An Accumulator Model Account of Semantic Interference in Memory Retrieval

Leendert van Maanen (leendert@ai.rug.nl)
Hedderik van Rijn (D.H.van.Rijn@rug.nl)
University of Groningen, Artificial Intelligence
Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands

Abstract

To explain latency effects in picture-word interference tasks (Glaser & Düngelhoff, 1984, cf. Stroop, 1935), cognitive models need to account for both interference and SOA effects. As opposed to most ACT-R models, which model the time course in a ballistic manner, the RACE model (Retrieval by ACcumulating Evidence) presented here accounts for semantic interference *during* the retrieval interval. This way, RACE offers a better account of semantic memory retrieval latencies in different interference and SOA conditions. We describe the architectural assumptions of RACE and simulations of a Glaser and Düngelhoff experiment, illustrating the validity of the RACE model.

Introduction

Often, symbolic models of cognition can be thought of as giving a 'stroboscopic' account of cognition. On a high level of analysis, the stroboscopic models reflect a continuous process, but on the level of a single 'flickering', these models provide a discrete account of a cognitive process. In most cases, this higher level of analysis is sufficient to understand cognitive functioning, but in some tasks, a closer examination might prove to be beneficiary.

As an example, consider the way retrieval of memory chunks is modeled in ACT-R: Retrieval latency is based on the activation of the to-be-retrieved memory chunk, as in $RT_i = Fe^{-A_i}$ (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004). If a retrieval request is made to the declarative memory system, the activations of all chunks are compared, and the highest is selected for retrieval. The latency is calculated according to the above equation and, after the appropriate amount of time has passed, retrieval of that chunk is reported. Even if new information is presented between the retrieval request and the actual retrieval, the retrieval result and latency cannot be influenced.

However, many experiments show that information that is presented shortly before or after a target stimulus can influence both the timing and accuracy of the task at hand (e.g., MacLeod, 1991; Neely, 1991). In a picture-word interference task for example, subjects respond slower in the picture-naming task when a distractor word is presented, even if the distractor word is presented shortly *after* the target stimulus.

A candidate explanation for this phenomenon comes from the field of choice behavior modeling. In sequential sampling models of simple choice behavior, the choice between candidates is modeled by competition between candidates. Sequential sampling is based on the idea that choosing one option over the other is based on sampling of inherently noisy neural representations of these choices, until one has sampled enough evidence to be chosen (Ratcliff & Smith, 2004). The RACE model presented in this paper is very similar to a specific instance of sequential sampling models: The leaky competing accumulator model of Usher and McClelland (2001).

RACE is implemented using the same basic principles as the leaky competing accumulator model: (a) it consists of a set of non-linear stochastic accumulators, all of which represent one memory chunk that can be chosen. (b) The accumulator units are activated by external input and recurrent activation, but are decreased by lateral inhibition and decay. However, the actual implementation of some aspects differs, most importantly different activation and evidence accumulator functions, both of which have been adapted to fit RACE in the ACT-R framework.

RACE Architecture

RACE stands for Retrieval by ACcumulating Evidence, which reflects both the accumulation of evidence for memory representations and the competition between memory chunks during retrieval: The comparison with a race between chunks seems appropriate in this respect.

The activation levels of memory chunks in RACE consist of two components: A long-term component that governs the global activation of chunks and a short-term component that comes into play during the retrieval process. The long-term component is represented by the ACT-R base-level activation equation (Anderson et al., 2004),

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right),$$

where t_j is the time since the jth presentation of a memory chunk and d is the parameter that controls decay, which is fixed at 0.5. The idea is that memory decays over time unless attention is shifted to a memory chunk and its memory traces are strengthened.

RACE's short-term component, called context activation (C_i), is continuously computed from the moment that a request for retrieval of a chunk is made. The context activation of chunks changes as a consequence of positive and negative influences from other chunks. Chunks from the same chunk type inhibit each other, thereby competing for context activation increase. Chunks of different chunk types excite each other, spreading their activation in the classical sense (Collins & Loftus, 1975). Thus, by continuously updating positive and negative spreading activation, some chunks may reach a level of activation at which retrieval can take place.

The context activation can be described as a system of two dependent equations:

$$C_i(t) = C_i(t-1) + e^{\beta E_i(t)-1} - d^{\text{context}} \cdot \ln(t) \quad (1)$$

$$E_i^k(t) = \sum_{j \notin k} e^{A_j(t-1)} S_{ji} - \sum_{l \in k} e^{A_l(t-1)} S_{li} \quad (2)$$

As stated earlier, these equations incorporate the basic assumptions of Usher and McClelland (2001), and are adapted to the ACT-R framework.

The system functions as follows: At every time step, positive (first component of eq. 2) and negative (second component of eq. 2) associative values towards a memory chunk are computed, and the difference is calculated. This is called the *net evidence* ($E_i^k(t)$) of chunk *i* of chunk type *k* at a certain time. Since relative, not absolute activation values are what count in ACT-R, an exponential scaling is applied to calculate net evidence. There are two types of sources of activation in RACE: Chunks (*l* in eq. 2) of the same chunk type (*k*) spread negative activation to each other, while chunks (*j*) of different types spread positive activation. This is similar to neurobiological findings from which it is clear that lateral inhibition between cortical representations of visual stimuli (Kastner, De Weerd, Desimone, & Ungerleider, 1998) as well as excitatory projections to other cortical layers (Callaway, 1998) exist. Note that most ACT-R models do not place constraints on the functional role of chunk types (although it does play a role in production compilation, Taatgen, 2005).

At each point in time, the net evidence determines the context activation growth (eq. 1). Context activation increases exponentially according to the amount of net evidence and a scaling factor β. If net evidence is negative (i.e., more inhibition than excitation), then growth is negative. At all time steps, evidence decays with $d^{\text{context}} \ln t$ (eq. 1), in which t is the time since the start of the accumulation and d^{context} a decay parameter. This way, context decay in RACE resembles decay in the ACT-R optimized learning equation.

The activation of a chunk at any time is the sum of base-level and context activation, plus a small normally distributed noise sample. A chunk is retrieved if this total activation crosses the context threshold[1]. The retrieval latency is defined as the time between the retrieval request and the time that the total activation of a matching chunk reaches the context threshold.

If no evidence is sampled, context activation decays away. Therefore continuous positive reinforcement from evidence is necessary for successful retrieval, and absence of positive evidence results in prolonged retrieval latencies or failures.

Picture-Word Interference

One of the most well-known experimental paradigms in cognitive psychology is the Stroop-task (Stroop, 1935; cf. Dyer, 1973), where, in the original setup, subjects have to either name the color a word is written in, or read the word, which is always a color name. It turns out that naming the color is much more difficult than reading the word – especially if color and word of a single stimulus do not correspond – as is reflected in increased reaction times and decreased accuracy in the color naming condition. The Stroop-task can be regarded as an instance of a more general class of experiments that demonstrate interference effects in various naming tasks between pictoral stimuli and word-form stimuli. These experiments are generally called picture-word interference experiments (Glaser & Düngelhoff, 1984; cf. MacLeod, 1991). In the case of the Stroop-task, the pictoral stimulus is the word color.

We tested the RACE model in a picture-word interference task, using two tasks and four different conditions, similar to the experimental setup in (Glaser & Düngelhoff, 1984, Experiment 1). One task consisted of reading a word (target stimulus) while a picture is presented as distractor; the other task consisted of naming the depicted item (target stimulus), while a word is presented (distractor). In both tasks, the distractors were presented on different SOAs (stimulus onset asynchronies). If a distractor was presented at negative SOA, it was presented *before* the target stimulus. At positive SOAs, the distractor was presented *after* the

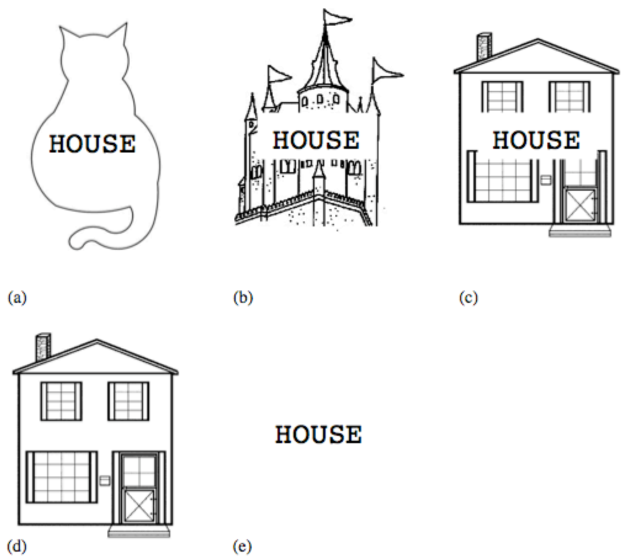

Figure 1: Example stimuli at zero SOA. (a) Incongruent condition, (b) category-congruent condition, (c) concept-congruent condition, (d) neutral condition in picture naming, (e) neutral condition in word reading.

[1] The context threshold is a different concept from the retrieval threshold in default ACT-R. Where the retrieval threshold determines the *minimum* activation at which a chunk *may be* retrieved, the context threshold determines the amount of activation at which a chunk *is* retrieved.

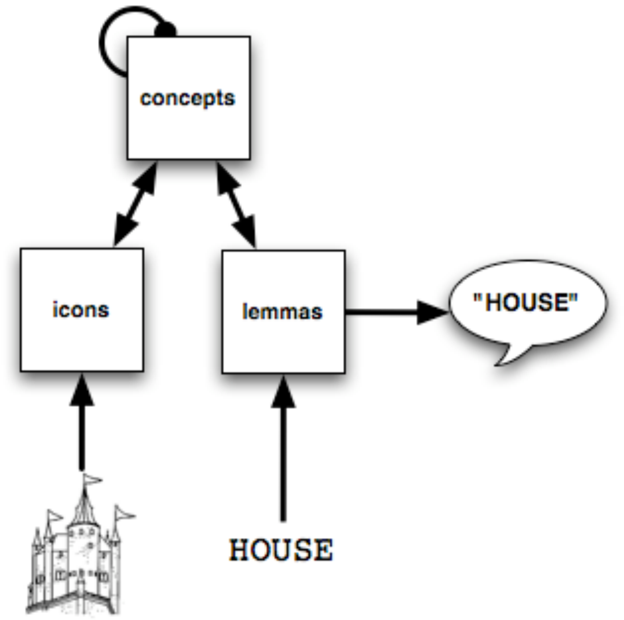

Figure 2: Processing route for pictures and words in the picture-word interference model. The route for words is shorter, since words do not require concept retrieval. Interference takes place between concept type chunks

target stimulus. The first condition was one in which both distractor and target stimulus refer to the same concept. This is referred to as the concept-congruent condition. In two other conditions, distractor and target stimulus refer to different concepts. In the category-congruent condition the concepts belong to the same semantic category (e.g., a picture of a house and the word "castle" were presented. Figure 1 presents example stimuli), in the incongruent condition the concepts do not belong to the same semantic category (e.g., a picture of a house versus word "cat"). As control condition only target stimuli were presented, meaning that either only pictures or only words were presented.

Glaser and Düngelhoff (1984) found that interference is highest in the category-congruent condition, which is known as the *semantic gradient effect*. They also showed facilitation in the concept-congruent condition, meaning that latency is decreased when both target and distractor stimuli refer to the same concept. A third effect they report is a clear asymmetry between the picture-naming task and the word-reading task. The semantic gradient and facilitatory effect virtually disappear in the word-reading task, but they are visible in the picture-naming task.

Picture-Word Interference Model

In an earlier model of picture-word interference effects (Roelofs, 1992), multiple types of memory were included in the model. Analogous to this spreading activation model of lemma retrieval, our model of picture-word interference comprises three chunk types: Icons (Roelofs: object-form), lemmas and concepts. The concept chunks can be regarded as representations of semantic properties. Chunks of the icon type represent iconographic instances of the stimuli. Chunks of the lemma type can be regarded as sets of both orthographic and syntactic properties of a word. Note that this is a simplification of Roelofs' (1992) model, in which the response selection stage (choosing a lemma) is followed by response programming and execution stages. Since our focus is on the retrieval process, these vocalization aspects of the task are not included in our model of picture-word interference. Positive activation is spread between chunks of different types. That is, icons spread to concepts and vice versa, and lemmas spread to concepts and vice versa. As in Roelofs' (1992) model, no direct spreading activation was allowed between lemmas and icons. The concept chunks also have negative associations between them and spread negative activation to each other.

At different SOAs, distractor stimuli were presented to the model, except in the control condition, in which only a target stimulus was presented. We tested the same four conditions as in Experiment 1 of Glaser and Düngelhoff (1984). The only deviation from the original experiment was the neutral or control condition: Glaser and Düngelhoff presented the subjects with a non-word distractor and a non-picture distractor respectively in the picture-naming and word-reading control condition. These were chosen in such a way as to minimize the amount of picture or word processing as possible. Assuming a successful operationalization by Glaser and Düngelhoff, we simulated this condition by not presenting a distractor in the neutral condition.

In the concept-congruent condition, the distractor consisted of a word stimulus referring to the same concept as the target, but of a different stimulus type. When activation spread through the model, both distractor and target would thus activate the same chunks, but not in the same order or strength. In the category-congruent condition and the incongruent condition, the distractor and the target referred to different concepts. However, in the category-congruent condition, the associations between the chunks representing these concepts were higher than in the incongruent condition. The assumption here is that concepts from the same category have a higher semantic relatedness than unrelated concepts.

The distractors were presented at SOAs of -400, -300, -200, -100, 0, +100, +200, +300, and +400ms, similar to the original Glaser and Düngelhoff experiment. The stimuli presentations were modeled as a fixed increase in activation of the lemma or icon type chunks during the period that a stimulus was presented.

Since the task was a verbalization of either the picture name or the word, a trial was finished when the stimulus-designated lemma was retrieved or after 2s, indicating a retrieval failure.

In the picture-naming task, the model behaves as follows: In the concept-congruent condition with negative SOAs, a distractor word is presented *before* the target

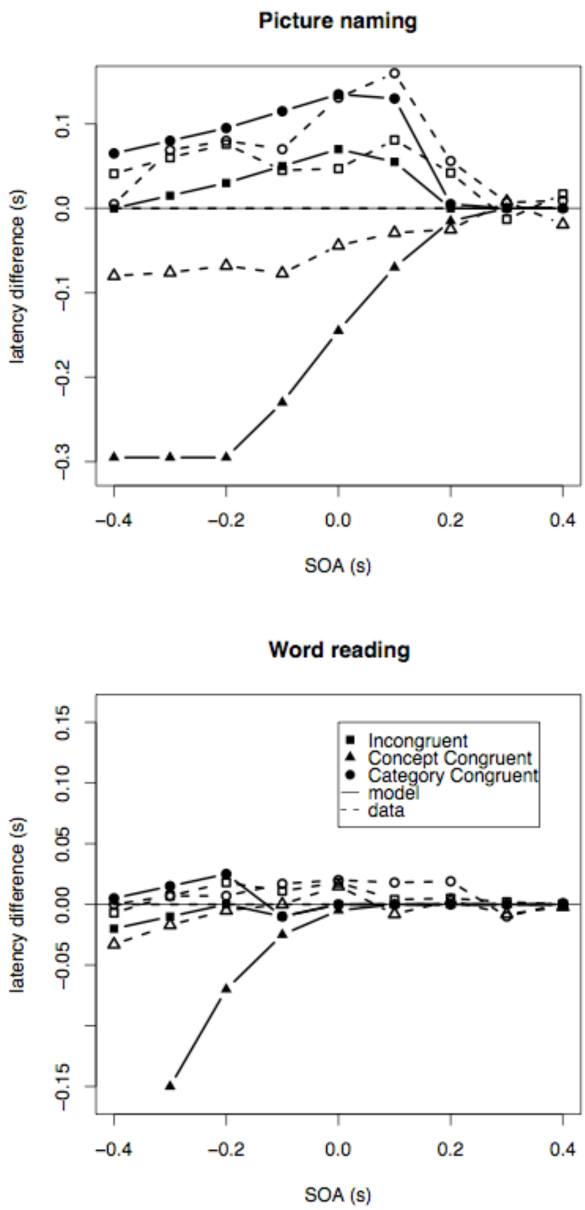

Figure 3: Simulation results and experimental data (Experiment 1, Glaser & Düngelhoff, 1984).

picture. The word activates a lemma chunk, which increases the activation of the associated concept chunks, but inhibits the activation increase of other lemma chunks. The increased activation of the concept chunks increases the activation of the associated icon chunks. Thus, immediately after the distractor is presented, all chunks that should be activated in naming the picture (one icon, one concept, and one lemma chunk) have an increased activation. When the target is presented, all concept-congruent chunks have a higher activation as compared to the stimulus onset in the neutral condition, and thus a shorter retrieval latency. In the concept-congruent condition with positive SOAs, the same process occurs, but to a lesser extent since the lemma's activation has less time to influence the activation before the lemma is retrieved: The picture has already increased the lemma's activation before the word is presented.

In the incongruent and category-congruent situations (both at negative and positive SOAs), the activation of the lemma due to the word stimulus results in inhibition of the target chunks of the icon, concept and lemma type.

Results

Figure 3 summarizes the results of our simulation studies.[2] The figure represents the latency differences in different conditions relative to the neutral condition. Since our model only captures the retrieval part of the picture naming and word reading tasks, we can only compare the latency differences between the different conditions from the model to the data. The observed latencies from the data set also comprise timing effects from other subtasks, such as pronunciation or perceptual encoding.

Negative values in figure 3 mean faster retrieval than in the neutral condition, and positive values indicate slower retrieval than in the neutral condition. The qualitative effects observed in (Glaser & Düngelhoff, 1984, Experiment 1) can also be seen in the predicted latency differences from the RACE model. The semantic gradient effect can be observed by looking at the different relative latencies of the category- congruent en incongruent conditions. The higher latencies in the category-congruent condition as opposed to the incongruent condition indicate that higher associations between concepts result in stronger inhibition. Other simulations indicate that increasing associations between concepts increases latency.

The facilitatory effect in the picture naming task is also apparent, although the effect appears to be too large. Our explanation for this increased effect is that the activation of the target lemma chunk is too high if the target stimulus is presented, probably caused by too little decay after the previous retrieval initiated by the distractor stimulus. This may also explain the observed effect in the simulation of the word reading task.

When looking at the difference between the two simulated tasks (picture-naming and word-reading), the asymmetry observed by Glaser and Düngelhoff is also shown by RACE. We explain this asymmetry by two effects: The shorter processing route and the faster encoding of word type stimuli. As Roelofs (1992, but also Glaser & Glaser, 1989; La Heij, Happel, & Mulder, 1990) noted, pronouncing words does not require retrieving a concept from memory, therefore processing word type stimuli can be much faster than processing pictoral stimuli. In Figure 2 it can be seen that the route in our model from a picture to the associated lemma is much longer than from the word to the associated lemma: Two

[2] An R implementation of RACE and the picture-word interference model can be retrieved from http://www.ai.rug.nl/~leendert/RACE.html.

intermediate steps have to be taken (i.e., processing of icon and concept chunks), before the lemma chunk is retrieved. Thus, potentially interfering pictures do not activate lemma chunks before the target lemma is retrieved. Only at high negative SOAs an effect can be seen (Figure 3), because under that condition there is enough time for the distractor stimulus to activate the inhibiting distractor lemma and interfere with retrieval of the target lemma.

A difference in encoding speed is incorporated to account for the observation that word reading is a faster process than picture naming. Without this difference, pictures are processed nearly as fast as words, and picture naming in the incongruent condition is as fast as in the neutral condition. The faster encoding of words reinforces the effect that the lemma associated with a word is processed before the lemma associated with the picture stimulus is retrieved.

Discussion

We have shown that a sequential sampling model can account for the time course of memory retrieval during asynchronously presented stimuli. This is an extension of the results from Usher and McClelland (2001). They only investigated accumulators with equal onset times. Moreover, RACE can be regarded as an extension of the ACT-R theory of cognition. It combines the long-term base-level activation equation of ACT-R with a short-term context activation used for retrieval.

The general fit of our model of picture-word interference is quite reasonable, thereby indicating that the RACE equations can provide for semantic interference effect in memory retrieval. Retrieval in the concept-congruent condition seems to be too fast, however.

Also, our model accounts for facilitatory effects. In line with the findings from Glaser and Düngelhoff (1984), RACE predicts that semantic facilitation occurs if target and distractor both refer to the same concept.

In the past, ACT-R models of Stroop-like effects have been proposed (Altmann & Davidson, 2001; Lovett, 2001, 2005). The WACT model (Altmann & Davidson, 2001) seems similar to RACE at first sight, since it combines ACT-R with insights from Roelofs' work. WACT describes retrieval in a Stroop task as ballistic, but with a retry-mechanism that checks if retrieval was correct; If not, retrieval is retried. Thus, WACT accounts for inhibitory effects by multiple retrievals caused by retrieval failures. As such, WACT is a perfect example of a stroboscopic account of cognition. Retrieval latency for one trial can be the latency associated to one retrieval attempt, or two retrieval attempts, or many, but nothing in between. Therefore, the distribution of reaction times predicted by WACT is clustered around the time it takes for one or multiple retrieval attempts. This does not correspond with the general assumption that subjects' reaction times are normally distributed around the average reaction time.

Another model of Stroop effects NJAMOS (Lovett, 2001, 2005) suggests that the Stroop effects are due to utility differences in the production rules for word and picture recognition. If both a word and a picture are presented, it is more likely that the word recognition production rule will be selected. For small positive SOAs (e.g., +100ms), however, the effect predicted by NJAMOS is smaller than that observed in the data (Lovett, 2001). We suggest that not fully processed words at small positive SOAs might explain this difference between NJAMOS and data. Perhaps a combination of the utility-based explanation Lovett proposes combined with RACE will produce a better fit to the data.

The picture-word interference experiment shows that the RACE model can be a useful extension of the ACT-R architecture of cognition. However, one crucial feature of RACE is not supported by the ACT-R architecture. In RACE, all chunks in declarative memory spread activation to chunks of the same chunk type (lemma chunks to lemma chunks etc.). ACT-R assumes that only chunks that are presently in the buffers can spread their activation (Anderson et al., 2004).[3] Global spreading-activation was abolished because it appeared that no second-order priming effect exists, indicating that spreading of activation through declarative memory was not necessary (Anderson, 1990). However, new evidence suggest a second-order priming effect, although very weak, that cannot be explained by assuming only first-order associations between prime and stimulus (Becker, Moscovitch, Behrmann, & Joordens, 1997; Joordens & Besner, 1992). Therefore, we deem this deviation of ACT-R theory not unreasonable.

Experiments using subliminal primes indicate that priming may also occur when a prime is not fully processed (Abrams, Klinger, & Greenwald, 2002; Draine & Greenwald, 1998), which hints that priming already occurs before chunks in the buffers are fully identified. A dynamical activation mechanism such as RACE may provide accurate modeling accounts for this observation. In RACE, activation of chunks – either in the buffers or in declarative memory – always affects the activation of other chunks, even before the context threshold is reached and a chunk might be retrieved.

In the near future, we intend to test the RACE model on other well-described experiments that are aimed at retrieval from semantic memory.

The first step will be to try to model Glaser and Düngelhoff's Experiment 2 (1984), that describes the counterintuitive result from a word or picture categorization task. If the task is not to name the pictures

[3] In ACT-R 6.0, chunks in all buffers can spread activation, as opposed to ACT-R 5.0, in which only chunks in the goal buffer could be a source of activation.

or read the words, but instead a superordinate word is to be named, the observed asymmetry is flipped: Semantic facilitation can be seen in the category-congruent condition, in both word and picture categorization. We believe that this can be attributed to an increase of the activation of the superordinate concept, which is the same for target and distractor, because they belong to the same category (La Heij, Kuipers, & Costa, in press). Only for the incongruent distractor condition in the word categorization task interference can be observed. This might be explained by inhibition from the incongruent category concept.

References

Abrams, R. L., Klinger, M. R., & Greenwald, A. G. (2002). Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin & Review, 9*(1), 100-106.

Altmann, E. M., & Davidson, D. J. (2001). An integrative approach to Stroop: Combining a language model and a unified cognitive theory, *Proceedings of the 23rd annual meeting of the Cognitive Science Society*: Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1059-1082.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience, 21*, 47-74.

Collins, A. M., & Loftus, E. F. (1975). Spreading activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*(3), 286-303.

Dyer, F. N. (1973). Stroop phenomenon and its use in study of perceptual, cognitive, and response processes. *Memory & Cognition, 1*(2), 106-120.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 640-654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture-processing. *Journal of Experimental Psychology-General, 118*(1), 13-42.

Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word - implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 483-491.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. C. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282*(5386), 108-111.

La Heij, W., Happel, B., & Mulder, M. (1990). Components of Stroop-like interference in word reading. *Acta Psychologica, 73*(2), 115-129.

La Heij, W., Kuipers, J. R., & Costa, A. (in press). A further look at semantic context effects in lagnuage production: The role of response congruency. *Language and Cognitive Processes*.

Lovett, M. C. (2001). Selective attention and strategies: Not just another model of Stroop, *Proceedings of the Fourth International Conference on Cognitive Modeling*: Mahwah, NJ: Lawrence Erlbaum.

Lovett, M. C. (2005). A strategy-based interpretation of Stroop. *Cognitive Science, 29*(3), 493-524.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect - an integrative review. *Psychological Bulletin, 109*(2), 163-203.

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.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*(2), 333-367.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1-3), 107-142.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Taatgen, N. (2005). Modeling parallelization and flexibility improvements in skill acquisition: From dual tasks to complex dynamic skills. *Cognitive Science, 29*(3), 421-455.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550-592.