Accounting for subliminal priming in ACT-R

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

Abstract

This paper presents a cognitive model of a subliminal priming task, using Retrieval by ACcumulating Evidence (RACE), a model of declarative memory retrieval. RACE is implemented as an extension to the ACT-R architecture of cognition. First, we will discuss the exact implementation of RACE within the constraints imposed by ACT-R. Second, we will discuss the subliminal priming task that we modeled and present a cognitive model of this task that incorporates RACE.

Introduction

Successful behavior depends for a large part on having declarative knowledge available at the right time. Humans are therefore continuously retrieving declarative facts from long-term memory storage, based on their continuously updated perception of the environment. The continuous character of perception is reflected in the memory retrieval process, as can for instance be observed in the retrieval latencies of psychonomic experiments in which stimuli are asynchronously presented (e.g., picture-word interference, Glaser & Düngelhoff, 1984) or in experiments in which the presentation durations of stimuli are manipulated (e.g., subliminal priming, Marcel, 1983). A cognitive model of declarative memory retrieval should also reflect the continuous character of the input on which memory retrievals are based. However, current cognitive architectures such as ACT-R (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004) or Soar (Newell, 1990) cannot satisfactorily account for this (Van Maanen & Van Rijn, 2006; in press). Retrieval by ACcumulating Evidence (RACE) is a model that does describe the *process* of retrieving one or more chunks of information from memory. In RACE, memory retrieval is not considered ballistic, but is rather thought of as a process in which the likelihood that a piece of information will be needed for successful behavior is continuously estimated. Therefore, the likelihood estimate can be continuously adapted to the changing environment.

RACE can be perceived as an interaction of ideas from cognitive architectures that rely on symbol manipulation (Anderson et al., 2004; Newell, 1990) and ideas from sequential sampling models (Ratcliff & Smith, 2004; Usher & McClelland, 2001). The architectural nature is clear from the cognitive constraints imposed on RACE. In the current implementation of the theory, we constrained RACE by adopting the rational approach that is intrinsic to the ACT-R cognitive architecture (Anderson et al., 2004). However, the *subsymbolic* computations that drive declarative memory retrieval are rooted in sequential sampling.

This paper will describe how RACE is implemented in the ACT-R architecture of cognition and will present a RACE model of a subliminal priming task. We will discuss which features of RACE naturally align with ACT-R, and which features of RACE seem to contrast with ACT-R. We chose to implement RACE as an extension to ACT-R because of ACT-R's widespread use in the cognitive modeling world (see for instance the web site of the ACT-R community: http://actr.psy.cmu.edu). More importantly however, adopting an existing general approach towards cognition will reduce the proliferation of different cognitive theories (Newell, 1990), and will constrain theorizing about RACE. A third reason for choosing ACT-R as a modeling framework is that the way ACT-R defines retrieval latency has difficulties with modeling semantic interference (Van Maanen & Van Rijn, 2006; in press). Extending ACT-R with RACE might solve this issue.

ACT-R

A prominent theory that explains behavior at the symbol manipulation level is the ACT-R architecture of cognition (Anderson et al., 2004). Because RACE is implemented as an extension to ACT-R, we will give a very short overview of the architecture, concentrating on these aspects of the theory that relate to declarative memory retrieval.

ACT-R is a cognitive theory in which production rules operate on declarative memory and the environment. Production rules are conditions-actions pairs whose actions are executed if their conditions are met. To determine which production rule's actions will be executed, ACT-R contains a set of *buffers* of which the content is matched against the conditions of each production rule. If multiple production rules are applicable – meaning that, given the buffer contents, multiple sets of actions may be performed – the production rule with the highest utility will be selected, a process called *conflict resolution*. By default, the buffers represent the current goal of the system, the current perceptual state, and a declarative fact that is currently in the focus of attention, that is, that is recently retrieved from long-term memory. Other buffers may be defined if necessary for the task at hand (as has for instance been done for prospective time interval estimation, Taatgen, Van Rijn, & Anderson, in press). The content of a buffer is a chunk: a symbolic unit that represents a simple fact, such as *The capital of Canada is Ottawa*, or *The object I am attending is green and spherical*. Both these example chunks are declarative facts, but the first example can typically be found in the *retrieval buffer*, and represents a fact that has been retrieved from long-term memory, whereas the second example represents a visually observable fact of the world, and might be present in the *visual buffer*. In the context of this paper, we are primarily interested in the way ACT-R incorporates retrieval of chunks from long-term memory, although we not necessarily want to constrain RACE to declarative memory retrieval.

All chunks have an activation level that represents the likelihood that a chunk will be needed in the near future. The likelihood is in part determined by a component describing the history of usage of a chunk called the *base-level activation* (B_i in Equation 1).

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

In this equation, t_j represents the time since the jth presentation of a memory chunk and d is the parameter that controls decay, which in most ACT-R models is fixed at 0.5 (Anderson et al., 2004). The idea is that the activation of a chunk decays over time unless attention is shifted to that chunk and its activation is increased. This way, the base-level activation can be used to model both forgetting and learning effects (Anderson & Schooler, 1991).

The total activation is the sum of the base-level activation and another component describing the influence of the current context (*spreading activation*, Equation 2). The spreading activation component is the sum of strengths of association from chunks j to chunk i, weighed by W_{kj}, representing the importance of various buffers (k) and of associated chunks (j).

$$A_i = B_i + \sum_k \sum_j W_{kj} S_{ji} \tag{2}$$

A more detailed description of the ACT-R cognitive architecture is provided in (Anderson & Lebiere, 1998; Anderson et al., 2004).

RACE model of memory retrieval

RACE is a proposal for a new retrieval mechanism in ACT-R. In RACE, retrieval of a chunk is thought of as a process in which the likelihood that a chunk will be needed given the current context is continuously estimated. This is different from ACT-R, were the context can influence the retrieval of a chunk only at the onset of a particular retrieval request. Note that the continuous aspects of ACT-R's base-level learning equation (Equation 1) are retained in RACE. The continuous updating of context-based activation is similar to the account presented in the leaky competitive accumulator model described by Usher and McClelland (2001).

Also similar to ACT-R, the accumulation process in RACE is influenced by various sources of evidence. Increases in activation may be caused by the current context, which may be formed by the current buffer contents, or other chunks that are currently active. Via a spreading activation mechanism these chunks provide evidence for the likelihood that other chunks will be needed. That is, they increase the activation of these chunks.

Another source of evidence for the likelihood that a chunk will be needed is the history of usage of that chunk. Frequently or recently used chunks are more likely to be used again in the near future. In RACE, this is reflected by the starting point of the accumulation process. The level of activation at which accumulation starts is determined by the base-level activation of ACT-R, which reflects the frequency and recency of the usage of a chunk (Anderson & Schooler, 1991).

To preserve the temporal nature of the evidence for a chunk, the accumulated RACE activation is subject to continuous decay. Activation of a chunk thus decreases if not enough evidence for that chunk is present. Since the context may change over time, the accumulation process is not determined when a retrieval process is initiated (the retrieval onset), but may also change. Therefore, incoming information or the removal of information from the buffers may influence which chunk will be retrieved.

Activation values represent the *relative* likelihood that a chunk may be needed (Anderson & Lebiere, 1998), which means that the level of activation at which a chunk has been retrieved should also be defined *relative* to the activation of other chunks. Therefore, RACE uses a *retrieval ratio* that determines how much the activation of a particular chunk must stand out against the total activation of all competing chunks. This is analogous to the relative stopping rule described by Ratcliff and Smith (2004; cf., ACT-R's former competitive latency mechanism, discussed in Van Rijn & Anderson, 2003). If multiple chunks match the criteria of the retrieval request, the chunk that reaches the retrieval ratio first will be retrieved. In these cases, the eligible chunks compete for retrieval. If the activation levels of multiple chunks increase, the total activation of the system also increases, making it more difficult for a chunk to reach the retrieval ratio. This feature of RACE will prove to be important in explaining differences in retrieval latency, for example in the model of subliminal priming explained later in this paper.

So far, we described the general idea of the RACE model of memory retrieval. In this section, the exact implementation of RACE will be presented and how RACE relates to the ACT-R architecture.

The accumulated activation component of RACE is described by the following equation:

$$C_i(t+\Delta t) = d^{\text{acc}} C_i(t) + \beta \sum_{j \in k} C_j(t) S_{ji} \tag{3}$$

This equation reflects the idea that the accumulated activation of a chunk at a certain moment in time ($C_i(t+\Delta t)$) is determined by the level of accumulated activation one time step ago ($C_i(t)$), summed with spreading activation from other chunks; that is, the accumulated activation of other chunks ($C_j(t)$) weighed by strengths of association between these chunks and the chunk i (S_{ji}). At retrieval onset, accumulation starts with the history-based evidence, which is the current base-level activation. Thus

$$C_i(\text{retrieval onset}) = B_i(\text{retrieval onset}) \quad (4)$$

Accumulated activation decays away, the speed of which is controlled by the parameter d^{acc}. A smaller value of d^{acc} results in faster decay. The parameter β in Equation 3 controls the amount of influence of the context. Although in ACT-R activation can have a negative value, we have chosen in our current implementation to ignore the spreading activation from very small – that is, negative – activation values for reasons of computational efficiency.

By continuously updating spreading activation towards a chunk, the chunk may reach a level of activation at which retrieval can take place. The time at which retrieval takes place is the first moment after the start of accumulation at which the following inequality holds:

$$\frac{e^{A_i}}{\sum_j e^{A_j}} \geq \theta \quad (5)$$

This means that for a chunk to be retrieved (i in Inequality 5) the activation should be high with respect to all competing chunks (j). Because ACT-R activation values represent the *relative* likelihood that a chunk will be needed, an exponential scaling is applied to eliminate effects from possible negative values, as is common in ACT-R equations.

Perhaps a clarification is needed on the notions base-level activation (B_i, defined in Equations 1 and 4) and accumulated activation (C_i in Equation 3). To incorporate frequency and recency effects in the retrieval process, the accumulation of activation starts at the current level of base-level activation (Equation 4). During a retrieval process however, activation is estimated according to Equation 3. At retrieval, the base-level activation of the retrieved chunk is also increased to account for the recent encounter with the retrieved chunk, because at the next retrieval attempt the base-level activation is again used as the starting value of the accumulation process.

The question arises which of the two activations (B_i or C_i) is a better predictor of the likelihood that a chunk will be needed. We believe that at very short time intervals – such as the SOAs from the subliminal priming experiment discussed below – accumulated activation better aligns with the empirical data. However, at longer time intervals, base-level activation has been shown to give good predictions (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Anderson & Schooler, 1991). Because in the subliminal priming task and model described below prime and target are retrieved in a very small time window, focusing on accumulated activation only will suffice to model the priming effects. Therefore, for this model the base-level activation values were kept constant over all chunks.

Subliminal priming

In this section, we will discuss the task we modeled using RACE: a subliminal priming study by Marcel (1983). Also, we will discuss why this particular task is interesting given the specific nature of RACE. In subliminal priming tasks, primes are presented that are not consciously perceived by the participant. Usually, primes are presented for a very short period and are followed by a visual mask, so that participants can not discriminate between the presence and absence of a prime (Marcel, 1983; Merikle, Smilek, & Eastwood, 2001). Marcel (1983) showed that under these circumstances priming effects persisted. His Experiment 3 describes a Stroop-task in which words are presented as primes, and color patches are presented as cues. Participants had to respond to the color patches by pressing a button associated to one of the colors. He found the same kind of interference and facilitation as usual in the Stroop paradigm, but a smaller effect for the subliminal primes than for the consciously perceived primes (Figure 3 presents the latencies that Marcel observed). Marcel concluded that subliminal primes have an effect on latency, even though participants are not aware of their presence.

Four prime conditions were tested by Marcel (1983, Experiment 3): Color congruent, color incongruent, neutral, and no-word. In the congruent condition, the prime was the name of the target color, whereas in the incongruent condition the prime was the name of another color. In the neutral condition, the prime was a non-color word that was also unrelated to colors. The no-word condition presented the mask only. Thus, no prime was presented. The condition in which the prime was subliminal was called the unaware condition. In the aware condition, by contrast, the presentation duration was 400ms. Both prime and cue were presented at the same time.[1]

From a symbolic perspective, stimuli have to be considered as symbols in order to engage in cognitive processing. In ACT-R, this means that a stimulus has to be present in a buffer. However, stimuli that are presented for such short durations as common in subliminal priming paradigms do not reach the visual buffer. ACT-R assumes an attention shift to the stimulus before an object can be encoded as a symbolic chunk, which takes a certain amount of time, estimated at 185ms (Anderson, Matessa, & Lebiere, 1998). This exceeds the presentation duration

[1] In the original experiment, Marcel included also another condition with a Stimulus Onset Asynchrony between prime and cue. This condition is similar to the picture-word interference study by Glaser and Düngelhoff (1984), which has previously been discussed by Van Maanen and Van Rijn (2006; in press). Therefore, it is not included here.

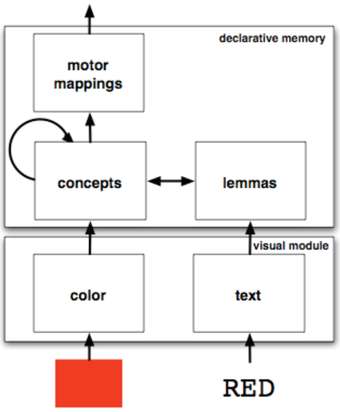

Figure 1: The flow of activation in the congruent condition of the subliminal priming model.

of the prime in the unaware conditions (which is 80ms at maximum, Marcel, 1983). In ACT-R models, stimuli that are presented for less than the time it takes to shift attention can therefore not influence central cognition. The way ACT-R deals with stimulus durations is all or none. Either the stimulus has been presented not long enough, and the stimulus is not perceived at all, or it is fully is perceived. Consequently, symbolic theories of cognition cannot account for subliminal priming data. The next section will show how RACE deals with the short presentation durations typical in subliminal priming tasks.

Subliminal priming model

The subliminal priming model comprises three chunk types, as outlined in Figure 1: Lemmas, concepts, and motor mappings. The concept chunks can be regarded as representations of semantic properties. Chunks of the lemma type can be regarded as sets of orthographic and syntactic properties of a word. The motor mapping chunks represent the information which button to press for which color.

Now, for example in a no-word condition, the cue (being a color patch) spreads activation to its associated concept, which spreads activation to the associated motor mapping resulting in a button press. A similar *flow of activation* will occur in the other conditions, albeit that because of the presentation of a prime word, lemma chunks will also be activated. The activation of multiple motor mapping chunks causes competition in RACE, because the retrieval ratio is harder to reach with multiple accumulating chunks.

Before the experiment, Marcel determined for each participant the critical presentation duration for which participants could not discriminate between presence and absence of a prime (see Marcel, 1983 for details of the procedure). The presentation durations he found ranged from 30 to 80 ms. We used the presentation duration as an extra parameter in fitting the model to the data, with the constraints that its value should be in the range that Marcel found and that the activation of the prime chunk would not exceed the retrieval ratio (Inequality 5). Because the primes in the original experiment were visually masked, we assume that the presentation duration is equal to the time that the prime is available to the visual system.

Table 1: Estimated parameter values for the subliminal priming model.

Parameter	Value
A_{color}	1.8
A_{text}	1.5
$\boldsymbol{\beta}$	**.255**
$\boldsymbol{d^{acc}}$	**.72**
$\boldsymbol{\Delta t}$	**5 ms**
$\boldsymbol{\theta}$	**.81**
aware presentation duration	400 ms
unaware presentation duration	70 ms

Table 1 presents all relevant parameters for the subliminal priming model. The presentation duration of primes in the aware condition is 400ms, as in the original experiment. The unaware presentation duration was estimated at 70ms, serving as the model's critical presentation duration. This duration depends on the RACE parameters presented in bold-face in Table 1. These parameters were not estimated for this experiment, but rather copied from a RACE model of picture-word interference (an updated version of Van Maanen & Van Rijn, 2006; in press). Hence, the only parameters presented here that were estimated for this model were the activation of the words (A_{text}) and of the color (A_{color}). The association values (S_{ji}) between chunks are presented in Figure 2.

Results

The results of the subliminal priming model are presented in Figure 3. We present here differences in latency relative to the no-word condition as this model only captures the memory retrieval process, which comprises the time course from the start of retrieval of a chunk until the retrieval of the motor mapping chunk. The model

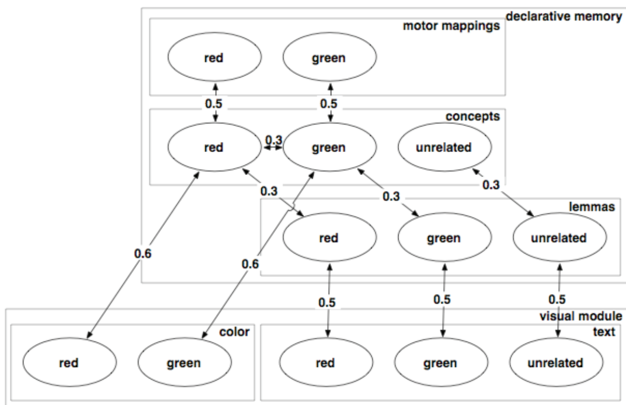

Figure 2: Associative values between different chunks in the subliminal priming model.

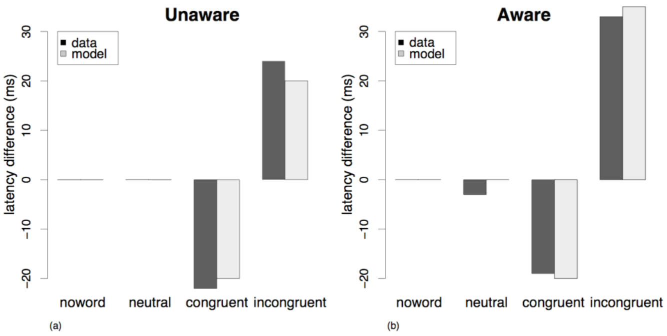

Figure 3: Comparison of the latencies found by Marcel (1983) (a) and the latencies predicted by the subliminal priming model (b). Shown here are the latency differences relative to the no-word condition.

captures quite nicely the effects observed in the data[2] by Marcel (1983) ($r^2 = 0.987$).

In the unaware conditions, only the target chunks reached the retrieval ratio, and no other chunks. Therefore, these chunks are the only ones that are consciously perceived by the model. The model thus remained unaware of all other chunks, as is required for these conditions.

An explication of how the activation flows through the model will be insightful. We split this up in four sections, each describing one condition.

Neutral

In the neutral condition, there is no competition between motor mappings, because there is no button associated with the neutral word. The activation cascades through the network similarly to the no-word condition, because no association exists between the neutral word and the target color, both at the lemma level and at the concept level. Therefore, the activation of the motor mapping associated with the target color increases similarly to the no-word condition because the activation of *all* the motor mapping chunks increases as in the no-word condition.

As an example, Figure 4 gives the activation accumulation in the neutral unaware condition. The activation of the neutral word lemma increases, but, due to the short presentation duration, it does not reach the retrieval ratio. This indicates that the neutral word does not reach awareness.

No-word

The no-word condition is similar to the Neutral condition, because no distractor stimulus is present, resulting in the same behavior of the model as in the Neutral condition. Because in the no-word condition, there is no distractor, there is no difference between aware and unaware.

[2] As the variance in the original data cannot be deduced from the published results, a sensible formal comparison is not possible.

Congruent

Both target and distractor stimuli activate the same concept: the color chunk directly, the text chunk mediated via the lemma chunk. Spreading activation towards the associated motor mapping chunk is therefore higher than in the Neutral and No-word conditions, resulting in faster retrieval.

Incongruent

Because both target stimulus and distractor activate a motor mapping, competition for retrieval takes place at the motor-mapping level. Higher activation for competing chunks means that it is more difficult to cross the retrieval ratio, leading to longer retrieval latencies. The effect is strongest in the aware condition, representing the longer presentation duration of the prime, and thus the longer accumulation of activation of prime-related chunks.

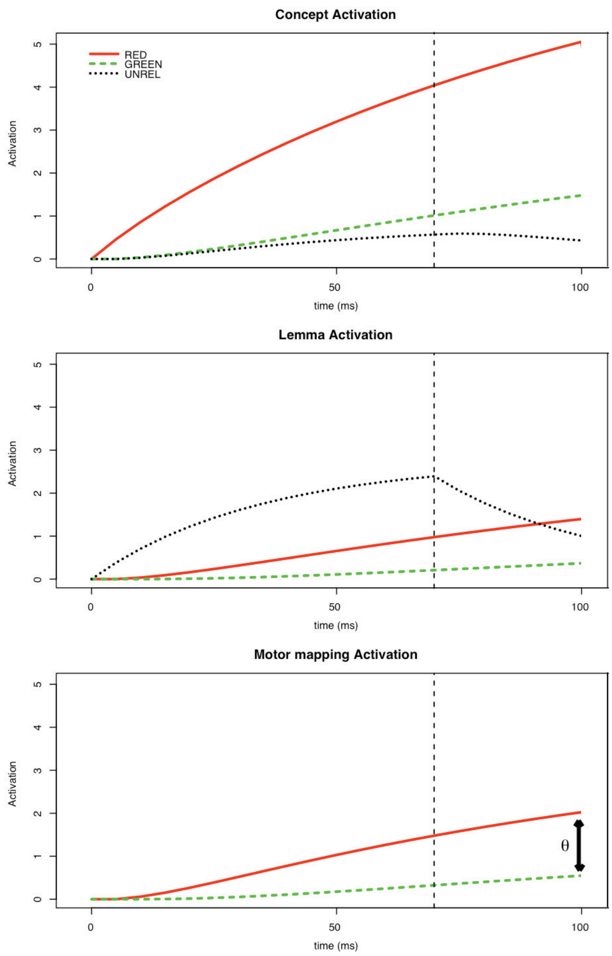

Figure 4: Activation accumulation in the Neutral Unaware condition. The stimuli are a red color patch and an unrelated prime word. The vertical dotted line indicates when the presentation of the prime ends. At this point, the activation of the related lemma has not crossed the retrieval ratio. After 100ms the red motor mapping chunk does cross the retrieval ratio.

Discussion

A difficult question when modeling cognitive tasks that deal with awareness is how awareness is defined within the model. We chose to set a strict boundary for awareness, the retrieval ratio. When a chunk reaches the retrieval ratio, it becomes available inside the buffers. We assume that people are aware of chunks that are currently in the buffers (Taatgen, submitted), and not aware of chunks that have not yet reached the activation needed to enter the buffers.

RACE involves a direct connection between information in the external world (that is, the visual module) and the activation values of declarative chunks in declarative memory. In this respect, RACE deviates from ACT-R, in which all visual information must be mediated by the visual buffer. However, since the visual buffer is associated with awareness as chunks appearing in the visual buffer enter the declarative system, another pathway must be present to account for the subliminal priming data modeled in this paper. We hypothesize that the connections in RACE from the visual module to the declarative memory module may represent part of the ventral visual pathway, that is known to involve connections from striate cortex (associated with ACT-R's visual module) to temporal brain regions (associated with the declarative memory module, Anderson et al., 2004).

The model of subliminal priming discussed in this paper demonstrates that RACE can account for the retrieval latencies observed by Marcel (1983, Experiment 3). By using standard RACE parameter values, the fit of our model to the data set of Marcel was quite good. In combination with previous models of declarative memory retrieval that use RACE (Borst & Van Rijn, 2006; Van Maanen & Van Rijn, 2006; in press), this suggests that RACE might be regarded as a general model of declarative memory retrieval. The added value of the RACE model is that it gives a rational account of how the process of declarative memory retrieval develops. Even the effects on declarative memory retrievals of changes in the world that last only milliseconds can now be taken into account.

References

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.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*(4), 341-380.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. R., Matessa, M., & Lebiere, C. (1998). The visual interface. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*(4), 703-719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396-408.

Borst, J. P., & Van Rijn, H. (2006). Memory decay problems in a level-repetition switch task model. In D. Fum, F. Del Missier & A. Stocco (Eds.), *Proceedings of the Seventh International Conference on Cognitive Modeling* (pp. 351-352). Trieste, ITA.

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.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*(2), 197-237.

Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*(1-2), 115-134.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard UP.

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

Taatgen, N. (submitted). Consciousness in the ACT-R architecture. To appear in *Oxford companion to consciousness*: Oxford UP.

Taatgen, N., Van Rijn, H., & Anderson, J. R. (in press). An integrated theory of prospective time interval estimation: The role of cognition, attention and learning. *Psychological Review*.

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

Van Maanen, L., & Van Rijn, H. (2006). An accumulator model account of semantic interference in memory retrieval. In D. Fum, F. Del Missier & A. Stocco (Eds.), *Proceedings of the Seventh International Conference on Cognitive Modeling* (pp. 322-327). Trieste, ITA.

Van Maanen, L., & Van Rijn, H. (in press). An accumulator model of semantic interference. *Cognitive Systems Research*.

Van Rijn, H., & Anderson, J. R. (2003). Modeling lexical decision as ordinary retrieval. In F. Detje, D. Dörner & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 207-212). Bamberg. GER.