

A Brain Model of the Relationship Between Semantic Memory and Working Memory in Semantic Cognitive Tasks

Alan Newberger (newbergr@princeton.edu)
Princeton University, Department of Philosophy
Princeton, NJ 08544 USA

Jack J. Gelfand (jgg@princeton.edu)
Princeton University, Department of Psychology
Princeton, NJ 08544 USA

Abstract

We have simulated the brain mechanisms involved in semantic processing tasks such as the differentiation of two sequential stimuli based upon recalled semantic features. In this paper we examine the relationship between perceptual inputs, working memory and semantic memory in these tasks. We propose that phase synchronous firing binds features in semantic memory with concepts in working memory, and that a phase comparison mechanism subserves the process of response selection. The model is consistent with the anatomy and physiology of the component brain circuits where known. This research is important because the relationship between working memory and long-term memory is a central component of many theories of cognition.

Semantic Memory and Working Memory in Cognitive Tasks

A link from long-term memory to working memory is required for those cognitive processes that require information in addition to that provided by the perceived

input. This is a fundamental organizational principle of many cognitive models such as ACT and SOAR (Anderson, 1993; Newell, 1990). In this paper we consider a simple task of this type and attempt to model the brain circuits involved. The task is one in which a question is posed about the semantic features of two sequentially presented stimuli with a short interval in between. This task might consist of a 1 second presentation of a picture of an bear, a 10 second blank screen and a 1 second presentation picture of an deer. The question posed in advance is which of these two animals hibernates. What we learn from modeling this simple process can be extended to tasks concerning more complex comparisons and relations between the two input stimuli.

The purpose of this project is to understand the overall operation of the brain at the systems level for higher cognitive functions. Where possible, we have tried to faithfully follow what is known about the anatomy and physiology of the brain regions involved. In other cases abstracted functional models were used.

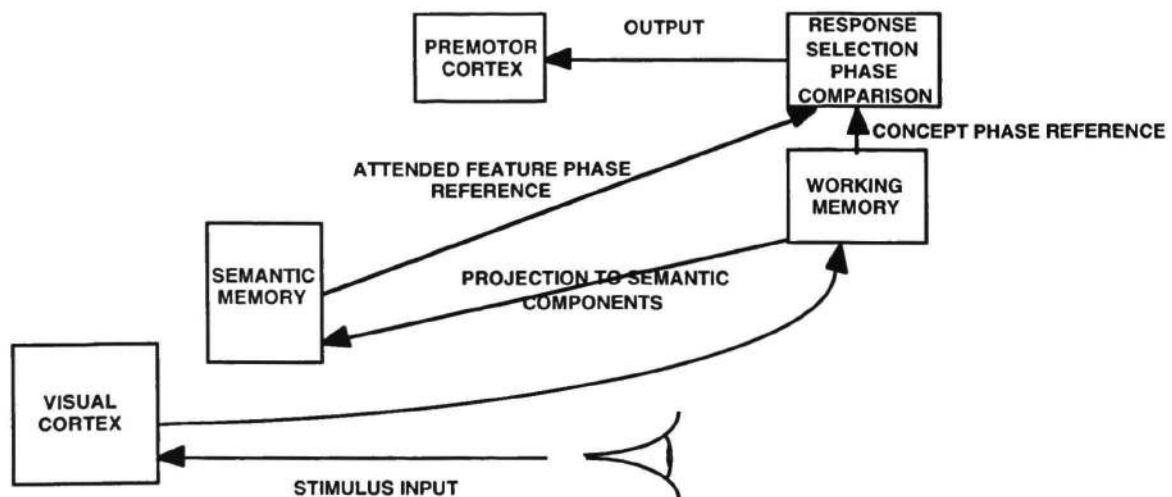


Figure 1: A schematic representation of the brain systems simulated in the model

Model Architecture

A schematic diagram of the overall system is shown in Figure 1. In the type of task simulated here, the stimulus is presented in picture form. This requires that the visual system identify the stimulus using the dorsal visual stream ending in the inferotemporal cortex (IT) (Ungerleider & Mishkin, 1982). There is a strong projection in monkeys from IT to prefrontal cortex through the uncinate fascicle and we assume that a similar projection exists in this model (Ungerleider, Gaffan & Pelak, 1989)

The principal feature of the system is that the perceived stimulus input is projected to working memory and then a projection from working memory to semantic memory excites those semantic features that are associated with the input stimuli. We utilize synchronous firing of neurons to bind these representations in the two memory systems and a phase comparison mechanism to select the appropriate output. There is considerable evidence that the working memory area associated with semantic functions is in left inferior prefrontal cortex and that it is closely linked to semantic memory areas in left temporal association cortex (Buckner, 1996; Jennings, McIntosh, Kapur, Zipursky & Houle, 1998; Nyberg, Cabeza & Tulving, 1996; Thompson-Schill, D'Esposito, Aguirre & Farah, 1997).

In this model a mechanism for persistent excitation in working memory excites representations of semantic features associated with the input stimuli for the duration of the task. An attentional circuit focuses attention on the semantic feature or features appropriate to the task. Excitation of these features plus the representations of the input stimuli in working memory are then transferred to a selection circuit in PFC whose output goes to premotor cortex. Significant projections exist from prefrontal cortex to premotor cortex to subserve the output connection (Lu, Preston & Strick, 1994).

Prefrontal Working Memory Circuits

Short term persistent response of prefrontal cortex neurons has been postulated as a mechanism for short-term memory (STM). The prefrontal cortex has specific reciprocal projections with the thalamus, resulting in local cortico-thalamic loops. When activated, these loops can sustain activity in frontal cortex neurons (Alexander, Crutcher & DeLong, 1990; Groenewegen & Berendse, 1994; Houk, 1995; Selemon & Goldman-Rakic, 1985). The details of these circuits and their functionality has been described by Gelfand et al. (Gelfand, Gullapalli, Raye, Johnson & Henderson, 1997)

The overall architecture for the postulated prefrontal working memory circuits and their relation to perceptual and semantic memory areas is shown in Fig. 2. This figure expands the portion of Fig. 1 from the highest visual area through working memory to semantic memory. It shows the relevant anatomical details mentioned in the discussion below. We assume that PFC, IT and semantic memory in temporal association cortex are organized as multiple groups of mutually inhibitory neurons which correspond to cortical columns (Goldman-Rakic, 1995; Mountcastle, 1978). These columns are indicated by groupings of neurons in the figure. The columns are not necessarily adjacent to each other in each cortical area but may be anatomically distributed.

We used a distributed representation over the input cortical columns with cortico-cortical interconnections between columns formed through Hebbian learning to encode the stimulus in visual cortex. In this representation, each column represents a class of features. Within each column, neurons or groups of neurons encode a particular feature in the class whose excitation is associated with that stimulus.

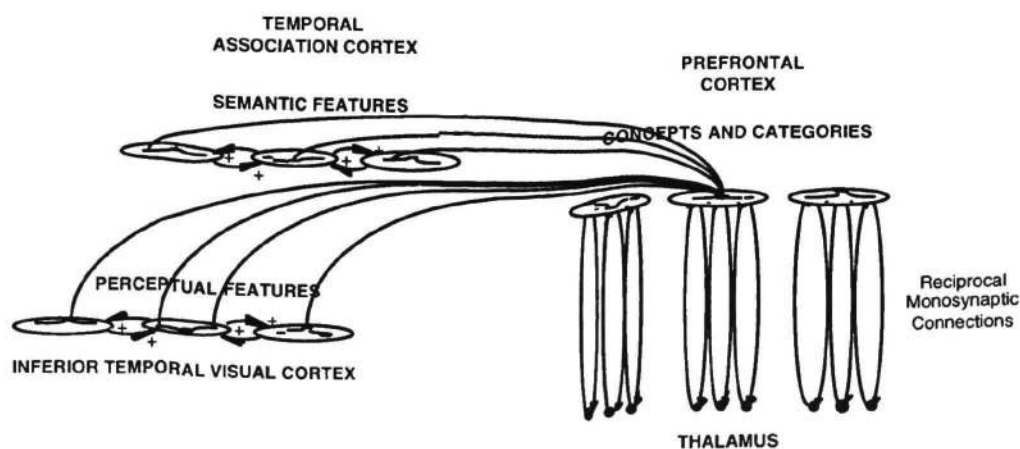


Figure 2: Schematic diagram of the projections between IT cortex, working memory and semantic memory.

We assume that these columns have groups of neurons representing concepts and that the columnar regions are organized as groups of concepts in categories. This organization is consistent with proactive inhibition and the release from proactive inhibition phenomena in short-term memory (Gelfand et al., 1997).

Binding of Semantic and Working Memory

In this simulation, the semantic representation of a concept is encoded as a constellation of features over cortical columns. For cognitive tasks that require the mediation of working memory, the individual features in semantic memory are activated by projections from neurons in working memory representing the concepts being processed. These projections are shown in Figure 2. Because oscillations in thalamocortical loops produce persistent concept-based storage in prefrontal cortex, these projections will produce persistent excitation of the semantic features associated with a particular concept. This referential process is not unlike pointers in symbolic systems and resembles the relationship between working memory and declarative memory in the ACT-R cognitive model (Anderson, 1993).

The representational capacity of neurons probably does not permit them to hold information complex enough to be construed as a pointer. So information in two separate cortical areas require some means of integration. This is a type of binding problem. One class of proposed solutions to the binding problem makes use of the bursting nature of neuronal activity. Neurons which fire in synchrony, and in phase, are to be regarded as referring to the same representational entity (Singer, 1993). We have chosen to use phase information to bind together concepts in working memory with component features in semantic memory.

In our model neurons representing different concepts in working memory each oscillate at a different phase. This is based upon Lisman and Idiart's proposal that nodes representing individual items in prefrontal cortex oscillate in distinct phase from one another (Lisman & Idiart, 1995). Concepts in working memory are bound to their semantic components by oscillating in the same phase. Through excitatory projections, a constellation of neurons in semantic memory representing the component semantic features of a concept are driven to fire synchronously with the neurons in working memory that excite them. This assumption is justified experimentally by the data of Sarnthein et al. who have shown that there is synchronous activity between prefrontal and posterior association cortex during cognitive tasks (Sarnthein, Rappelsberger, Shaw & von Stein, 1998).

Response Selection

Our task involves the differentiation of two stimuli based on a stored semantic feature. The appropriate feature in semantic memory must be retrieved and compared with the activated concepts in PFC. Because this relationship is based upon synchronous firing, we must use a phase comparison to execute this manipulation. We propose that a separate area of prefrontal cortex, distinct from working

memory, is responsible for this process. This is based upon the data of D'Esposito et al. who have found distinct areas of prefrontal cortex are responsible for short-term memory and for other cognitive processing (D'Esposito et al., 1997).

An attentional circuit is responsible for passing the phase of the appropriate feature to the response selection circuit. The response selection circuit performs a comparison of the phase of the selected feature with the phases of the concepts that are activated in working memory. It selects as its output that concept which is in phase with the task appropriate feature. This output feeds into the premotor cortex to initiate the subject's actual response. For each output from working memory there is a cell, or complex of cells, which performs this phase comparison. Phase comparator circuits of this kind have been found in the auditory system of the barn owl (Spence & Pearson, 1995).

Model Simulation

Our simulation is based in part on the previous work of Gullapalli and Gelfand, and uses the same network simulation system, written in C++ (Gelfand et al., 1997; Gullapalli & Gelfand, 1995). We chose to model the following specific task, diagrammed in Figure 3. A subject is first instructed to verbally respond as to which of 2 visual stimuli is more dangerous. Then, in succession, the subject is presented with a picture of an apple, and a picture of an alligator for 1 second each with a 5 second delay in between.

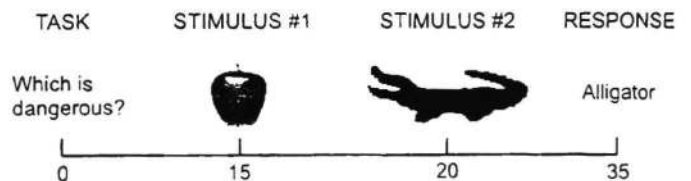


Figure 3: A schematic time line for the task simulation

We trained the visual system network to recognize 12 different objects, including apple and alligator, based upon visual features. Visual cortex in the model projects to working memory which because of the feedback loops between thalamus and PFC causes a persistent excitation after the stimulus is removed. The persistent excitation in the neurons encoding the first stimulus lasts for a period of time which overlaps with the second stimulus before decaying. The area of association cortex attributed to semantic memory had 33 features represented including the target feature, dangerous. We hand coded weights for the strengths of projections to semantic memory in a fashion that was consistent with the semantic definition of the stimuli.

Results

The stimuli apple and alligator were presented to the simulation in sequence. Figure 4 plots the activity of the prefrontal cortex modules during the simulation. Both apple and alligator activate during stimulus presentation, and persist with slow decay due to feedback loops. The oscillations of the concept nodes are not in phase with each other. The activity of the other nodes is present in Figure 4 but negligible. Figure 5 illustrates the output of the response selection module. When apple is activated, it does not excite the feature representation of dangerous, and so dangerous does not oscillate in phase with apple. The

central executive module whose attentional circuit is focused on the semantic feature, dangerous, has no output for the first half of stimulus presentation. In this case, no response selection nodes display excitation during the first stimulus.

When alligator is present, the semantic feature, dangerous, is excited causing the excitation passed through the central executive to serve as a reference in the phase comparator circuit. Because this oscillation is in phase with alligator, the alligator output is excited. Figure 5 illustrates this. The response selection output node whose input representing alligator begins to oscillate when the alligator stimulus is introduced.

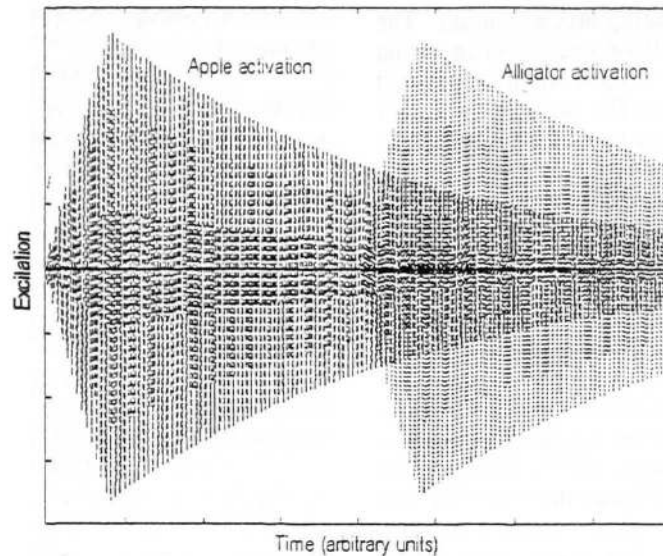


Figure 4: Excitation of the neuronal assemblies representing APPLE and ALLIGATOR in the PFC

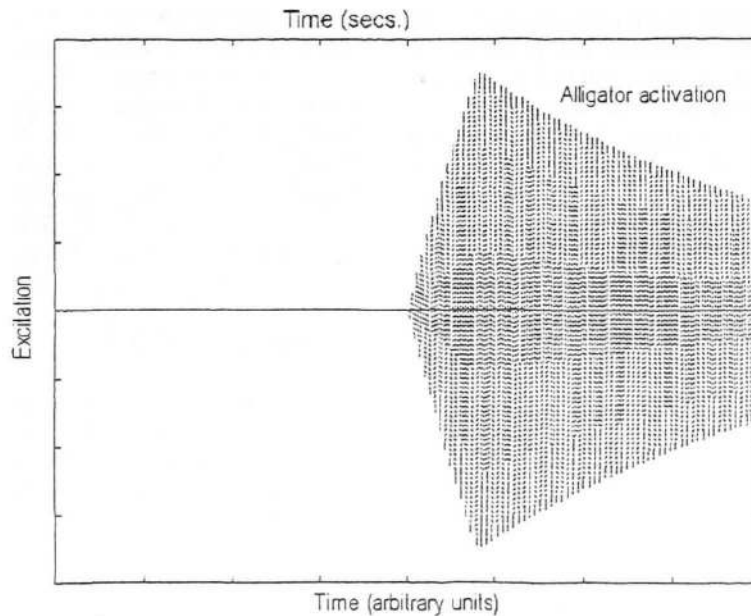


Figure 5: Output of response selection circuit to premotor cortex for the response, ALLIGATOR

Discussion

The routing of signals in the simulation in this paper is task dependent. For example in this simulation the initial statement of the task requires the circuit to pay attention to a particular semantic feature, in this case dangerousness, and transfer its phase to the phase comparator in the response selection circuit. Another task such as a rank ordering of stimuli would require different semantic knowledge and different manipulations. We presume that there must be a circuit that provides this function of problem specific routing and processing. This module would implement the central executive function postulated by Baddeley (Baddeley, 1986).

The primary observable property in human subjects that has some bearing on the correctness of our model is the time course of activation in the brain during task execution. The simulation predicts that there will be excitation in visual cortex immediately following the stimuli and a subsequent transfer of excitation to PFC for the working memory function. This excitation will then be transferred back to posterior regions associated with semantic memory. Finally excitation will then again go to PFC for the response selection circuit. This process has actually been observed by Dale for similar cognitive tasks using event related potential measurements (Dale, 1997).

We would also predict that synchrony of oscillations will be observed between posterior association cortex and PFC during semantic tasks. Synchrony has been reported for related tasks. Sarnthein has found synchronous oscillations for PFC and posterior association areas for delayed response tasks as mentioned earlier (Sarnthein, 1998). Desmedt and Tomberg report synchronous oscillation between PFC and parietal cortex during a finger shock perception task (Desmedt & Tomberg, 1994). We would expect that synchrony would also be found for semantic tasks of the type modeled here.

Our approach is similar to the work of Shastri in that we both postulate neural mechanisms which perform operations based on phase information. Shastri has proposed a functional similarity to circuits in hippocampus which can subserve rapid memory formation (Shastri, 1997). The purpose of our research is to examine mechanisms that subserve reasoning processes which become possible with the addition of working memory capability. This allows us to examine cognitive processes involving the manipulation of multiple stimuli presented sequentially. Also, the introduction of an ordering mechanism in the working memory circuit would allow for an account of language processing which is more general (Chomsky, 1957).

Acknowledgments

We acknowledge many helpful discussions with Mark D'Esposito, Vijay Gullapalli, Philip Johnson-Laird, Marcia Johnson, John Kounios, Carol Raye and Clay Spence.

References

- Alexander, G., Crutcher, M., & DeLong, M. (1990). Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In H. Uylings, C. Van Eden, J. De Bruin, M. Corner, & M. Feenstra (Eds.), The prefrontal cortex: its structure, function and pathology. Amsterdam: Elsevier.
- Anderson, J. R. (1993). Rules of the Mind. Hillsdale, NJ: Erlbaum.
- Baddeley, A. (1986). Working Memory. London: Oxford.
- Buckner, R. (1996). Beyond HERA: Contributions of specific prefrontal brain areas to long-term memory retrieval. Psychonomic Bulletin and Review, 3, 149-158.
- Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton & Co.
- D'Esposito, M., Ballard, D., Tang, J., Lease, J., Zarahn, E., & Aguirre, G. (1997). Maintenance and manipulation of information held in working memory: A fMRI study, Abstract 657.3, Abstracts, 27th Annual Meeting, Society for Neuroscience, New Orleans.
- Dale, A. (1997). Strategies and limitations in integrating brain imaging and electromagnetic recording, Abstract 4.3, Abstracts, 17th Annual Meeting, Society for Neuroscience, New Orleans.
- Desmedt, J., & Tomberg, C. (1994). Transient phase-locking of 40 Hz electrical oscillations in prefrontal and parietal human cortex reflects the process of conscious somatic perception. Neuroscience Letters, 168, 126-129.
- Gelfand, J., Gullapalli, V., Raye, C., Johnson, M., & Henderson, J. (1997). The Dynamics Of Prefrontal Cortico-Thalamo-Basal Ganglionic Loops and Short-Term Memory Interference Phenomena. In M. Shafto & P. Langley (Eds.), Proceedings of the 19th Annual Conference of the Cognitive Science Society, Stanford, CA., (pp. 253-258). San Diego, CA: Erlbaum.
- Goldman-Rakic, P. (1995). Cellular Basis of Working Memory. Neuron, 14, 477-485.
- Groenewegen, H., & Berendse, H. (1994). Anatomical relationships between the prefrontal cortex and basal ganglia in the rat. In A. Thierry, J. Glowinski, P. Goldman-Rakic, & Y. Christen (Eds.), Motor and Cognitive Functions of the Prefrontal Cortex. Berlin: Springer-Verlag.
- Gullapalli, V., & Gelfand, J. (1995). A model of the dynamics of prefrontal cortico-thalamo-basal ganglionic loops in verbal response selection tasks. In J. Grafman, K. Holyoak, & F. Boller (Eds.), Structure and Functions of the Human Prefrontal Cortex. New York: New York Academy of Sciences.
- Houk, J. (1995). Information processing in modular circuits linking basal ganglia and cerebral cortex. In J. Houk, J. Davis, & D. Beiser (Eds.), Models of information processing in the basal ganglia, (pp. 3-9). Cambridge: MIT Press.

Jennings, J., McIntosh, A., Kapur, S., Zipursky, R., & Houle, S. (1998). Functional network differences in schizophrenia: A rCBF study of semantic processing. Neuro Report, in press.

Lisman, J., & Idiart, M. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. Science, *267*, 1512-1515.

Lu, M.-T., Preston, J., & Strick, P. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. J. of Comparative Neurology, *341*, 375-392.

Mountcastle, V. (1978). An organizing principle for cerebral function: The unit module and distributed function. In G. Edelman & V. Mountcastle (Eds.), The mindful brain, . Cambridge: MIT Press.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, MA: Harvard University Press.

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieving: The HERA model. Psychonomic Bulletin and Review, *3*, 135-148.

Sarthein, J., Rappelsberger, P., Shaw, G. L., & von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during working memory tasks in humans. P.N.A.S., in press.

Selemon, L., & Goldman-Rakic, P. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. J. Neurosci., *5*, 776-794.

Shastri, L. (1997). A model of rapid memory formation in the hippocampal system. M. Shafto & P. Langley (Eds.), Proc. 19th Annual Conference of the Cognitive Science Society, Stanford, CA., (pp. 680-685). San Diego, CA: Erlbaum.

Singer, W. (1993). Synchronization of cortical activity and putative role in information processing and learning. Ann. Rev. Physiol., *55*, 349-374.

Spence, C., & Pearson, J. (1995). Do neurons in the barn owl couple to extracellular potentials ? In J. Bower (Ed.), Computational Neuroscience: Trends in Research. San Diego: Academic Press.

Thompson-Schill, S., D'Esposito, M., Aguirre, G., & Farah, M. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. P.N.A.S., *94*, 14792-14797.

Ungerleider, L., Gaffan, D., & Pelak, V. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. Experimental Brain Research, *76*, 473-484.

Ungerleider, L., & Mishkin, M. (1982). Two Cortical Visual Systems. In D. Ingle, M. Goodale, & R. Mansfield (Eds.), Analysis of Visual Behavior. Cambridge: MIT Press.