Modeling the fan effect during learning with a log-normal race model

Philine Link (p.m.link@uu.nl)[1], **Bruno Nicenboim**[2] **& Jakub Dotlačil**[1]
[1]Institute for Language Sciences, Utrecht University
[2]Department of Computational Cognitive Science, Tilburg University

Abstract

The fan effect is a well-known phenomenon in the study of associative memory. It refers to the finding that as the number of associations linked to a concept increases, retrieval becomes slower and less accurate. Modeling accounts have largely focused on reaction times or accuracy in isolation, and have primarily modeled testing-phase performance. In the present study, we examine whether fan effects already arise during learning and propose a unified account of reaction times and accuracy at the trial level using a Bayesian hierarchical log-normal race model. We report data from a Dutch fan experiment showing that fan effects are already observable during the learning phase and increase across repetitions. The hierarchical Bayesian framework allows us to capture both reaction times and accuracy simultaneously while accounting for individual differences. Our model successfully captures slower incorrect responses, consistent with predictions from activation-based theories such as ACT-R. Our results demonstrate that race models can reproduce key empirical patterns associated with fan effects during memory acquisition.

Keywords: log-normal race model; fan effect; associative memory; spreading activation; reaction times; accuracy

Introduction

Human memory is often characterized as a network of associations, where retrieving a concept activates related concepts through spreading activation (e.g., Collins and Loftus, 1975). A classic demonstration of this is the fan effect (Anderson and Reder, 1999; Anderson, 1974): when a concept is associated with multiple facts (high fan condition), retrieving an individual fact becomes slower and less accurate. This interference is created when participants learn, for example, the sentences "The captain is in the tower" and "The captain is in the pool". When asked to verify either fact, participants are slower compared to cases when they verify facts with just one association.

The fan effect has informed theories of memory retrieval, particularly ACT-R (Anderson et al., 2004), which explains the phenomenon through activation-based competition. In ACT-R, activation spreads between concepts that share retrieval cues. When activation is distributed across many associations (high fan), each individual association receives less activation, resulting in slower and less accurate retrieval. Importantly, ACT-R makes explicit, quantitative predictions about how this competition affects both response times (RTs) and accuracy within one framework.

While the fan effect has been studied extensively at the behavioral level (e.g., Anderson, 1974; Anderson and Reder, 1999; Bunting et al., 2004; Cantor and Engle, 1993), several theoretical questions remain unaddressed. First, most modeling efforts focus on summary statistics rather than trial-level dynamics, not capturing how competition unfolds during individual retrieval attempts. Second, incorrect responses, which ACT-R predicts result from competitors winning the race for retrieval, receive limited attention. Third, existing work has primarily examined the testing phase of fan experiments, leaving unclear whether fan effects already emerge during learning, when associations are still being acquired.

Approaching the fan effect from a computational modeling perspective allows us to address these gaps. Race models, which describe retrieval as a competition between accumulators, provide a framework for simultaneously capturing RTs and response accuracy (Rouder et al., 2015). The log-normal race model is particularly well-suited for studying the fan effect, as it directly implements ACT-R's assumption that activation determines both retrieval speed and success (Nicenboim and Vasishth, 2018). Critically, a hierarchical Bayesian implementation of this model allows us to quantify the uncertainty in our parameter estimates, capture individual differences in how fan effects manifest across participants, and model both RT distributions and accuracy jointly at the trial level.

In the present work, we apply a hierarchical Bayesian log-normal race model to data from the learning phase of a fan experiment, thereby modeling RTs and accuracy in combination on a trial-by-trial basis. We model responses in the learning phase for two reasons: First, in this phase, participants choose between four response options that are presented simultaneously. We can therefore model the decision process as a race where the fastest accumulator determines both the response and the RT, with fan size and response correctness affecting accumulation rates. And second, modeling the learning phase allows us to determine whether interference effects already emerge while associations are still being learned by participants. To our knowledge, this is the first study to fit such a model directly to individual participant data from the learning phase. We show that the model accurately captures fan effects, both on RTs and accuracy, while also accounting for incorrect responses as outcomes of the race for retrieval. Our work demonstrates that an activation-based competition between response options occurs even during the initial learning of the associations.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Person	Location	Example sentences
a	A	*The captain is in the tower*
b		*The salesman is in the tower*
a	B	*The captain is in the pool*
c		*The soldier is in the pool*
...		
e	E	*The painter is in the cave*
f		*The cowboy is in the cave*
e	F	*The painter is in the village*
f		*The cowboy is in the village*
e	G	*The painter is in the barn*
g		*The judge is in the barn*
e	H	*The painter is in the hotel*
g		*The judge is in the hotel*
...		
Re-paired foils:		*The painter is in the tower*
		The painter is in the pool

Table 1: Example of stimuli translated to English. Lowercase letters refer to a person item (e.g., a = *captain*), uppercase letters refer to a location item (e.g., A = *tower*). Here, person fan is varied (a = fan 2, e = fan 4) while location fan remains at fan 2. The foil sentences were created with new pairings of lexical items that were used in the target sentences.

Fan Experiment

In classic fan experiments, participants learn a number of simple facts, e.g., *The hippie is in the park* (Anderson, 1974), in which concepts, such as *hippie* and *park* occur in multiple different facts. This creates the associative fan which denotes the number of different sentences linked to a given concept. It has been shown that an increase in fan size results in an increase in RTs and a decrease in accuracy when participants are asked to retrieve the studied sentences (Anderson and Reder, 1999).

Materials

We conducted a fan experiment in Dutch with sentences that followed the pattern "The [person] is in the [location]". An example of a stimulus set can be seen in Table 1. The person and location words were repeated to create two fan conditions: a fan 2 condition with two sentences and a fan 4 condition with four sentences. In total, participants studied 24 sentences: eight sentences with both concepts set to fan 2 and 16 sentences in which either persons or locations were set to fan 4. Whether the fan 4 condition applied to persons or locations was counterbalanced between participants. Stimulus sets were randomly generated from a list of person and location words for each participant.

Participants

We collected data from 150 (lab, N=50; online through Prolific, N=100) Dutch native speakers residing in the Netherlands. All participants gave informed consent and were compensated for their participation. The study was approved by the local ethics committee.

Procedure

The experiment consisted of three parts: an exposure phase, a learning phase, and a testing phase. In the exposure phase, each stimulus sentence was presented once. Participants proceeded to the next sentence by pressing a key. Then, they completed two iterations of the learning phase in which they had to respond to two questions for each sentence they were exposed to before: one question inquiring about the location ("*Where is the [person]?*") and the other about the person ("*Who is in the [location]?*"). The questions were presented in random order and participants responded by clicking on one of four response options. One of the presented response options was correct. Only if participants responded to all 48 questions correctly could they move on to the second iteration of the learning phase. If a question was responded to incorrectly, it was presented again at the end of the current iteration of the learning phase, otherwise it was dropped. There was no time restriction on responses in the learning phase. Feedback (correct/incorrect) and the correct sentence were presented after each response. In the following testing phase, participants saw a sentence and had to indicate as quickly and accurately as possible whether it was a sentence they had studied before (target) or a new sentence (foil) which was a novel combination of the previously studied person and location words. In total, participants saw 48 sentences per testing iteration (24 targets, 24 foils) and they could take a break after 16 responses.

Methods and results

Here, we present accuracy and RTs recorded in the learning phase of the fan experiment as this is the data that we will model. However, we also observed the fan effect in the testing phase: it manifested as longer RTs and lower accuracy for higher fan conditions.

From the data from the learning phase, we removed RTs faster than 300 ms and slower than 20 seconds. Additionally, we removed RTs outside of 3 SDs from the mean. In total, we excluded 3.3% of the data. Table 2 shows the mean accuracy and mean RTs for the learning phase.

To examine whether the fan effect was visible in the learning phase, we fitted a linear mixed-effects model to the log-transformed RT data with the structure shown in (1). The variables *fan_pers* and *fan_loc* denote the number of associations (2 or 4) for the person and location in each observation, respectively. Response error (correct/incorrect) is indicated by *feedback* and the *learning_rep* refers to whether the response was given in the first or the second iteration of the learning phase. Random intercepts were included for participants (*pp_num*) and individual questions (i.e., specific fact

Person fan	Location fan	Accuracy (SE)	RT correct (SE)	RT incorrect (SE)
2	2	0.47 (0.007) 0.62 (0.008)	4304 (47) 4245 (48)	4753 (46) 4654 (63)
2	4	0.45 (0.007) 0.59 (0.008)	4294 (47) 4374 (49)	4658 (43) 4655 (58)
4	2	0.47 (0.007) 0.60 (0.008)	4594 (49) 4653 (52)	5094 (47) 5071 (65)

Table 2: Mean accuracy and RTs (in ms) with standard errors (SE) in the learning phase. Rows show fan size conditions, the values shown refer to repetition 1 (top) and repetition 2 (bottom) of the learning phase.

probes).

$$\begin{aligned}\log(\text{rt}) \sim\ & \text{fan\_pers} + \text{fan\_loc} + \text{feedback} + \text{learning\_rep} \\ & + (1 \mid \text{pp\_num}) + (1 \mid \text{question})\end{aligned} \tag{1}$$

The model in (1) revealed significant effects of response error ($\beta = 0.123$, SE = 0.007, t = 18.46), location fan ($\beta = 0.022$, SE = 0.005, t = 4.55), person fan ($\beta = 0.021$, SE = 0.005, t = 4.39) and learning repetition ($\beta = 0.017$, SE = 0.007, t = 2.56). Longer RTs are associated with incorrect responses and higher fan sizes for both person and location items.

Additionally, we fitted two separate models with the same structure (see 2) on log-transformed RTs for each repetition of the learning phase using only correct trials to study how fan effects changed across learning. Since the log-normal race model is fit to log-transformed RTs, we used this outcome variable when specifying the following linear mixed-effects models.

$$\begin{aligned}\log(\text{rt}) \sim\ & \text{fan\_pers} + \text{fan\_loc} \\ & + (1 \mid \text{pp\_num}) + (1 \mid \text{question})\end{aligned} \tag{2}$$

In the first iteration of the learning phase, both person fan ($\beta = 0.022$, SE = 0.009, t = 2.53) and location fan ($\beta = 0.018$, SE = 0.009, t = 2.04) significantly predicted log(RT). These effects increased in the second iteration, with person fan ($\beta = 0.031$, SE = 0.009, t = 3.37) and location fan ($\beta = 0.049$, SE = 0.009, t = 5.43) both showing stronger effects.

As previously stated, the fan effect could also be observed in the testing phase, which successfully replicates the fan effect in Dutch. More interestingly, however, we already observed the fan effect in the learning phase, with a stronger effect occurring in the second iteration of learning. At this point in the experiment, participants have limited knowledge of the experimental items and their associations, yet their memory retrieval is already affected by the number of associations of the retrieval target.

Modeling

Since the fan effect is one of the fundamental phenomena supporting ACT-R's theory of declarative memory, we use a computational model, namely the log-normal race model (Rouder et al., 2015), which aligns with ACT-R's central assumptions about how activation influences the time and accuracy of retrieval from memory (Nicenboim and Vasishth, 2018).[1]

The ACT-R activation model as a Log-Normal Race Model

In the ACT-R activation model, it is assumed that targets and foils are stored in declarative memory. During retrieval, the element i with the highest activation is retrieved and its retrieval time T_i is a function of its activation, A_i, treated as a normally distributed random variable (F is a free time scaling parameter).

$$T_i = Fe^{-A_i} \tag{3}$$

As is common in approaches linking ACT-R activation and race models, we can re-conceptualize activation as evidence accumulation. Activation then represents the rate of accumulation: the first chunk that accumulates evidence fastest up to a certain threshold is retrieved (Nicenboim and Vasishth, 2018; Van Maanen et al., 2012). The ACT-R activation model also allows us to capture response accuracy. This is based on the simple assumption of race models that, in case multiple elements compete for retrieval, the element that accumulates evidence fastest up to the threshold is the winner. This also provides us with the formula for the retrieval time of the "winner", T_w, which is simply the retrieval time using the highest activation among competing elements in I, i.e.,

$$T_w = Fe^{-\max_{i \in I} A_i}. \tag{4}$$

Let us go more concretely through what happens in a single trial. In the learning phase of the fan experiment, which we model, participants choose one of four response options on each trial n, such that one option is the target, and the three other options are the foils. We model each trial as a race between four accumulators, one for each response option. For the model to capture the correct response, the target accumulator has to finish faster than the three other accumulators representing foils. The accumulation of evidence is fixed within a trial for every accumulator, but it can vary from trial to trial, as we will see below.

Finally, we rewrite the formula (3) as

$$T_i = Fe^{-A_i} = e^{\log(F) - A_i}. \tag{5}$$

Since A_i is a normally distributed random variable, $\log(T_i)$ is a normally distributed random variable shifted by $\log(F)$, or, in other words, T_i is a log-normally distributed random variable.

$$\begin{aligned}\log(T_i) &\sim \text{Normal}(\log(F) - A_i, \sigma) \\ T_i &\sim \text{LogNormal}(\log(F) - A_i, \sigma)\end{aligned} \tag{6}$$

[1] All code for modeling can be found at: https://github.com/jakdot/fan-effect-learning-cogsci.git

More concretely, since we work with four retrievals per trial n, three of which are foils and one is the target, we can spell out all corresponding "potential" retrieval times $T_{f1,n}, \dots T_{f3,n}$ (for foils) and $T_{t,n}$ (for the target) as shown in (7).

$$\begin{aligned} T_{t,n} &\sim \text{LogNormal}(\log(F) - A_{t,n}, \sigma) \\ T_{f1,n} &\sim \text{LogNormal}(\log(F) - A_{f1,n}, \sigma) \\ T_{f2,n} &\sim \text{LogNormal}(\log(F) - A_{f2,n}, \sigma) \\ T_{f3,n} &\sim \text{LogNormal}(\log(F) - A_{f3,n}, \sigma) \end{aligned} \tag{7}$$

The transformed parameter $\log(F)$ is a scaling parameter. Its chosen value is immaterial since other parameters in the activation can offset it, as long as we make sure that retrieval time is never negative. We set it to the constant $\log(F) = 10$ throughout the whole experiment.

We now turn our attention to activation, i.e., the rate of evidence accumulation. Note that we distinguish evidence accumulation for targets, $A_{t,n}$ and for foils $A_{f1,n} \dots A_{f3,n}$. Finally, following Rouder et al. (2015), we assume that the noise parameter is the same for all retrievals, irrespective of what accumulator is used.

We now discuss in detail how evidence accumulation is specified. Targets and foils are treated separately. For the target, we define $A_{t,n}$ as

$$\begin{aligned} A_{t,n} &= \exp(\alpha + u_{subj[n],1}) - \\ &\quad \exp(\beta_{fan} + u_{subj[n],2}) \cdot fan_{t,n}. \end{aligned} \tag{8}$$

For each of the three foils, the evidence accumulations $A_{f1,n} \dots A_{f3,n}$ are defined as

$$\begin{aligned} A_{f1,n} &= \exp(\alpha + u_{subj[n],1}) - \\ &\quad \exp(\beta_{fan} + u_{subj[n],2}) \cdot fan_{f1,n} - \\ &\quad \exp(\beta_{penalty} + u_{subj[n],3}) \\ A_{f2,n} &= \exp(\alpha + u_{subj[n],1}) - \\ &\quad \exp(\beta_{fan} + u_{subj[n],2}) \cdot fan_{f2,n} - \\ &\quad \exp(\beta_{penalty} + u_{subj[n],3}) \\ A_{f3,n} &= \exp(\alpha + u_{subj[n],1}) - \\ &\quad \exp(\beta_{fan} + u_{subj[n],2}) \cdot fan_{f3,n} - \\ &\quad \exp(\beta_{penalty} + u_{subj[n],3}). \end{aligned} \tag{9}$$

Let us now break this down. First, the model has a fixed effect of $fan_{t,n}$ (for targets) and $fan_{f1,n} \dots fan_{f3,n}$. This simply represents the fan size of each of the four choices per trial without distinguishing between person and location words: the fan size of the target (where one word in the target could vary between the fan size of 2 or 4, and the other word is always of the fan size 2) and the fan size of the foils (where, again one word could be of the fan size 2 or 4 and the other word is always of the fan size 2). For each target/foil combination, its fan is the sum of log of its person fan and log of its location fan, following the standard ACT-R procedure.

Then, there are several parameters:

- The intercept parameter α is the baseline accumulation, and corresponds to base-level activation in ACT-R. It is shared by targets and foils and is adjusted by by-subjects random effects $u_{subj[n],1}$, to model differing baseline rates per subject. The accumulation with the by-subject adjustment is exponentiated to ensure that the baseline accumulation is always positive, as standardly assumed in ACT-R (Anderson and Lebiere, 1998).
- The fixed-effect parameter β_{fan} is a change in accumulation due to increased fan sizes of targets or foils.[2] The parameter is shared across targets and foils and is adjusted by by-subjects random effects $u_{subj[n],2}$. The final value is exponentiated to ensure that the accumulation penalty due to increased fan is always positive.
- The fixed-effect parameter $\beta_{penalty}$ is a penalty parameter that reduces the drift rate for incorrect response options, reflecting participants' bias toward the target.[3] As before, it is adjusted by by-subjects random effects and exponentiation ensures that accumulation penalty for choosing foils is positive.

Obviously, for each trial n we only observe the time associated with the winner w of the race process, whichever it is (the target, or one of the three foils). The time per trial is thus

$$T_{w,n} \sim \text{LogNormal}(\log(F) - A_{w,n}, \sigma). \tag{10}$$

However, by the fact that the other candidates were not the winners, we also get some information about them: we know that their evidence accumulation was not faster than that of the winner. We can calculate the probability of this happening by integrating out the time, which is at least as large as the time of the winner.

$$\mathbb{P}(T_{s \neq w,n} \geq T_{w,n}) = \int_{T_{w,n}}^{\infty} \text{LogNormal}(T | \log(F) - A_{s,n}, \sigma) dT \tag{11}$$

Lastly, next to the retrieval process, the model includes a non-decision time parameter T_{nd} to account for encoding and motor response processes, so the observed RT is the addition of the two.

$$RT_n = T_{w,n} + T_{nd} \tag{12}$$

Before we turn to the modeling results we want to point out some intuitions that the model should capture. First, higher fan sizes reduce the evidence accumulation leading to longer RTs. The penalty parameter similarly reduces drift rates for foils, making the correct response more likely to win the race. However, incorrect responses can still win for two reasons. First, due to trial-by-trial variability captured in the noise parameter σ, a foil can occasionally sample a higher drift rate and win the race. Second, fan effects can favor certain foils: when the response has a high fan (i.e., fan 4) but a foil has a low fan (i.e., fan 2), the foil benefits from faster evidence accumulation. If the benefit of a lower fan outweighs the penalty for incorrect responses, incorrect information can be retrieved.

[2] This parameter is closely related to the log-transformed attentional weighting parameter, W, in ACT-R.

[3] This parameter is closely related to the log-transformed match scale parameter, P, in ACT-R.

Implementation of the log-normal race model

We fit the log-normal race model to data from the second repetition of the learning phase in a Bayesian statistical framework (Nicenboim et al., 2025) using the probabilistic programming language Stan (Carpenter et al., 2017) with the `cmdstanr` package (Gabry et al., 2025) in R.

We used weakly informative priors to regularize parameter estimates while allowing the data to determine the posterior distributions:[4]

- $\alpha \sim \text{Normal}(0, 2)$
- $\beta_{fan} \sim \text{Normal}(-1, 2)$
- $\beta_{penalty} \sim \text{Normal}(-1, 2)$
- $\sigma \sim \text{Normal}(0.5, 0.2)$ truncated to positive values.

Furthermore, we constrained the non-decision time to be positive and less than the minimum observed RT: $T_{nd} \sim \text{Normal}(300, 100)$, truncated at $(0, \min(RT))$. For by-subject random effects, we assumed the parameter $\tau \sim \text{Normal}(0.1, 0.1)$ truncated at zero for the group-level standard deviations.

The posteriors were sampled using 4 chains, with 2,000 draws per chain, half of which constituted the warm-up phase. The Rhat values of all the model parameters were below 1.011 and ESS values were greater than 800 for Bulk ESS and greater than 1,400 for Tail ESS, indicating convergence and reliability of the sampled results. We also verified that we could recover the parameters from the non-hierarchical version of the model using the fake-data simulation (see Chapter 18, Nicenboim et al., 2025).

Results

To inspect the fit to the data, we collected 100 random draws from full model simulations and compared them to the observed data. Figure 1 compares the model and observed response time distributions. The model approximates the observed data well. The density plots show that the model captures the slowdown of incorrect responses across both fan sizes. The violin plot shows a close overlay of predicted and observed distributions, demonstrating that the model captures the effect of accuracy and RTs. The fan size effect is also approximated well, albeit the effect is small and thus harder to observe in this figure.

Figure 2 shows a quantile probability plot comparing observed and predicted response proportions and RT quantiles, split by accuracy and fan condition. First, higher fan size reduces accuracy (shifting correct responses leftward and incorrect responses rightward), and the model captures this effect. Second, the increased fan size leads to a slowdown from the median quantile onward, while its effect seems absent or even reversed for the lower quantiles. The model correctly captures this pattern. Third, incorrect responses are slower than correct responses across all quantiles, both in the model and in the observed data. Finally, while the fit is not perfect across all quantiles, the model captures both accuracy and RT patterns well.

[4]Recall that the α and β parameters are exponentiated in the model formula, hence both positive and negative numbers are transformed into positive values.

Parameter	Mean	SD	5%	95%
T_{nd}	12.927	10.592	0.885	33.995
α	1.457	0.034	1.402	1.514
$\beta_{penalty}$	0.773	0.031	0.722	0.825
β_{fan}	0.029	0.013	0.009	0.051
σ	0.657	0.005	0.649	0.665

Table 3: Posterior summaries of selected parameters in the model.

The estimates of the selected model parameters are shown in Table 3. T_{nd}, given in ms, represents the non-decision time. α is the logarithm of the base-level activation, shared across targets and foils. The foil penalty $\beta_{penalty}$ represents the decrease in log-activation due to the incorrect match, i.e., recall of the foil instead of the target. The fan penalty β_{fan}, shared across targets and foils, models the fan effect in logarithmic scale. Finally, σ is the noise parameter.

The parameter posteriors are narrow, suggesting the data are informative about these parameters. The foil penalty $\beta_{penalty}$ is greater than the fan penalty β_{fan}, i.e., the incorrect match decreases activation more than an increased fan size. This is intuitively plausible: the target should be preferred over a foil even if the former has a higher fan than the latter. However, ACT-R treats the $\beta_{penalty}$ and β_{fan} parameters as independent and does not necessarily predict this effect. The non-decision time is very short, which is likely related to our conservative RT cutoff of 300 ms. Finally, the base-level activation might seem quite high. However, this parameter should not be interpreted in isolation. It scales with the scaling parameter F, which we also set quite high, at $\log(F) = 10$.

Discussion

In this study, we modeled the fan effect during the learning phase of a fan experiment using a hierarchical Bayesian log-normal race model. Using such a model has several benefits over traditional modeling approaches. First, the hierarchical structure allows us to capture individual differences in how fan effects emerge across participants while simultaneously estimating population-level effects. Second, the Bayesian framework provides full posterior distributions over all parameters, allowing us to quantify uncertainty in our estimates. Third, it enables us to model RTs and accuracy jointly at the trial level. This reveals patterns that would be obscured by analyzing summary statistics alone, such as the relationship between fan size and the shape of RT distributions.

The first question we addressed was whether fan effects already emerge before participants have fully learned the associations between concepts. Our descriptive analyses confirm that fan effects are present in both iterations of the learning phase and increase with the progress in learning. This

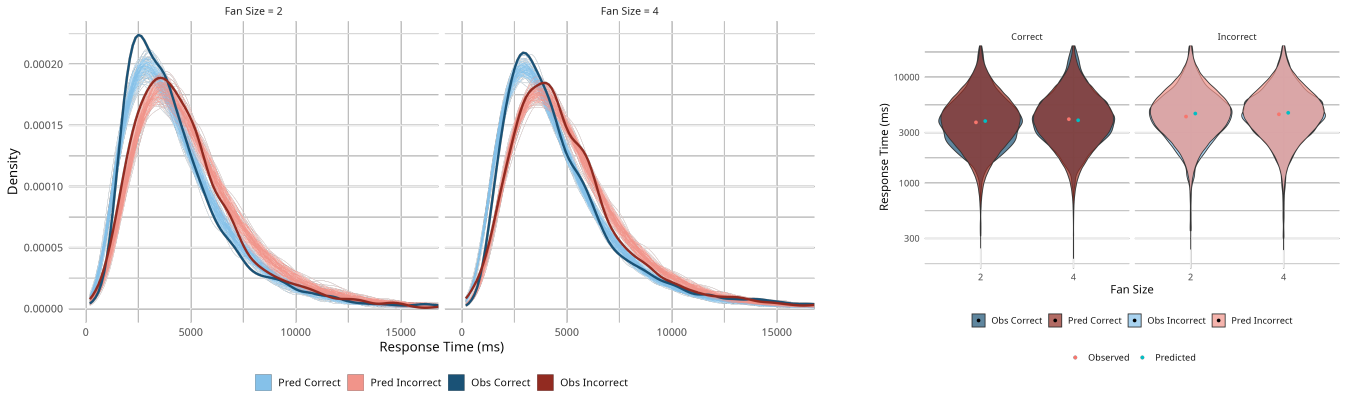

Figure 1: Posterior predictive checks comparing observed response time distributions to simulated data. Left: The density plots split by fan effect. Dark lines represent observed data, light lines represent simulated data. Right: The violin plot overlays predicted and observed distributions and compares observed and predicted medians.

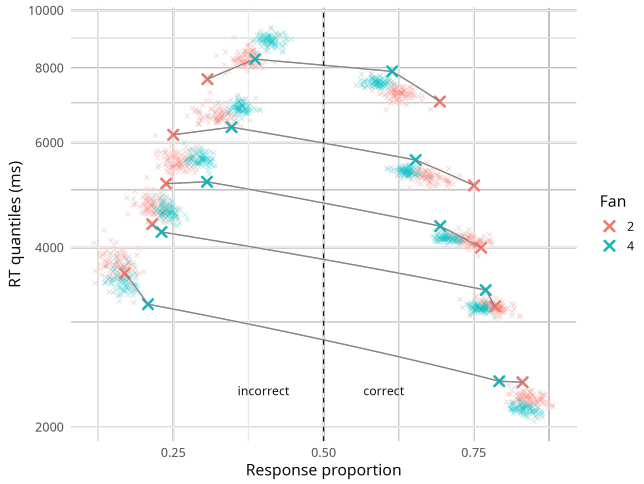

Figure 2: Quantile probability plot, showing response proportions of incorrect and correct responses, split by fan, and reaction times (RT) for 0.1, 0.3, 0.5, 0.7 and 0.9 quantiles. Large boldface crosses represent observed accuracies and RT means, small light crosses are model predictions. Incorrect responses have response proportions below 0.5 and appear on the left, correct responses have response proportions above 0.5 and appear on the right. The lines within each quantile are used to highlight the fan and accuracy differences.

suggests that interference builds as associations between the concepts are strengthened. The log-normal race model successfully captures these effects. This demonstrates that the competition between associated concepts can account for fan effects, even during learning.

Further, we aimed to model RTs and accuracy jointly. The posterior predictive checks of the log-normal race model show that it captures both patterns in the data, including a slowdown for incorrect responses. This confirms the prediction made by ACT-R that incorrect responses arise from a competition between response options. The incorrect option can win either because a fan advantage outweighs the penalty for being incorrect, or because of variability in drift rates.

Lastly, we aimed to model the fan effect at the process level using a race model. As we show in Figure 2, the model can, in addition to capturing overall accuracy and RT patterns, also capture the shape of their distributions.

Limitations and future work

The present work has several limitations that point to directions for future research. First, we did not account for mouse position in our analysis. Since participants responded by clicking on one of four response options, the physical distance to the correct response may have influenced RTs. This could confound fan effects with motor preparation and execution costs. Mouse position could be included in future models.

Second, our model does not account for the previously selected response options or word frequency, which are factors known to affect memory retrieval. Future models could include these as predictors of drift rates.

Third, the log-normal race model could be compared to other models of memory retrieval, such as the direct-access model (Van Dyke and McElree, 2011). The direct-access model assumes that retrieval can bypass the competitive interference mechanism and that longer RTs are a result of revisions of incorrectly retrieved responses. Comparing different models could demonstrate whether the assumption of a race between retrieval competitors is necessary to account for the observed behavioral patterns.

Finally, we only model data from the learning phase. It remains to be seen whether the same model can be fit equally well to the testing phase, when participants have fully acquired the associations and interference effects are presumably stronger.

Acknowledgments

The research reported in this paper was supported by the European Research Council (ERC), grant 101088098 - MEMLANG. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

We thank Leendert van Maanen for his role in the conceptualization and supervision of the experiment underlying this work.

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.

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought*. Psychology Press.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474.

Bunting, M. F., Conway, A. R., & Heitz, R. P. (2004). Individual differences in the fan effect and working memory capacity. *Journal of Memory and Language*, *51*(4), 604–622.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1101.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

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

Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2025). *Cmdstanr: R interface to 'cmdstan'* [R package version 0.9.0, https://discourse.mc-stan.org]. https://mc-stan.org/cmdstanr/

Nicenboim, B., Schad, D. J., & Vasishth, S. (2025). *Introduction to Bayesian data analysis for cognitive science*. CRC Press.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.

Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., & Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *80*(2), 491–513.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263.

Van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive science*, *36*(1), 62–101.