Gallagher 1990), (Goldman-Rakic et al. 1992). D1-receptor activation changes the efficacy of calcium-activated afterhyperpolarizing potassium channels (Vergara et al. 1998), which leads to an altered signal transmission compared to unmodulated neurons. Basically, a modulated neuron resets quickly, and can process many signals in quick succession faithfully, while the unmodulated neuron shows signal dampening effects over a period of up to 400 ms. In the latter case, we have signal masking effects, where the strength of a synaptic input that is needed to induce firing diminishes with the interval to the previous spike, while this "relative refractory period" is reduced for a modulated neuron (Scheler & Fellous 1999). The detailed single-neuron models used in this work and their physiologically motivated parameters are given in appendix A.

The effects of D2-receptors have been less clearly characterized. Their main effects are a reduction of the size of excitatory postsynaptic potentials by limiting glutamate efflux and a depolarization of inhibitory neurons, i.e. an increase in the amount of inhibition. There are considerably more D1-receptors in rat and monkey prefrontal cortex than D2-receptors, and D2-receptors need a much higher level of dopamine to become active (Goldman-Rakic et al. 1995). In our model D2 activa-

Figure 4: Oscillatory model of working memory. The bottom row shows the spontaneous activity of weakly (link density(ld)=0.01, synaptic weights(sw)=0.05) connected neurons. The top row shows the activity of more strongly connected neurons (ld=0.02, sw=0.2). The center part shows the sequential activation of seven attractors. Input is received only during the first 50 ms. The activity of inhibitory neurons is not shown.

vation is mainly needed to stabilize network activity at higher levels of D1-activation.

When we look at the oscillatory memory buffer that represents the set of active memory items, we note that they are tightly synchronized and have comparatively small delays. As long as single neurons in the attention module and in the working memory module have the temporal properties of modulated neurons, the network is stable and inputs from the attention module can be directed either at active items or at semi-active groups. (cf. Fig. 5).

With neurons that have temporal dependencies in their signal processing, the precise coincident timing of individual spikes within the oscillatory cycle is lost. This leads to a "sprinkler-like" activation of memory items by the attentional module with a consequent loss of either attentional boosting and stabilizing of memories or of fast and reliable switches. Instead, at a low dopamine level, we would expect a reduced focus with consequent impaired performance on delayed decision tasks, and an increased (involuntary) distractability, such as performance on the Stroop or Eriksen tasks (Cohen, Dunbar, & McClelland 1990),(Servan-Schreiber et al. 1998) (cf. Fig. 6).

Neuromodulation of Word Meaning

Access

Concerning the neuromodulation of word meaning processes, data have been primarily collected from the ob-
similar view see (Cohen, Dunbar, & McClelland 1990)). We may assume that a whole activated structure from long-term memory is "loaded into" working memory, i.e. that it is activated and attentionally controlled from working memory. This would involve a change of representation to an oscillatory attractor type network (cf. (Cohen & Knudsen 1999) for analogous representational changes in auditory perception).

This change of representation could be brought about by the activation of a specialized architecture as shown above, and the recruitment of neurons to support that representational format. In that sense, strongly activated neurons will become oscillators, while weakly activated neurons will become semi-active neurons. The LTM structure shown in Fig. 2 can thus be transformed into a representation as in Fig. 4.

The representation of tasks that guide attentional processes is replaced by a contextual representation of cues that guide the meaning selection process. These may now boost certain weak aspects of the representation and automatically suppress certain strong features. A compromised attentional system, which has non-modulated neurons, i.e. a lack of dopamine in the prefrontal cortex, will fail to provide this attentional switch fast and effortlessly in all cases.

A certain difficulty with this system is of a theoretical nature. With the current separation of attentional selection module and working memory rehearsal ("storage") system, it does not become clear why there should be a change in representation at all. The specific nature of working memory representations probably reflects a more integrated function of storage and selection than described here.

We have assumed that an activated meaning structure consists only of semantic features so far, yet word meaning activation under conditions of altered catecholaminergic activity also shows patterns of interference from phonological storage and episodic memory. Presumably a spatial segregation of processing areas for different types of information is compromised by a lack of attentional selection of relevant information. This closeness of different types of associational relations actually resembles an earlier developmental stage, when a more effortful, attention-intensive processing with a greater prefrontal contribution is present (Krasnegor, Lyon, & Goldman-Rakic 1997). We may assume that the failure to distinguish between e.g. phonological and semantic associations in schizophrenia is due to the involvement of larger areas in prefrontal cortex and at the same time their comparatively compact representations (with respect to parietal cortex). Then the selection mechanisms for semantic information, which we have shown to produce only weak preferences at a low dopaminergic tone may be overridden by an equally weak preference for a phonologically associated representation.

Conclusion
We have presented a neural network model of selective attention and semantic memory that employs single neurons with the basic characteristics of integrate-
and-fire neurons, but adaptable to different levels of dopaminergic activation.

Using this type of model neuron, we have built a two-module network consisting of a short-term storage device and a central executive function network which implements the function of directing attention. The short-term storage function was realized by an oscillatory buffer consisting of sequences of activation of tightly synchronized groups of neurons. We propose that activated meaning structures from long-term memory are present in short-term memory during the tasks of word meaning understanding and generation. The functioning of working memory depends on a specific dopaminergic tone that corresponds to an optimal activation of D1-receptors and their counterbalance by D2-receptors. This dopaminergic tone is reflected as certain temporal signal processing characteristics of single neurons. We could show that the functions of stability of representation, selectivity and inhibition of interference effects depend on these neuronal properties. Specific impairments in word meaning access observed in schizophrenics and attributed to the general low dopaminergic tone in prefrontal cortex can thus be explained by the given model.

Appendix A: Single-Neuron Models

The basic single-neuron model adopted in this work is that of an integrate-and-fire unit, i.e. a model without dendritic structure. Temporal varying properties are rendered by parameters for resistance (R) and driving force (E) (cf. (Stevens & Zador 1998)).

\[ V = \Delta V \times \exp\left(\frac{t - t_0}{\tau \times R(t) \times (1 + E_2(t))}\right) - E(t) \]

\[ E(t) = E_0 \times \exp\left(\frac{t - t_0}{\tau_E}\right) \]

\[ R(t) = R_0 \times \exp\left(\frac{t - t_0}{\tau_R}\right) \]

\[ E_2(t) = E_2^0 \times \exp\left(\frac{t - t_t}{\tau_{E2}}\right) \]

\( E_2^0 \) is initially set to 0. Immediately after each firing event, \( E_2^0 \) is increased by \( \Delta E_2 \) until a limit of \( E_2^{max} \) is reached. Several different variants of the basic model are used in the simulations:

- For inhibitory neurons, a basic, non-temporal formulation of the model was used. Inhibitory neurons show low accommodation effects and are presumably not modulated by D1-receptors.
- Excitatory (pyramidal) cells were modeled using measurements from rat prefrontal slices. We distinguish between a baseline condition (no dopamine receptor activity) and a modulated condition (D1-receptor activity). Furthermore certain intermediate values have been chosen, which may be indicative of partial receptor activity.

- the single-neuron model used in Fig. 2:
  \[ \tau = 20 \text{ ms}, \tau_R = 450 \text{ ms}, \tau_E = 125 \text{ ms}, R_0 = 1.8, E_0 = -8 mV, \Delta E_2 = 0.9, \tau_E2 = 12.5 \text{ ms}, E_2^{max} = 2.0, \theta = -44 mV, \text{ synaptic delay = 2 ms}. \]

  In comparison to the measured baseline condition, this has a lower accommodation effect (\( \tau_R \) is longer, \( \tau_E \), \( \tau_E2 \) are shorter, and \( E_2^{max} \) is lower), which serves to keep up activity in the chosen network structure.

- the single-neuron model used in Fig. 4:
  \[ \tau = 22.5 \text{ ms}, \tau_R = 37.5 \text{ ms}, \tau_E = 135 \text{ ms}, R_0 = 8, E_0 = -1.1 mV, \Delta E_2 = 0.0, \tau_E2 = 250 \text{ ms}, E_2^{max} = 0.0, \theta = -46 mV, \text{ synaptic delay = 2 ms}. \]

  These are the parameter settings that provide a good fit with the measured input-output function of a D1-modulated neuron.

Appendix B: Architecture of the working memory model

For the model used in Fig. 3 and Fig. 4 the following link densities (ld) and synaptic weights (sw) for functional groups of neurons were used: neurons within a representational attractor: ld=1.0 sw=0.8, between attractors: ld=1.0 sw=1.0, between attractors and semi-active neurons: ld=0.02 sw=0.2, from inhibitory neurons to attractors: ld=1.0 sw=2.0, from attractors to inhibitory neurons: ld=1.0 sw=2.0, between attractors and background neurons: ld=0.01 sw=0.05.

For the attentional model (Fig. 5 and 6) the following additional architectural parameters have been used: between attentional neurons and semi-active neurons: ld=1.0 sw=0.2, from inhibitory neurons to attractors: ld=1.0 sw=-2.0, from inhibitory neurons to semi-active neurons: ld=1.0 sw=-1.5, from attentional neurons to attractors: ld=1.0 sw=2.0, from attentional neurons to semi-active neurons: ld=1.0 sw=1.0, between attractors: ld=1.0 sw=1.0, between semi-active and background neurons: ld=1.0 sw=-2.0, between semi-active neurons: ld=1.0 sw=-2.0, from semi-active neurons to attractors: ld=1.0 sw=0.6, from background neurons to semi-active neurons: ld=1.0 sw=2.0, from background neurons to attractors: ld=0.01 sw=0.05.

For simplicity the increase of inhibitory activity due to D2-receptor activation has been modelled (incorrectly) by a greater synaptic strength (indicated by BL for the baseline, non-modulated case and D for the dopamine-modulated case). The net result of more inhibition in the network for the modulated condition, however, is the same.

References


Vergara, C.; Latorre, R.; Marrion, N.; and
