RACE for retrieval: Competitive effects in memory retrieval

Leendert van Maanen & Hedderik van Rijn
Artificial Intelligence, Groningen University

When a question is stated such as "What is the capital of Australia?", various answers start competing for retrieval from declarative memory. If a hint is given during this retrieval process ("The name of the capital begins with the letter C."), retrieval may be facilitated, but if a distractor is presented ("Amsterdam"), retrieval may be inhibited. A well-known example of these kinds of effects is Picture-Word Interference, a task similar to the Stroop-task (Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Schriefers *et al.*, 1990). In these tasks, it has been shown that SOAs have differential effects. For example, presenting a drawing that depicts a concept 50 ms after a word-form of that concept has appeared, speeds up processing of that word in comparison with a neutral condition, whereas in other conditions SOAs might have a negative effect.

The ACT-R Latency Equation as it is defined now, $RT_i = Fe^{-A_i}$ (Anderson *et al.*, 2004), cannot account for these phenomena, as it suggests that retrieval latency *only* depends on the state of the buffers and declarative memory at the exact time of retrieval *onset*, both reflected in A_i. This is easiest demonstrated in the condition where a facilitating word is presented at a short SOA after the picture is shown. After a retrieval request, the chunks that match the request are identified and the one with the highest activation is selected. The Latency Equation determines how long it will take to complete that retrieval, and takes the current level of activation at retrieval onset into account. Thus, at present in ACT-R, the presentation of an interfering stimulus after retrieval onset simply does not influence the calculated latency. Likewise, when another stimulus is presented before retrieval onset, retrieval latency depends (in part) on the spreading activation from the stimulus in a sensory buffer to the to be retrieved chunk: higher levels of activation result in shorter latencies. However, as this can only explain a speed-up, this does not comply with the observation that a condition in which a distractor from the *same category* as the target stimulus is presented has a *larger retrieval latency* than a condition in which an *unrelated distractor* is presented (Glaser & Düngelhoff, 1984). The intuition at least is that concepts of the same category have higher inter-associations than unrelated concepts, which in ACT-R would lead to higher activation levels and shorter latency.

A solution to this issue might be to regard the retrieval process as an instance of a sequential sampling mechanism (Ratcliff & Smith, 2004). Sequential sampling models follow the hypothesis that a neural representation of a stimulus is inherently variable or noisy, and in order to retrieve the required representation, enough samples of the stimulus representation have to be accumulated. Thus, sequential sampling models offer a mechanism that allows for a specification of the time course of retrieval. At retrieval onset, sequential sampling of evidence will allow for the activation of chunks to increase, until at least one chunk's activation has crossed a threshold. Using an adapted version of the leaky competitive accumulator model for perceptual choice (an example of a sequential sampling mechanism, (Usher & McClelland, 2001)), we show that the ACT-R activation function can be extended to account for the *time course of activation* without changing current mechanisms. We will refer to this new set of mechanisms as Retrieval by ACcumulating Evidence (RACE).

In RACE, the time of retrieval is defined as the time at which the activation of a chunk crosses a threshold. We assume that after initiating a retrieval request, the activation of matching chunks is updated per time step. The function underlying this updating has two components. A long-term activation component that is identical to the current ACT-R activation formula (henceforth base-level activation), and a short-term, more volatile activation component that represents the current context (context activation). The total activation is calculated by summing both activation components, and retrieval is finished when this summed quantity reaches a fixed threshold (the context activation threshold $\theta^{context}$). As in the leaky accumulator models, the context activation is based on "evidence ticks". At each time step, if positive context evidence outweighs negative context evidence, the amount of evidence increases. This evidence, gathered during the retrieval phase, is subject to decay. Each piece of evidence can be seen as a contribution to the context activation of that chunk, similar to how the prior occurrences of a chunk contribute to its base-level activation. But different from occurrences of a chunk, pieces of evidence cannot be considered having an infinitely high activation, since the chunk is not (yet) retrieved. Therefore we modeled the volatile context activation using the ACT-R Optimized Equation:

$$C_i(t) = \ln(n/(1-d^{\text{evidence}})) - d^{\text{evidence}} \cdot \ln(T)$$

.

This equation represents the Power Law of Learning (Anderson & Lebiere, 1998). As this equation represents the built-up of evidence, instead of T representing the age of prior occurrences, T is a representation of the age of the sequential sampling process.

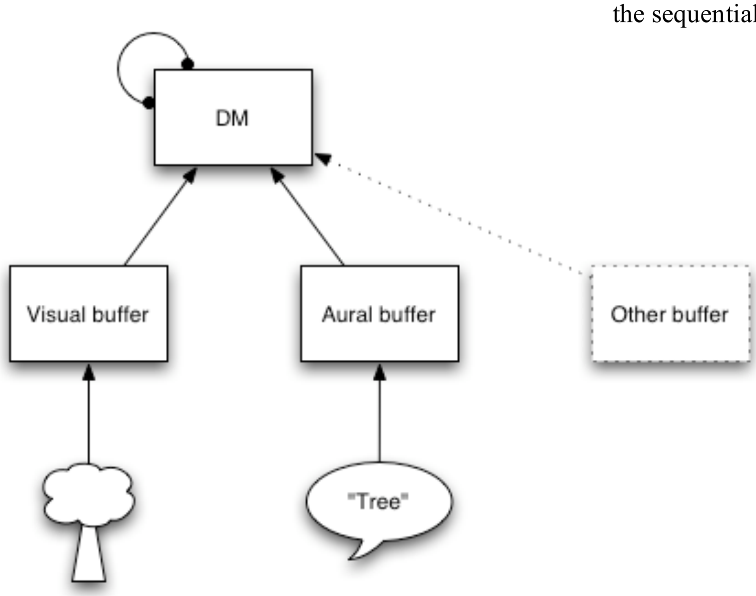


Figure 1. The RACE model. Chunks in the buffers spread positive context activation to chunks in declarative memory. Chunks in declarative spread negative association to each other (only during a retrieval process).

Whether a certain point in time is associated with evidence depends on the amount of context activation coming from the current buffers, context activation (currently only inhibition) coming from other chunks in declarative memory, and base-level activation (Figure 1): If at a certain time step excitatory context activation minus inhibitory context activation crosses a threshold, that time step is associated with positive evidence, and context activation increases. A lack of positive evidence can be considered inhibition as the context activation decays strongly. The name RACE reflects how different chunks compete on this basis of their accumulated evidence for retrieval.

Using the RACE approach, the ACT-R latency equation can be rewritten as:

$$L_i = t \ni \left[(B_i(t) + C_i(t)) \geq \theta^{\text{context}}\right]$$

Note that $\sum WS$, the term reflecting context activation in the default ACT-R equations, is discounted for in the C_i component, as the slots of the goal buffer spread context activation to the associated chunks. In this approach, the source of positive activation is still the chunk available in the buffers, and filled slots in the sensory buffers still inhibit retrievals as in the fan experiments.

We conducted two experiments with RACE. In experiment 1 we compared the predicted latencies of RACE with the predicted latencies from the ACT-R Latency Equation. In experiment 2 we fitted data from a Picture Word Interference experiment.

<u>Experiment 1</u>
Numerous models in ACT-R have shown that default ACT-R provides accurate predictions of retrieval latency (Anderson *et al.*, 2004). To

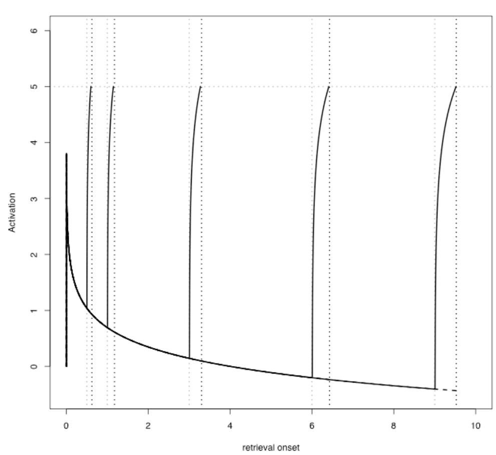
ensure that our approach does not invalidate these results, we show that our model predicts the same latencies as the ACT-R Latency Equation in a non-competitive condition. In this experiment we fitted the ACT-R retrieval latency for different times after the chunk to be retrieved is presented. In other words, we will predict the ACT-R latency for different base-level activation levels.

We first chose reasonable parameters for the ACT-R Latency Equation. These values were not updated in the optimization process, to ensure a fair comparison between the models. The crucial parameters in our model were the context activation threshold $\theta^{context}$, and the evidence threshold $\theta^{evidence}$. These indicate whether a chunk is retrieved ($\theta^{context}$) and whether evidence may be sampled ($\theta^{evidence}$). The evidence threshold was noisy with a standard deviation of σ=0.3. $d^{evidence}$ turned out to be less important in this condition because due to the absence of other chunks no inhibition was present, and evidence was sampled at almost every time step. However, because the age variable in the context activation function is in ms, the decay parameter and the time step frequency are related: If the frequency is high, so is the chance of sampling evidence (more opportunities), and decay may be higher. All parameters are presented in Table 1.

Table 1. Parameters experiment 1

RACE Parameters:	
$d^{evidence}$	0.1
$\theta^{context}$	5
time step frequency (f)	1000 Hz
$\theta^{evidence}$	1.7
σ	1
ACT-R Parameters:	
$d^{base\text{-}level}$	0.5
F	0.35

For a fixed set of retrieval onsets we calculated both the latency predicted by ACT-R and the prediction of our model. The retrieval onsets were chosen to ensure that different base-level activation levels were tested (0.5, 1.0, 3.0, 6.0, 9.0, 12.0, 15.0 seconds after chunk presentation). The experiment was performed 30 times, and the results were averaged. The results are shown in Figure 2.



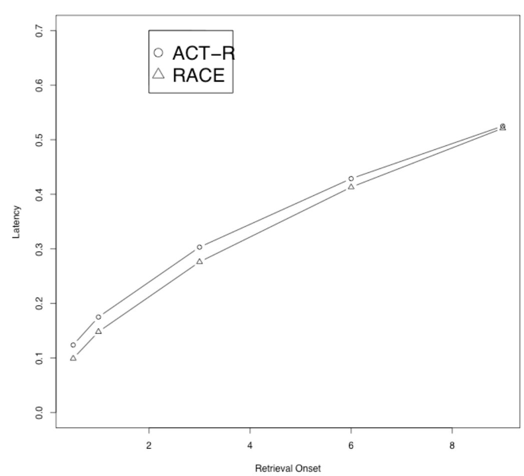


Figure 2. The retrieval process at different time steps (left) and associated latencies (right). Left: the grey dotted lines depict retrieval onsets, the black dotted lines depict predicted ACT-R latency, and the black solid lines depict activation. The simulation was terminated after reaching the retrieval threshold.

This simulation shows that in a single chunk retrieval task where similarity-based interference does not play a prominent role, the model predicts similar latencies to the classical ACT-R retrieval latency function. Therefore, we assume that this extension does not limit ACT-R in its ability to capture the already modeled phenomena.

Experiment 2

Picture Word Interference experiments are characterized by a double stimulus paradigm, in which a distractor stimulus is presented at different SOAs from the target stimulus, and subjects are asked to name the target stimulus (Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Schriefers *et al.*, 1990). Distractors will interfere with the naming process, unless it is indicative of the same concept as the target stimulus, in which case its presentation will facilitate naming (see Figure 3, left, adapted from (Glaser & Düngelhoff, 1984)). We will refer to this condition as *concept congruent*. The conditions in which distractors interfere with the naming process will be referred to as *category congruent*, indicating that distractor and target are concepts from the same category, ensuring a high association, and *incongruent*, indicating that distractor and target are unrelated, ensuring that no association exists.

The addition of a short-term component makes it possible to predict priming effects at short SOAs. As soon as a secondary stimulus is present in one of the buffers, this stimulus will start influencing the primary stimulus and thus cause the interference patterns typically observed in Picture Word-like experiments. When the evidence threshold of a buffer chunk is crossed, evidence is sampled and context activation increases. This increase leads to a higher probability that evidence of positively associated chunks will be sampled. A higher context activation level of a chunk in declarative memory decreases in turn the likelihood that the evidence thresholds of competing chunks are crossed. Less evidence leads to an increased retrieval latency. This process accounts for the latencies observed in the Picture-Word paradigm. Moreover, a highly associated chunk will decrease the likelihood that the evidence thresholds of competing chunks are crossed further than an unassociated chunk, with as a result a higher retrieval latency. To keep the model simple and focus on the behavior of the concept component of the activation, base-level activation was set at a fixed constant of -2.5. The evidence threshold was set high enough that without external stimulation, evidence sampling was unlikely, and decay was set high enough that in the event of spontaneous evidence, context activation would quickly decay back to base-level.

Table 2. Parameters experiment 2

RACE Parameters:	
$d^{evidence}$	0.5
$\theta^{context}$	0
f	1000 Hz
$\theta^{evidence}$	1.7
σ	1
B_i	-2.5

We ran our model through a series of retrieval experiments. The model was set to retrieve concepts from memory that were highly associated with the presented items. At different SOAs (-400ms, -200ms, 0ms, 200ms, 400ms) distractor items were presented, and the latency was recorded. The results are presented in figure 3 (right).

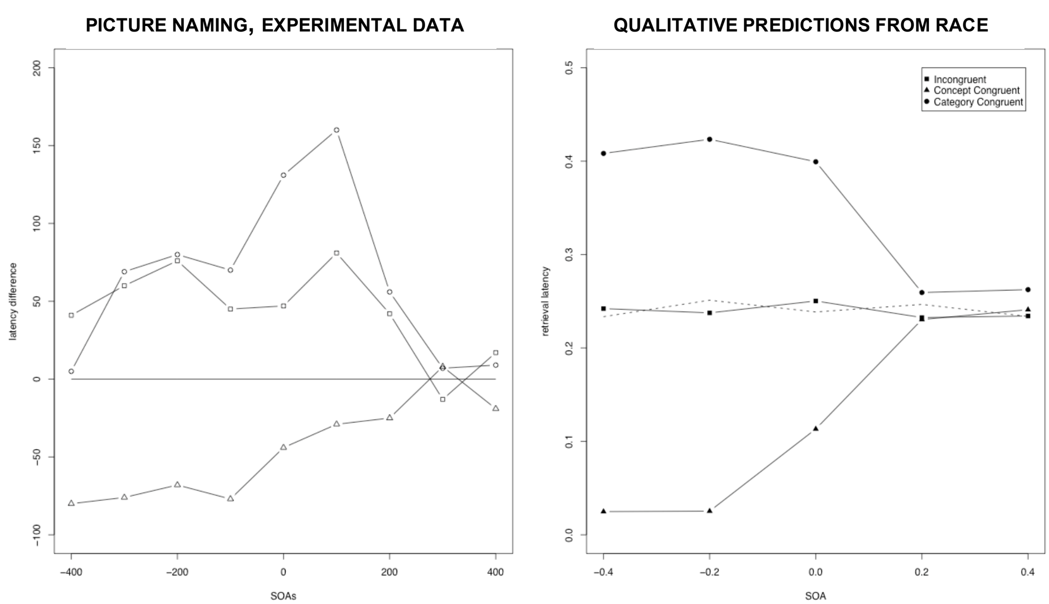


Figure 3. The left panel shows data from a Picture Naming Experiment by Glaser & Düngelhoff (1984). Distractors were visually presented. The right panel shows results from the RACE model.

Figure 3 (right) shows that both the facilitating and interfering effects are strongest at negative SOAs. In those trials the distractor has more time to accumulate evidence, thereby influencing the target stronger when it is presented. At 0 or positive SOAs, the influence of the distractor is less as, on retrieval onset of the target, it does not have any already gathered evidence. For the concept congruent (facilitating) condition, a similar pattern is observed in the data, but for the interfering condition the data shows a peak at SOA=100ms, and smaller values at negative SOAs. The current version of our model does not capture this, but could easily be extended to explain this effect: An explanation is that when a chunk is retrieved, accumulation of evidence stops, and the context activation quickly decays. However, the chunk itself is retrieved, and therefore has a higher base-level activation. If the prior stimulus was category congruent, this higher base-level slightly inhibits the target chunk, but not as much as when the distractor was still being processed. Therefore, the category congruent conditions at long negative SOAs will still show longer latencies than the neutral condition, but not as long as on positive SOAs. If the prior stimulus was concept congruent, the retrieval has increased the base-level of the concept that also needs to be retrieved for the target. Therefore, these conditions will still show significantly shorter latencies than the neutral condition. In our model, however, base-level activation effects are not taken into account, but instead the context activation remains active, prolonging the interfering effect.

Another difference between the two panels of Figure 3 is the effect of the incongruent distractor condition. In the left panel, this condition shows slower reaction times than the neutral condition (which is the condition in which no distractors are present). In our model, the neutral and incongruent distractor conditions have a similar effect as we did not aim at explaining these differences. However, the model can easily be expanded to explain this effect. The expected increased retrieval latency in the incongruent distractor condition might be explained by assuming weak associations between the target and the incongruent distractor. The same effect as in the category congruent condition occurs, but to a lesser extent. Another possibility is that the activation of the incongruent concept is influenced by the number of competing chunks (i.e., the fan-effect). In the neutral condition (dotted line in figure 3), only one chunk is in the buffers. In the other conditions, more chunks are present in the buffers. An extra competitor would mean that the "total amount" of activation is divided by one more, having a negative effect on the activation of the chunks and thus on the latency. A third, equivalent option would be to calculate the *Luce Ratio* (Luce, 1959), indicating that the probability that a chunk will be retrieved depends on its part of the total activation. An extra competitor would mean higher total activation, and would in that way affect the latency of the target chunk, a mechanism which is implemented in the ACT-R competitive latency equation. We are planning to test these hypotheses in a next implementation of the model.

Without much emphasis on optimization, RACE shows the most important phenomena observed in Picture-Word like tasks: a facilitating effect for the concept congruent stimulus, and a convergence of the effects to the neutral condition at positive SOAs. Besides the explanatory power for these previously unexplained effects, RACE is still compatible with current ACT-R latency predictions.

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., Lebiere, C. (1998). *The Atomic Components of Thought.* Mahwah, NJ: Lawrence Erlbaum.

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.

Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.

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

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture-word interference studies. *Journal of Memory and Language, 29*(1), 86-102.

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