Running head: COGNITIVE PROCESSES UNDERLYING THE TRUTH EFFECT

Cognitive Processes Underlying the Repetition-Based Truth Effect: A Diffusion Model Study

Annika Stump[12], Lukas Schumacher[3], Andreas Voss[2], & Karl Christoph Klauer[1]

[1]Institute of Psychology, University of Freiburg, Germany

[2]Institute of Psychology, Heidelberg University, Germany

[3]Institute of Psychology, University of Basel, Switzerland

Author Note

This research was financially supported by a research grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK 2277 "Statistical Modeling in Psychology". We have no conflicts of interest to disclose.

The data and code are publicly accessible at https://github.com/LuSchumacher/truth_ddm.

Annika Stump and Lukas Schumacher contributed equally to this work.

Author Contributions:

Annika Stump: Conceptualization; Data curation; Formal analyses; Investigation; Methodology; Project administration; Software; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Lukas Schumacher: Data curation; Formal analyses; Methodology; Visualization; Writing – original draft; Writing – review & editing.

Andreas Voss: Conceptualization; Data curation; Methodology; Writing – review & editing.

Karl Christoph Klauer: Conceptualization; Methodology; Writing – review & editing.

Correspondence concerning this article should be addressed to Dr. Annika Stump, Institute of Psychology, University of Freiburg, Engelbergerstraße 41, 79085 Freiburg im Breisgau, Germany, email: annika.stump@psychologie.uni-freiburg.de, phone: +49-761-2032467.

Abstract

People are more likely to judge repeated (vs. novel) information as true. This illusory truth effect is commonly attributed to increased processing fluency, with perceptual and semantic aspects of repetition assumed to enhance perceptual and conceptual fluency, respectively. While prior research has shown that the truth effect is affected by the temporal gap between presentations, the cognitive mechanisms underlying this effect – and how they are shaped by retention interval – remain insufficiently understood. In Experiment 1 ($N = 75$), we manipulated repetition (repeated vs. new statements) and retention interval (10 minutes vs. 1 week) within participants. Using diffusion modeling, we found that repetition increased drift rate, reduced non-decision time, and induced a bias toward "true" responses – suggesting contributions from both conceptual and perceptual fluency. These effects diminished with increasing retention interval. In Experiment 2 ($N = 81$), participants judged the validity of one-word answers to previously shown questions, allowing separate measurement of reading times and tighter control over non-decision time. When modeling only the answer-related response times, repetition did not affect drift rate, but the repetition-enhanced response bias remained. Reading times for repeated questions were reduced, indicating facilitated encoding. An additional analysis using total trial reaction times (reading + response time) revealed a repetition effect on drift rate, suggesting that information accumulation may begin during question reading. Together, these findings indicate that the illusory truth effect results from a combination of perceptual facilitation, response bias, and increased speed of information accumulation supporting the truthfulness of a statement – each contributing at different stages of the decision process.

Keywords: truth judgments, information repetition, truth effect, retention interval, cognitive processes, diffusion model

Cognitive Processes Underlying the Repetition-Based Truth Effect: A Diffusion Model Study

The simple repetition of information can lead to increased familiarity and perceived credibility of information – this phenomenon is known as the illusory truth effect (Dechêne et al., 2010; Hasher et al., 1977). The effect has been consistently observed in over 80 replications since its initial experimental demonstration by Hasher et al. (1977), establishing its robustness. Current literature suggests that the experience of relatively high processing fluency, resulting from repetition, plays a central role in the processes underlying the truth effect: When judging the validity of information, individuals tend to rely on their subjective feeling of ease during the information processing (Dechêne et al., 2010). In truth effect experiments, fluency is typically manipulated by presenting novel and repeated information (i.e., information presented in a previous experimental phase, the so-called exposure phase). Garcia-Marques et al. (2016) argue that this manipulation, which involves semantic and verbatim repetition, contributes to both conceptual and perceptual processing fluency (see also Garcia-Marques et al., 2015). In this terminology, perceptual processing fluency varies depending on how easily the stimuli are physically perceived, whereas conceptual processing fluency refers to the ease of processing caused by semantic-based manipulations; furthermore, it has to be taken into account that different fluency manipulations can interact (for a more detailed description, see Alter & Oppenheimer, 2009). From a cognitive modeling perspective, the repetition-based truth effect may involve distinct underlying mechanisms that differentially influence how information is processed and evaluated in terms of its truthfulness. Specifically, information repetition may facilitate encoding processes (i.e., perceptual fluency), resulting in a response bias toward “true” (vs. “false”), whereas semantic repetition may enhance the accumulation of evidence in favor of a “true” rather than a “false” judgment. Previous studies have already shown that both the unique manipulation of perceptual fluency (e.g., via manipulating visual contrast; Reber & Schwarz, 1999) and conceptual fluency (e.g., via the presentation of a semantically related text prior to the judgment phase; Arkes et al., 1991) can cause an illusory truth effect. However, findings also indicate that manipulations of conceptual processing fluency have a stronger effect on judged truth compared to manipulated perceptual fluency (Parks & Toth, 2006).

The time (i.e., the retention interval) between information presentations significantly impacts the magnitude of the repetition-based truth effect (e.g., Henderson et al., 2021). A body of research suggests that the retention time affects different affective and cognitive processes underlying the illusory truth effect (e.g., Nadarevic & Aßfalg, 2017; Schnuerch et al., 2021; Silva et al., 2017; Stump et al., 2022). Silva et al. (2017) systematically manipulated processing fluency via repetition with different conceptual und perceptual features (e.g., paraphrases vs. verbatim repetitions; contradictory vs. non-contradictory statements) as well as the retention interval length (same experimental session vs. 1 week). Experiment 1 showed that verbatim repetitions were judged as true as paraphrases in both judgment phases (same experimental session vs. 1-week

retention interval). In Experiment 2, the authors observed an illusion of falseness for contradictory repetitions, which maintained perceptual characteristics of the prior presented statements, only after the short (10-minute) interval. This pattern of results seems to be well in line with previous findings by Garcia-Marques et al. (2015), demonstrating that participants were more likely to rate contradictory statements as false compared to novel statements after a relatively short interval of a few minutes but were more likely to rate them as true after a longer (1-week) interval. However, in the study by Silva et al. (2017) participants perceived contradictory paraphrases still as less true compared to original statements after the longer interval. The authors concluded that a time-related reduction of conceptual fluency may enable perceptual processing fluency to take over when judging the truthfulness of information.

Taken together, there may be a decreasing strength of the involved memory traces as a function of time, influencing the magnitude of the truth effect and its underlying processes. However, it remains unclear which cognitive components exactly are altered when processing the same information repeatedly, and whether and how the specific mechanisms are influenced by time.

The diffusion decision model (DDM, Ratcliff, 1978) provides a mathematical framework for an integrated analysis of responses and response times (RTs) from binary decision tasks that allows insights into different processing components. Therefore, it offers an excellent method for gaining deeper insights into cognitive mechanisms underlying the repetition-based illusory truth effect.

In the following, we present a concise introduction to the DDM before describing our hypotheses (for a more detailed description of diffusion decision modeling, see Voss et al., 2013). The model assumes an underlying process that reflects the accumulation of evidence for potential decisional outcomes. This process evolves within a corridor bounded by two thresholds and terminates when either the upper or lower threshold is hit. In DDM analyses, several psychologically meaningful model parameters are estimated from the distributions of RTs (Voss et al., 2004). In its simplest form, the DDM consists of the following parameters: the starting point, non-decision time, threshold separation, and drift rate. The starting point represents an a-priori bias towards a particular response (if the starting point is not centered between the two thresholds). Threshold separation captures individual decision caution, where a small (large) separation indicates a more liberal (conservative) response style characterized by fast and inaccurate (slow and accurate) responses. The drift rate quantifies the speed of information processing, with larger drift rates indicating faster information accumulation for the decision associated with the upper threshold. Finally, the non-decision time accounts for the duration of all processes prior and after the decision process (e.g., encoding and motor response).

We argue that the drift rate, as a measure of information processing speed, primarily reflects the conceptual processing fluency, whereas the non-decision time, which comprises the time for encoding, should primarily map perceptual processing fluency. Therefore, we expect an increased

drift rate (indicating faster information accumulation for the judgment "true") and a decreased non-decision time (mapping reduced encoding time) during the processing of repeated (vs. novel) information. Previous research suggests that manipulations of conceptual processing fluency have a stronger effect on judged truth compared to manipulated perceptual fluency (Parks & Toth, 2006). Thus, it remains an open question how strongly perceptual processing fluency contributes to the repetition-induced truth effect. If perceptual processing fluency – indicated by a reduction in non-decision time – plays a central role in the mechanisms underlying the repetition-induced truth effect, we would expect not only a reduction in non-decision time but also a bias toward "true" responses. Furthermore, we expect that the effect of repetition on drift decreases as a function of retention interval length (due to a time-related reduction in semantic-based processing fluency). To which extent the retention interval length influences the effect of repetition on non-decision time and the starting point is an open question, considering the inconsistency of prior research findings.

Open Practices Statement

With respect to Experiment 1, we analyzed data from a larger study which was designed to address several research questions. Informed consent was obtained from all study participants. This study was approved by the ethics committee of Heidelberg University (Faculty of Behavioural and Cultural Studies) and conducted in accordance with the Declaration of Helsinki. Experiment 2 aimed to replicate and extend the findings from Experiment 1 with tighter control over the non-decision time. In the following, we describe all procedures and measures that are relevant to the present hypotheses. Experiment 1 was not preregistered; Experiment 2 has been preregistered (https://osf.io/uz5fx/resources). The data and code of both experiments are publicly accessible at https://github.com/LuSchumacher/truth_ddm.

Experiment 1

In this experiment, participants were asked to judge the truthfulness of various statements. Processing fluency was systematically manipulated by presenting 50% of the statements during a prior exposure phase. This repetition-based truth effect study employed two retention intervals (10 minutes vs. 1 week) and was designed to meet the requirements for DDM analysis, namely speeded binary truth judgments.

Method

Participants and Design

A total of 81 participants were recruited through the recruitment software hRoot (Bock et al., 2014). Data from three participants who did not attend the second experimental session, two participants who aborted the experiment, and one participant who encountered a technical error were excluded from the analysis. The final sample included 75 participants (61 female; 13 male; 1 diverse), aged between 18 and 31 years (M = 21.63, SD = 2.96). The majority of participants (68%) were non-psychology students. Participants received either course credit or a compensation of 15 Euros (approximately 16 USD) for the study participation.

The study design comprised the within-subject factors *factual truth* (correct vs. incorrect), *repetition status* of presented statements (new vs. repeated), and *retention interval length* (10 minutes vs. 1 week after the exposure phase).

Material

The main statement material comprised 120 statements in German language, organized into four sets. To ensure that the actual truth status of the items was unknown for most participants, we selected statements that have been carefully pretested (Nadarevic, 2010). All statements were of similar length to ensure comparable processing time. Furthermore, only statements which have been perceived as affectively neutral were used to avoid affective reactions. Each of the four item sets consisted of 15 true and 15 untrue statements, with similar mean truth ratings and standard deviations across the sets (set A: $M = 4.06$, $SD = 1.20$; set B: $M = 4.06$, $SD = 1.18$; set C: $M = 4.05$, $SD = 1.18$; set D: $M = 4.05$, $SD = 1.20$). During both experimental judgment phases, two sets of statements were used: one set with novel statements and one set with statements already shown in the exposure phase. The assignment of different item sets to the experimental phases was counterbalanced across subjects.

Procedure

After obtaining informed consent, electrodes to record emotional responses via EMG were placed on the participants' faces. Then, the computer-based experiment started. In the first experimental phase, items from two statement sets (each including 15 true and 15 false statements) were presented trial by trial in random order, and participants categorized them into one of six predefined knowledge categories (e.g., geography) by pressing the corresponding number keys. To mitigate potential primacy and recency effects, six additional buffer statements were presented at the beginning and end of this phase. Each trial began with the display of a fixation cross (1 s), followed by a statement that was presented in the center of the screen until the participant responded by pressing a number key. As for every following task, there was no time limit given, but participants were instructed to respond as quickly as possible while minimizing errors. Subjects were informed prior to the exposure phase that both true and false information would be presented.

After a retention interval of 10 minutes, during which participants engaged in a non-verbal filler task, the first judgment phase began. In this phase, 30 new and 30 repeated statements were presented trial-wise in random order.[1] The 30 new items were selected from one of the previously unused sets, while the 30 repeated statements were taken from one of the sets presented during the exposure phase. Again, a fixation cross was shown (1 s) prior to the statement presentations.

[1] Experiment 1 was part of a larger research project in which facial electromyography (fEMG) activity was recorded. To ensure precise synchronization between stimulus onset, muscle activity, and behavioral responses, response times were recorded using LUX sensors that detected light signals presented in the screen corners (not visible to participants). Because of LUX sensor errors in certain trials, these had to be excluded. This concerned 1.8% of all trials in the first judgment phase and 1.4% in the second judgment phase.

Each statement remained on the screen until participants made their truth judgment by pressing the key "W" ("wahr"; German for true) or "F" ("falsch", German for false). After each validity judgement, participants indicated their subjective confidence in their former judgment using the number keys (from 1 "very uncertain" to 6 "very certain").

Exactly 7 days after the first experimental session, the second judgment phase of the experiment took place. The procedure mirrored that of the first judgment phase, except for the use of different statement material. Specifically, this phase comprised the statements from the exposure phase that were not used in the first judgement phase as well as from the final unused item set.

Results

To account for the hierarchical structure of the data, we employed a multilevel modeling approach for all generalized linear models (GLMs). Analyses were conducted in R (version 4.3.1) using the brms package (Bürkner, 2017). Bayes factors (BFs) were computed using the Savage–Dickey density ratio method (Wagenmakers et al., 2010). DDM parameters were estimated using a hierarchical Bayesian framework (Vandekerckhove et al., 2011), implemented directly in Stan (Stan Development Team, 2024). Since we are dealing with relatively long and variable RTs, we additionally incorporated inter-trial variability in the non-decision time parameter, extending the model beyond the four core parameters, drift rate (v), threshold separation (a), starting point bias (β), and non-decision time (τ). Trial-wise non-decision time was defined as

$$\tau_t = \tau + \tau_{\text{var}}\zeta, \quad \zeta \sim U(0,1),$$

with τ representing the lower and $\tau + \tau_{\text{var}}$ the upper bound of the uniformly distributed non-decision time. All parameters ($v, a, \beta, \tau, \tau_{\text{var}}$) were estimated hierarchically and allowed to vary across conditions. Specifically, we employed a regression-based parameterization in which each parameter was modeled as a linear function of two effect-coded predictors (factual truth and repetition status) and their interaction, yielding four group-level estimates per parameter. Informative priors were specified for all group-level parameters, as detailed in Appendix A. Posterior samples were drawn using the Hamiltonian Monte Carlo algorithm (Betancourt, 2018), with four chains of 4,000 iterations for GLMs and 6,000 for DDMs, discarding the first 50% as warm-up. Convergence was assessed via the $\hat{R}$ statistic (Gelman & Rubin, 1992) and visual inspection of trace plots.

To ensure that the model was identifiable, we conducted a DDM parameter recovery analysis. Results showed excellent recovery for all parameters, confirming that the model could reliably recover the true values; full details are provided in Appendix B.

To test hypotheses about the DDM parameters, we computed BFs for the posterior distributions of the regression weights corresponding to the experimental manipulations statement repetition, factual truth, and their interaction. All effects were tested against zero. During estimation, some parameters with natural constraints (i.e., threshold, bias, and non-decision time

variability) were reparametrized onto unconstrained scales to improve sampling efficiency. For clarity and interpretability, all group-level DDM parameter summary statistics are reported on the original (constrained) scales. BFs for regression weights associated with constrained parameters are computed on the unconstrained scale; accordingly, their summary statistics are reported on the unconstrained scale. Model fit was evaluated via simulations from the posterior predictive distribution: we drew 200 samples from the joint posterior, simulated new data sets, and computed summary statistics for choice proportions and RT quantiles (medians and 95% credible intervals (CIs)), which were then compared with those from the observed data.

Before turning to the inferential analyses, we report descriptive response rates for the experimental conditions (separately for each session). In Session 1, the proportion of "true" (vs. "false") responses were .725 (*SD* = .447) for repeated true statements, .528 (*SD* = .499) for novel true statements, .749 (*SD* = .434) for repeated false statements, and .552 (*SD* = .498) for novel false statements. In Session 2, the corresponding proportions were .634 (*SD* = .482) for repeated true statements, .541 (*SD* = .499) for novel true statements, .650 (*SD* = .477) for repeated false statements, and .598 (*SD* = .491) for novel false statements. Descriptively, repetition was associated with higher proportions of "true" responses compared to novel statements across factual truth conditions in both sessions.[2]

The truth effect as a function of retention interval length

First, we analyzed participants' responses with a Bayesian logistic multilevel model that included the predictors session, repetition status, and factual truth, as well as all possible two- and three-way interactions. All predictors were effect coded. Figure 1 presents the posterior and prior distributions of the model parameters along with corresponding Bayes factors, and Figure 3 shows the observed choice frequencies.

We found evidence for a main effect of session (b = –0.10, 95% CI [–015, -0.05], BF_{10} = 11.74), indicating less "true" responses in the later session. Replicating the illusory truth effect, participants were substantially more likely to judge repeated (vs. novel) information as true (b = 0.31, 95% CI [0.25, 0.37], BF_{10} > 1000). This effect was moderated by session, as indicated by a credible interaction (b = -0.15, 95% CI [-0.21, -0.09], BF_{10} > 1000), suggesting that the repetition effect was stronger after a short retention interval. There was no strong evidence for a main effect of factual truth (b = -0.07, 95% CI [-0.11, -0.02], BF_{10} = 1.18). All remaining interaction effects were negligible with all BFs < 0.05. To account for differences in retention interval and to enhance comparability with prior work – most of which focuses on a single judgment phase – we conducted the following analyses separately for both judgment phases (i.e., after the short – 10 minutes – and long – 1 week – retention interval).

[2] To assess whether participants engaged in meaningful truth evaluation, we examined whether independent pretest truth ratings (Nadarevic, 2010) predicted responses using multilevel logistic regression models. Across both sessions, higher pretest plausibility significantly increased the likelihood of a "true" (vs. "false") judgment (session 1: b = .879, p < .001; session 2: b = 1.098, p < .001), indicating systematic sensitivity to item-level plausibility.

DDM modeling results

Convergence. $\hat{R}$ statistics for all model parameters were below 1.01, indicating convergence. Furthermore, the trace plots did not show any problematic behavior, which further supports the convergence of the models.

Absolute model fit. To assess absolute model fit, we compared the observed data to simulations from the posterior predictive distribution generated by our model. Figure 2 shows RT quantiles for both observed and re-simulated data, broken down by experimental session and statement repetition status. Overall, the re-simulated RT quantiles closely matched the observed data across all conditions, indicating strong agreement between the model and the empirical data. Choice data showed similarly strong alignment between observed and predicted response probabilities (Figure 3). The close match between the model predictions and the observed data supports the conclusion that the model provides an excellent fit to both RT and choice behavior.

Parameter estimates. To examine how factual truth, repetition status, and their interaction influenced the DDM parameters, we analyzed the posterior distributions of the corresponding regression weights. Positive weights indicate higher parameter estimates for repeated compared to new statements, true compared to false statements, or a stronger repetition effect for true than for false statements. When reporting regression weights, we refer to the median of the posterior distribution. Figure 4 presents the estimated regression weights for each DDM parameter (rows) across the three effects (columns). Table 1 provides summary statistics of the posterior distributions for each group-level DDM parameter per condition.

Inferences are based on posterior medians and 95% credible intervals, complemented by Bayes factors (BFs) computed via the Savage–Dickey density ratio method.

In Session 1, the drift rate for new statements was close to zero for both true and false statements (see Table 1), suggesting that participants were equally likely to respond “true“ or “false“. Repetition increased drift rates (false statements: median = 0.19, 95% CI = [0.09, 0.29]; true statements: median = 0.10, 95% CI = [–0.02, 0.20]), indicating a greater probability to respond “true“ and faster responses when a statement had been seen before. This resulted in a credible main effect of repetition (median = 0.08, 95% CI = [0.03, 0.13], BF_{10} = 6.05). Neither the main effect of factual truth (median = -0.02, 95% CI = [-0.07, 0.03], BF_{10} = 0.06) nor the interaction (median = -0.03, 95% CI = [-0.08, 0.02], BF_{10} = 0.11) were credible. In Session 2, the repetition effect on drift vanished (median = 0.02, 95% CI = [-0.082, 0.07], BF_{10} = 0.07), likely because drift rates for new false statements were already shifted above zero (median = 0.15, 95% CI = [0.06, 0.24]). Other effects on the drift rate were again not credible in Session 2.

Median group-level threshold values were relatively stable across conditions and sessions, with posterior medians ranging from 2.03 to 2.33 (see Table 1). In Session 1, there was weak evidence that thresholds were slightly higher for repeated than new statements (median = 0.11, 95% CI = [0.03, 0.19], BF_{10} = 2.13). This effect was absent in Session 2 (median = 0.00, 95% CI = [–0.08, 0.08], BF_{10} = 0.08). No other threshold effects were supported in either session (all BFs

< 0.33).

Across all conditions and sessions, participants showed a bias toward “true“ responses, as indicated by starting point bias estimates ranging from 0.52 to 0.69 (see Table 1). In Session 1, this bias was stronger for repeated compared to new statements (median = 0.27, 95% CI = [0.18, 0.36], $BF_{10} > 1000$). No substantial evidence for this effect was found in Session 2 (median = 0.09, 95% CI = [0.01, 0.17], $BF_{10} = 1.09$). Other effects on starting point bias were not credible in either session (all BFs were below 0.30).

Lower bounds of non-decision time ranged from 1.47 to 1.75 seconds across conditions and sessions (Table 1). In Session 1, repetition effects were observed (median = –0.11, 95% CI = [–0.17, –0.06], $BF_{10} = 99.92$), indicating longer non-decision processes for new compared to repeated statements. This effect was not present in Session 2 (median = –0.05, 95% CI = [–0.10, 0.00], $BF_{10} = 0.36$). No effect of truth or the interaction was found (all $BF_{10} < 0.07$).

Variability estimates ranged from 1.61 to 1.95 seconds. In Session 1, repetition again played a role (median = –0.26, 95% CI = [–0.41, –0.11], $BF_{10} = 32.32$), with greater variability for new compared to repeated statements. This effect was absent in Session 2 (median = –0.04, 95% CI = [–0.20, 0.12], $BF_{10} = 0.19$). No other effects were supported (all $BF_{10} < 0.24$).

We additionally examined whether repetition effects differed between the two sessions (short vs. long retention interval). This comparison was based on the posterior distributions of the respective effects. A clear difference emerged for the bias parameter (median = 0.18, 95% CI = [0.06, 0.30], $BF_{10} = 7.71$), whereas the repetition effects for drift rate and non-decision time were descriptively weaker after the longer retention interval (drift rate: median = 0.05, 95% CI = [-0.01, 0.12], $BF_{10} = 0.16$; non-decision time: median = -0.06, 95% CI = [-0.14, 0.01], $BF_{10} = 0.20$; variability in non-decision time: median = -0.22, 95% CI = [-0.45, -0.01], $BF_{10} = 1.03$; threshold separation: median = 0.11, 95% CI = [0.00, 0.22], $BF_{10} = 0.45$). Overall, this pattern, together with the observation that credible repetition effects on the theoretically relevant parameters – bias, drift rate, and non-decision time – were observed only after the short but not after the long retention interval, points to a modest but consistent reduction of repetition effects across all theoretically relevant parameters.

A potential concern is whether our model can reliably distinguish repetition effects arising from changes in starting point versus changes in drift rate, especially given the substantial estimated variability in non-decision time. To address this, we conducted an additional quantile-based analysis (see Appendix C) comparing model-predicted RT distributions when the repetition effect was imposed exclusively on either starting point or drift rate, calibrated to produce the same increase in “true” responses. Despite large non-decision time variability, the predicted RT quantiles differed qualitatively between these mechanisms, with starting-point changes shifting early quantiles and the median, and drift-rate changes primarily affecting the upper tail of the distribution. This diagnostic demonstration supports the interpretability of the DDM parameters in capturing distinct cognitive mechanisms underlying repetition effects.

Discussion. Results of Experiment 1 indicate that verbatim repetitions led to (a) increased drift rates, suggesting faster accumulation of information supporting the truthfulness of a presented statement, (b) reduced non-decision time, indicating facilitated perceptual processing, as well as (c) a response bias toward "true" (vs. "false"). All effects were noticeably reduced after a longer retention interval (10 min vs. 1 week).

Experiment 2

Experiment 2 aimed to replicate and extend the findings from Experiment 1, with tighter control over non-decision time. As in the first experiment, participants were asked to judge the truthfulness of various pieces of information, while processing fluency was systematically manipulated by presenting 50% of the statements during a prior exposure phase. In the judgment phase of this experiment, the information was presented in a new format, consisting of a question followed by a one-word answer. Reading times were recorded for the questions, and response times were measured for the one-word answers. This repetition-based truth effect experiment employed a single retention interval of one day.

Method

Participants and Design

A total of 83 participants were recruited through the recruitment software SONA and attended both experimental sessions. All subjects had been prescreened to speak German as their native language and to have normal or corrected-to-normal vision. Data from two participants who entered different personal identification codes in experimental sessions 1 and 2 were excluded from the analysis. The final sample included 81 participants (62 female; 16 male; 3 diverse), aged between 19 and 57 years (M = 24.24, SD = 6.56). The majority of participants (68%) were non-psychology students. Participants received either course credit or a compensation of 10 Euros (approximately 11 USD) for the study participation. The study design comprised the within-subject factors *factual truth* (correct vs. incorrect) and *repetition status* of presented statements (new vs. repeated).

Material

The main statement material comprised 120 statements in German language, organized into two sets. As in Experiment 1, we selected statements that have been carefully pretested (Nadarevic, 2010) to ensure that the actual truth status of the items was unknown for most participants. Furthermore, only statements which have been perceived as affectively neutral were used to avoid affective reactions. Each of the item sets consisted of 50% correct/incorrect statements, with similar mean truth ratings and standard deviations across the sets (set A: M = 4.06, SD = 1.19; set B: M = 4.06, SD = 1.19). During the experimental judgment phase, two sets of statements were used: one set with novel statements and one set with statements already shown in the exposure phase. For the judgment phase, the original statements were changed to a format consisting of a question and a one-word answer (e.g., "What is the capital of Bahrain?" – "Manama"). The assignment of the item sets to the experimental phases was randomized across participants.

Procedure

After obtaining informed consent, the computer-based experiment started. In the first experimental phase, statements from one of the two item sets were presented trial by trial in random order, and participants categorized them into one of six predefined knowledge categories (e.g., geography) by pressing the corresponding number keys. The whole procedure during this phase mirrored that of the first experiment.

After a retention interval of one day, 60 new and 60 repeated pieces of information (i.e., pairings of questions and one-word answers) were presented during the experimental judgment phase. Within the instruction prior to this judgment phase, the participants were reminded that half of the information presented during the first phase of the experiment was true/false. A fixation cross was shown (1 s) prior to each presented question. Since reading speed can vary substantially from person to person, the respective questions were presented until the subjects indicated via keypress (space bar) that they had finished reading. This was followed by the presentations of the corresponding one-word answers, which in turn were shown until their validity was judged. Subjects were informed that RTs would be recorded, accordingly participants were instructed to press the space bar immediately after they had finished reading each question. With respect to the corresponding one-word answers, subjects were instructed to make their truth judgments (via key press) as quickly as possible, while avoiding unnecessary mistakes. Each one-word answer remained on the screen until participants made their truth judgment by pressing the key "S" or "K". Following each validity judgement, participants had to indicate their subjective confidence in their former judgment using the number keys (from 1 "very uncertain" to 6 "very certain"). After each certainty rating, participants were instructed to return their thumbs on the space bar, and their index fingers on the letter keys "S" and "K" to be able to respond as quickly as possible in the next trial.

Results

To analyze the data from Experiment 2, we first performed a Bayesian multilevel lognormal regression on the reading times with *repetition status* and *factual truth* as (effect coded) predictors. Then, we applied the same modeling approach as used in Experiment 1, with the exception that Experiment 2 included only one judgments phase (1 day after first exposure). Figure 7 presents the observed response proportions, and Figure 5 shows the posterior and prior distributions of the model parameters, along with the corresponding BFs, derived from the multilevel logistic regression.

Analysis of the reading times revealed that – on average – individuals were faster to read repeated compared to new statements (b = -0.02, 95% CI [-0.03, -0.01], BF_{10} > 1000). No difference in reading times between factual true and false statements was found (b = 0.00, 95% CI [-0.01, 0.00], BF_{10} = 0.01). Further, the interaction between both factors (factual truth and repetition status) was not credible as well (b = 0.01, 95% CI [-0.01, 0.00], BF_{10} = 0.01).

As in Experiment 1, we observed a robust illusory truth effect: participants were more likely

to judge repeated statements as true than novel ones ($b = 0.18$, 95% CI [0.13, 0.24], $BF_{10} > 1000$). Also, there was no credible main effect of factual truth ($b = 0.00$, 95% CI [–0.05, 0.104], $BF_{10} = 0.02$) and no evidence for an interaction between factual truth and repetition status ($b = 0.01$, 95% CI [–0.03, 0.06], $BF_{10} = 0.03$).

Again, before turning to the inferential analyses, we report descriptive response rates for the experimental conditions. The proportion of "true" (vs. "false") responses was .648 ($SD = .478$) for repeated true statements, .561 ($SD = .496$) for novel true statements, .640 ($SD = .480$) for repeated false statements, and .569 ($SD = .495$) for novel false statements. Descriptively, repetition was associated with higher proportions of "true" responses compared to novel statements across factual truth conditions.[3]

Diffusion Modeling Results (Based on One-Word Answer Response Times)

Convergence. All model parameters exhibited $\hat{R}$ statistics below 1.01, indicating good convergence. In addition, inspection of the trace plots revealed no concerning patterns, providing further support for stable and reliable model estimation.

Absolute model fit. As shown in Figures 6 and 7, the model predictions closely matched the observed RT and choice data, indicating a strong absolute fit.

Parameter estimates. Figure 8 presents the posterior distributions of the regression weights corresponding to the effects (columns) on each DDM parameter (rows). Summary statistics of the posterior distributions (medians and 95% credible intervals) of each condition-wise group-level DDM parameter are reported in Table 2. Inferences are based on posterior medians and 95% credible intervals, complemented by Bayes factors (BFs) computed via the Savage–Dickey density ratio method.

On average, participants' drift rates were centered around zero, ranging from –0.04 to 0.03. We found no evidence for any of the three effects of interest, with all BFs < 0.07, indicating that the efficiency of evidence accumulation was not influenced by statements' repetition status or factual truth. Group-level threshold separation estimates were stable across conditions and comparable to Experiment 1, ranging from 2.02 to 2.09. Accordingly, no evidence for any effects of interest was observed (all BFs < 0.14).

Consistent with Session 1 of Experiment 1, participants exhibited an a priori bias toward "true" responses, with estimates ranging from 0.58 to 0.66 across conditions. Strong evidence indicated that repetition status affected this bias, with a stronger bias for repeated compared to new statements (median = 0.12, 95% CI = [0.08, 0.17], $BF_{10} > 1000$). No effects of factual truth or the interaction were observed (all BFs < 0.18).

As expected, non-decision times were shorter than in Experiment 1. Lower-bound

[3] As for the data of Experiment 1, to assess whether participants engaged in meaningful truth evaluation, we examined whether independent pretest truth ratings (Nadarevic, 2010) predicted responses using multilevel logistic regression models. Again, higher pretest plausibility significantly increased the likelihood of a "true" (vs. "false") judgment ($b = .619$, $p < .001$), indicating systematic sensitivity to item-level plausibility.

estimates ranged from 0.63 to 0.66 seconds across conditions, showing minimal differences between conditions. The variability parameter was also smaller than in Experiment 1, ranging from 0.25 to 0.58. Unlike the clear repetition effect on variability observed in Session 1 of Experiment 1, evidence in Experiment 2 was weak (median = –0.11, 95% CI = [–0.20, –0.02], BF_{10} = 1.36). Similarly, no evidence was found for effects of factual truth or the interaction on non-decision time variability (both BFs < 0.22).

Discussion. Reading times were systematically reduced for repeated (vs. new) questions, indicating faster encoding when information is presented repeatedly. Furthermore, results of Experiment 2 indicated a repetition-induced response bias toward "true" (vs. "false"). None of the other DDM parameters were substantially influenced by repetition when analyzing reaction times for one-word answers only. These findings may appear somewhat unexpected, given the clear effect of repetition on drift rate after the short interval in Exp. 1. However, we argue that the information accumulation process might already have begun while participants were reading the questions – prior to the presentation of the corresponding one-word answers. This seems plausible, as no unexpected answers were presented; that is, for repeated questions, the one-word answers always matched the information from the exposure phase. Second, we observed a repetition effect on the response bias, but not on the non-decision time parameter. This suggests that early perceptual encoding may have been less strongly facilitated by repetition in Experiment 2 than in Experiment 1 – possibly due to the considerable change in visual format between exposure and test phases in Exp. 2 (i.e., from full statements to a question-answer format). Nonetheless, reading times were systematically reduced for repeated questions, indicating more fluent processing and potentially an earlier onset of evidence accumulation in support of the truthfulness of the statements. To examine this possibility, we conducted an additional analysis in which we re-ran our DDM model using total reaction times per trial (i.e., the sum of the reading time for the question and the response time for the respective one-word answer).

Diffusion Modeling Results (Based on Total Response Times)

Convergence. All model parameters showed $\hat{R}$ values below 1.01, indicating good convergence. Furthermore, an inspection of the trace plots revealed no problematic patterns, supporting the stability and reliability of the model estimates.

Absolute model fit. As shown in Figures 9 and 10, the model predictions closely matched the observed RT and choice data, indicating a good absolute fit. However, compared to the model fit based on the one-word answer reaction times, the fit was slightly worse. Specifically, the model tended to underestimate fast responses and overestimate slow responses.

Parameter estimates. Figure 11 displays the posterior distributions of the regression weights corresponding to the effects (columns) on each DDM parameter (rows). Summary statistics of these posterior distributions, medians and 95% credible intervals, for each condition-wise group-level DDM parameter are reported in Table 3. Inferences are based on posterior medians and 95% credible intervals, complemented by Bayes factors (BFs) computed via the Savage–Dickey

density ratio method.

Consistent with the findings from Experiment 1, repetition increased drift rates (median = 0.06, 95% CI = [0.03, 0.09], $BF_{10} > 1000$), indicating a higher speed of information accumulation in support of a statement's truthfulness. The data did not provide substantial evidence for either a main effect of factual truth ($BF_{10} = 0.04$) or the interaction effect ($BF_{10} = 0.03$). Group-level estimates of threshold separation were stable across conditions and comparable to previous analyses (all BFs < 0.07). We also found no effect of repetition on response bias ($BF_{10} < 0.08$). Similarly, no effects of factual truth or its interaction with repetition were observed (all BFs < 0.10). Non-decision times differed only minimally between conditions (all BFs < 0.11). As in Experiment 1, we found a clear repetition effect on non-decision time variability (median = -0.12, 95% CI = [-0.19, -0.05], $BF_{10} = 8.52$). Again, no strong evidence was found for effects of factual truth or the interaction on non-decision time variability (both BFs < 0.16).

Discussion Additional analyses of Experiment 2, using total response times, replicated the clear effect of repetition on drift rate. Apart from variability in non-decision time, none of the other DDM parameters were credibly affected by repetition. These results support our theoretical assumption that the information accumulation process likely already begins while participants are reading the questions, prior to the presentation of the corresponding one-word answers.

General Discussion

The aim of the present study was to examine the extent to which repetitions impact different cognitive processes during truth evaluations. The DDM offers a mathematical framework for an integrated analysis of participants' responses and RTs from binary judgments, where the model's parameters allow insights into different components underlying the judgment process. Since prior research indicates that the retention interval length substantially influences the repetition-induced truth effect (e.g., Henderson et al., 2021), we were furthermore interested in time-related changes in the effects of repetition on cognitive processes. For this purpose, we conducted a truth effect study (Experiment 1) including two judgment phases (10 minutes and 1 week after first exposure) that fulfilled the requirements for DDM analyses (i.e., speeded binary truth judgments). Experiment 2 was conducted to replicate and extend the findings from Experiment 1, with a tighter control of non-decision time. As in Experiment 1, participants judged the truthfulness of statements, with fluency manipulated by pre-exposing 50% of them. However, this time, each piece of information was presented in the form of a question followed by a one-word answer during the judgment phase, enabling a separate measurement of reading times and tighter control over non-decision time when analyzing RTs for validity judgments.

The results of Experiment 1 show that verbatim repetition led to (a) increased drift rates, indicating faster accumulation of information supporting the truthfulness of a presented statement (related to higher conceptual processing fluency), (b) reduced non-decision times, indicating facilitated perceptual processing as well as (c) a response bias toward "true" (vs. "false"). In line with previous research demonstrating time-related reductions of the illusory truth effect (e.g.,

Henderson et al., 2021), all repetition effects were noticeably reduced after a longer retention interval (10 min vs. 1 week). These time-related findings, which are evident in all repetition-based effects on our DDM parameters, suggest that both conceptual and perceptual processing fluency are decreasing as interval length increases.

The results of Experiment 2 provide a more nuanced picture of how repetition influences the cognitive mechanisms underlying truth judgments. When modeling only the response times for the one-word answers, repetition led to a response bias toward judging statements as "true", but it did not credibly influence the other model parameters, including drift rate. Reading times for repeated (vs. new) questions were systematically reduced, indicating facilitated processing of previously encountered information.

To account for this pattern of results, we argued that the process of information accumulation might already begin while participants are reading the questions – prior to the presentation of the corresponding one-word answers. That is, when the onset for diffusion modelling is set to a later point, the evidence accumulated so far will then be incorporated into the starting point. This interpretation seems plausible for two reasons. First, no unexpected answers were presented in the experiment: For repeated questions, the one-word answers were entirely predictable and consistent with the information provided during the exposure phase. Second, we observed an effect of repetition on bias but not on the non-decision time parameter. At the same time, reading times were systematically reduced for repeated questions, suggesting that these were processed more fluently and may have triggered early accumulation of evidence supporting their truthfulness. To test this idea, we conducted an additional analysis in which we re-ran the DDM using total reaction times per trial – that is, the sum of question reading time and answer response time. This approach allowed us to better capture the full temporal scope of evidence accumulation in the context of truth evaluation.

This additional analysis revealed a clear repetition effect on drift rate, replicating the finding from Experiment 1. At the same time, there was no effect of repetition on non-decision time and starting point in this analysis. This suggests a diminished role of perceptual fluency in facilitating encoding in Experiment 2, which agrees well with the facts that (a) perceptual fluency has been linked to verbatim repetition as explained in the Introduction and (b) that repetition was verbatim in Experiment 1, but not in Experiment 2 due to the shift to a question-plus-answer format in the judgment phase in that experiment. Conceptually, early evidence accumulation toward the upper boundary prior to the one-word answer onset can be formally absorbed into the starting point when the model is time-locked to the answer. However, a pure starting-point shift would imply a one-time boost without affecting the ongoing integration of truth-relevant evidence. In contrast, the drift-rate effect observed in the total-RT model suggests that repetition alters the accumulation process itself rather than merely providing an initial boost. Taken together, the results support the idea that the repetition-based truth effect is driven, at least in part, by increased speed of information accumulation (i.e., higher drift rates), which may reflect enhanced

conceptual fluency during the processing of repeated information. While repetition may also facilitate perceptual encoding, as evidenced by repetition effects on non-decision time (Exp. 1) and reduced reading times (Exp. 2), our findings suggest that perceptual fluency and an associated response bias toward "true" do not, by themselves, shape the mechanisms underlying the repetition-based truth effect. Instead, distinct perceptual and conceptual fluency components appear to contribute to different aspects of the decision process.

Based on findings indicating that remembering information that has been presented can increase subjective truth (e.g., Nadarevic and Erdfelder, 2025), one might argue that the faster information accumulation rather represents facilitated recollection processes (vs. conceptual processing fluency). However, our pattern of results does not suggest that the repetition-based changes in drift rates are related to explicit recollection of statements. We explicitly informed our participants prior to the exposure phases that both true and false statements would be presented.[4] If subjects explicitly remembered individual statements from the experimental exposure phase, this should not be a valid cue for the evaluation of truth. In fact, earlier research findings by Arkes et al. (1989) even suggest that the truth effect is reduced when repeated statements are attributed to the experiment. In contrast, our results clearly show that the illusory truth effect, as well as repetition-based changes in drift rates, are substantially increased after a shorter retention interval, i.e., when memory strength should be even more pronounced (compared to longer retention interval lengths). However, the idea that repetition-induced faster information accumulation is associated with increased conceptual processing fluency is consistent with the observed pattern of results and can furthermore be successfully integrated with other theoretical accounts in the field of the illusory truth effect. According to the referential theory proposed by Unkelbach and Rom (2017), a higher number of coherent mental activations during the processing of a repeated (vs. new) statement drives the repetition-based truth effect. In the context of this theory, repetition-induced faster information accumulation in favor of the judgment "true," along with an accompanying feeling of cognitive ease (fluency), can be viewed as the result of a greater number of coherent mental activations while evaluating repeated information.

It is worth highlighting that the data presented in this study were gathered from a sample of primarily young German university students. Consequently, further investigation is needed to ascertain the extent to which these effects can be generalized to other populations. Nevertheless, we assume that our results, reflecting fundamental cognitive mechanisms underlying the truth effect, are probably not limited to the group investigated in our study.

From a modeling perspective, it is important to acknowledge that the interpretation of DDM parameters in terms of specific cognitive processes is necessarily theory-driven. Although parameters such as drift rate, non-decision time, and starting point bias are commonly associated

[4] In Experiment 2, we furthermore reminded the participants directly before the judgment phase that true and false statements were presented in the first part of the experiment.

with information accumulation, encoding processes, and response bias, respectively, these mappings are not process-pure and may reflect multiple underlying mechanisms. Accordingly, our conclusions should be understood as theoretically guided interpretations of parameter changes rather than direct evidence for specific cognitive processes.

Conclusion

With the present research, we demonstrated that repetition influences the evaluation of truth through multiple cognitive mechanisms, and that these mechanisms vary as a function of retention interval length. In Experiment 1, repetition led to increased drift rates, reduced non-decision times, and a response bias toward "true," indicating that both conceptual and perceptual fluency contributed to the illusory truth effect shortly after exposure. These effects diminished after a longer retention interval. In Experiment 2, when modeling only the response times for one-word answers, repetition no longer affected drift rates or non-decision times, but a response bias toward "true" responses persisted. Additionally, reading times for repeated questions were reduced, indicating facilitated perceptual encoding. These findings suggest that repetition enhances perceptual fluency during question reading, but that its influence on conceptual fluency may already manifest prior to the onset of the one-word answers. To test this interpretation, we conducted an additional analysis using total trial response times (i.e., reading times + response times). This analysis replicated the repetition-induced increase in drift rates, supporting the idea that repetition contributes to faster information accumulation.

Taken together, our results suggest that the repetition-based truth effect arises from a combination of enhanced perceptual encoding, a subsequent bias toward affirming repeated statements, as well as increased speed of information accumulation supporting the truthfulness of a statement. Our findings furthermore highlight that these components contribute at different stages of processing, reinforcing the value of temporally sensitive modeling approaches.

Transparency

Funding: This research was financially supported by a research grant from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK 2277 "Statistical Modeling in Psychology".

Author Contributions:

Annika Stump: Conceptualization; Data curation; Formal analyses; Investigation; Methodology; Project administration; Software; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Lukas Schumacher: Data curation; Formal analyses; Methodology; Visualization; Writing – original draft; Writing – review & editing.

Andreas Voss: Conceptualization; Data curation; Methodology; Writing – review & editing.

Karl Christoph Klauer: Conceptualization; Methodology; Writing – review & editing.

Conflict of interest: All authors declare that they have no conflict of interest.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology, 27*(6), 576–605. https://doi.org/10.1016/0022-1031(91)90026-3

Betancourt, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv:1701.02434 [stat]. https://doi.org/10.48550/arXiv.1701.02434

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, *71*, 117–120. https://doi.org/10.1016/j.euroecorev.2014.07.003

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Garcia-Marques, T., Silva, R. R., & Mello, J. (2016). Judging the truth-value of a statement in and out of a deep processing context. *Social Cognition*, *34*(1), 40–54. https://doi.org/10.1521/soco.2016.34.1.40

Garcia-Marques, T., Silva, R. R., Reber, R., & Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology*, *56*, 126–129. https://doi.org/10.1016/j.jesp.2014.09.015

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Henderson, E. L., Simons, D. J., & Barr, D. J. (2021). The Trajectory of Truth: A Longitudinal Study of the Illusory Truth Effect. *Journal of Cognition*, *4*(1), 1–23. https://doi.org/10.5334/joc.161

Nadarevic, L. (2010). *Die Wahrheitsillusion* [The illusory truth effect]. Berlin: Verlag Dr. Köster.

Nadarevic, L., & Aßfalg, A. (2017). Unveiling the truth: warnings reduce the repetition-based truth effect. *Psychological Research*, *81*, 814–826. https://doi.org/10.1007/s00426-016-0777-y

Parks, C. M., & Toth, J. P. (2006). Fluency, Familiarity, Aging, and the Illusion of Truth. *Aging, Neuropsychology, and Cognition*, *13*(2), 225–253. https://doi.org/10.1080/138255890968691

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*(3), 338–342. https://doi.org/10.1006/ccog.1999.0386

Schnuerch, M., Nadarevic, L. & Rouder, J.N. (2021). The truth revisited: Bayesian analysis of individual differences in the truth effect. *Psychonomic Bulletin & Review*, *28*, 750–765. https://doi.org/10.3758/s13423-020-01814-8

Silva, R. R., Garcia-Marques, T., & Reber, R. (2017). The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth. *Consciousness and Cognition*, *51*, 53-67. http://dx.doi.org/10.1016/j.concog.2017.02.016

Stan Development Team. 2024. Stan Reference Manual, 2.36.0. https://mc-stan.org

Stump, A., Rummel, J., & Voss, A. (2022). Is it all about the feeling? Affective and (meta-) cognitive mechanisms underlying the truth effect. *Psychological Research*, *86*, 12–36. https://doi.org/10.1007/s00426-020-01459-1

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44–62. https://doi.org/10.1037/a0021765

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2020). Loo: Efficient leave-one-out cross-validation and waic for bayesian models [R package version 2.4.1]. https://mc-stan.org/loo/

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & cognition*, *32*, 1206–1220. https://doi.org/10.3758/BF03196893

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Table 1

DDM group-level parameter posterior distribution summary statistics for Experiment 1

	Repetition status			
	Repeated		New	
	Factual truth		Factual truth	
	False	True	False	True
	Session 1			
Drift rate	0.19 [0.09, 0.29]	0.10 [-0.02, 0.20]	-0.02 [-0.13, 0.08]	0.00 [-0.10, 0.11]
Threshold separation	2.33 [2.17, 2.50]	2.17 [1.99, 2.35]	2.09 [1.92, 2.26]	2.03 [1.87, 2.19]
Starting point bias	0.66 [0.62, 0.71]	0.69 [0.64, 0.74]	0.57 [0.52, 0.61]	0.53 [0.48 0.58]
Non-decision time	1.53 [1.43, 1.64]	1.51 [1.40, 1.61]	1.75 [1.64, 1.86]	1.74 [1.63, 1.85]
Non-decision time variability	1.61 [1.13, 2.05]	1.84 [1.36, 2.33]	2.25 [1.80, 2.70]	2.22 [1.78, 2.66]
	Session 2			
Drift rate	0.13 [0.03, 0.23]	0.12 [0.02, 0.21]	0.15 [0.06, 0.24]	0.01 [-0.09, 0.10]
Threshold separation	2.14 [2.96, 2.31]	2.24 [2.08, 2.41]	2.24 [2.06, 2.41]	2.14 [1.96, 2.32]
Starting point bias	0.58 [0.54, 0.62]	0.57 [0.53, 0.61]	0.52 [0.48, 0.56]	0.54 [0.50 0.58]
Non-decision time	1.51 [1.41, 1.61]	1.47 [1.37, 1.58]	1.59 [1.49, 1.70]	1.59 [1.49, 1.70]
Non-decision time variability	1.94 [1.43, 2.42]	1.89 [1.43 2.34]	1.86 [1.35, 2.35]	2.11 [1.63, 2.60]

Note. The values represent the median and lower and upper boundary of the 95% credibility interval for each parameter, condition and session separately.

Table 2

DDM group-level parameter posterior distribution summary statistics for Experiment 2 (based on one-word answer reaction times).

	Repetition status			
	Repeated		New	
	Factual truth		Factual truth	
	False	True	False	True
Drift rate	-0.03 [-0.09, 0.04]	0.03 [-0.03, 0.10]	-0.04 [-0.10, 0.03]	-0.03 [-0.10, 0.03]
Threshold separation	2.04 [1.92, 2.17]	2.09 [1.97, 2.22]	2.02 [1.90, 2.15]	2.08 [1.96, 2.21]
Starting point bias	0.66 [0.63, 0.68]	0.64 [0.61, 0.66]	0.59 [0.57, 0.62]	0.58 [0.55, 0.61]
Non-decision time	0.63 [0.58, 0.67]	0.66 [0.61, 0.70]	0.66 [0.61, 0.71]	0.65 [0.59, 0.70]
Non-decision time variability	0.40 [0.14, 0.64]	0.25 [-0.01, 0.50]	0.51 [0.24, 0.76]	0.58 [0.31, 0.83]

Note. The values represent the median and lower and upper boundary of the 95% credibility interval for each parameter, and condition separately.

Table 3

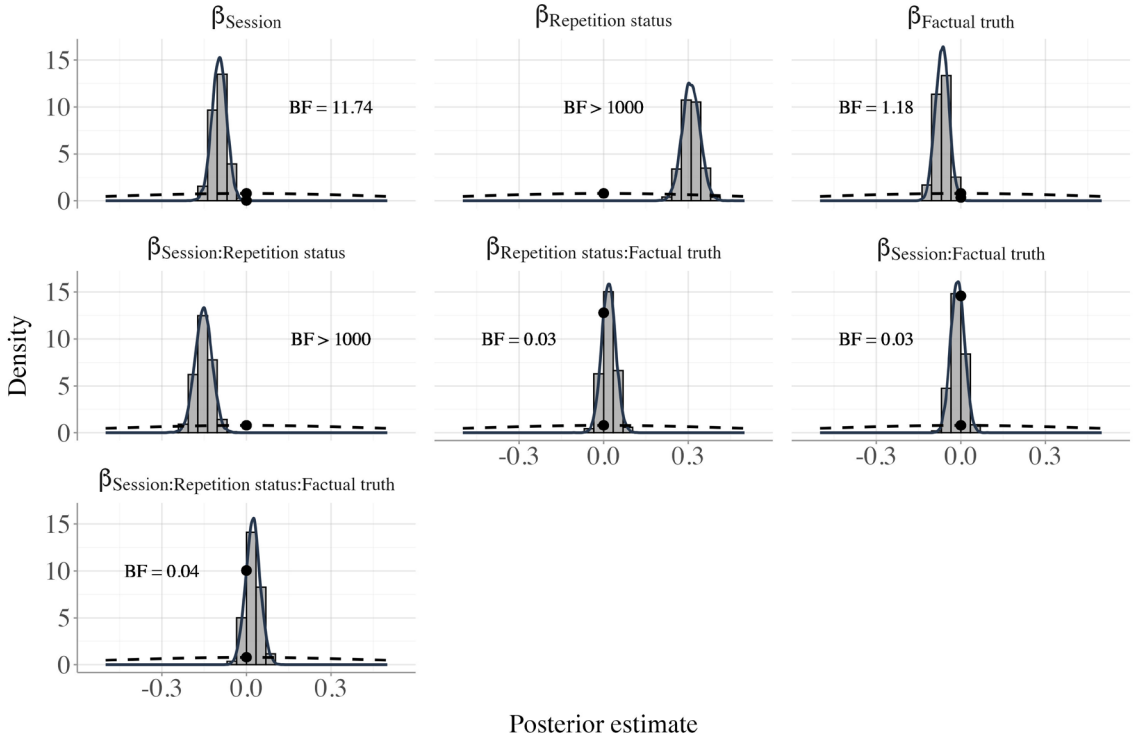
DDM group-level parameter posterior distribution summary statistics for Experiment 2 (based on total reaction times).

	Repetition status			
	Repeated		New	
	Factual truth		Factual truth	
	False	True	False	True
Drift rate	0.19 [0.12, 0.26]	0.23 [0.16, 0.29]	0.09 [0.02, 0.15]	0.09 [0.02, 0.16]
Threshold separation	2.69 [2.55, 2.83]	2.67 [2.53, 2.81]	2.66 [2.52, 2.80]	2.64 [2.51, 2.78]
Starting point bias	0.53 [0.50, 0.56]	0.52 [0.49, 0.55]	0.52 [0.49, 0.55]	0.51 [0.48, 0.54]
Non-decision time	0.75 [0.69, 0.80]	0.78 [0.72, 0.83]	0.81 [0.75, 0.86]	0.81 [0.75, 0.86]
Non-decision time variability	3.37 [3.09, 3.64]	3.34 [3.07, 3.62]	3.51 [3.25, 3.79]	3.66 [3.37, 3.94]

Note. The values represent the median and lower and upper boundary of the 95% credibility interval for each parameter, and condition separately.

Figure 1

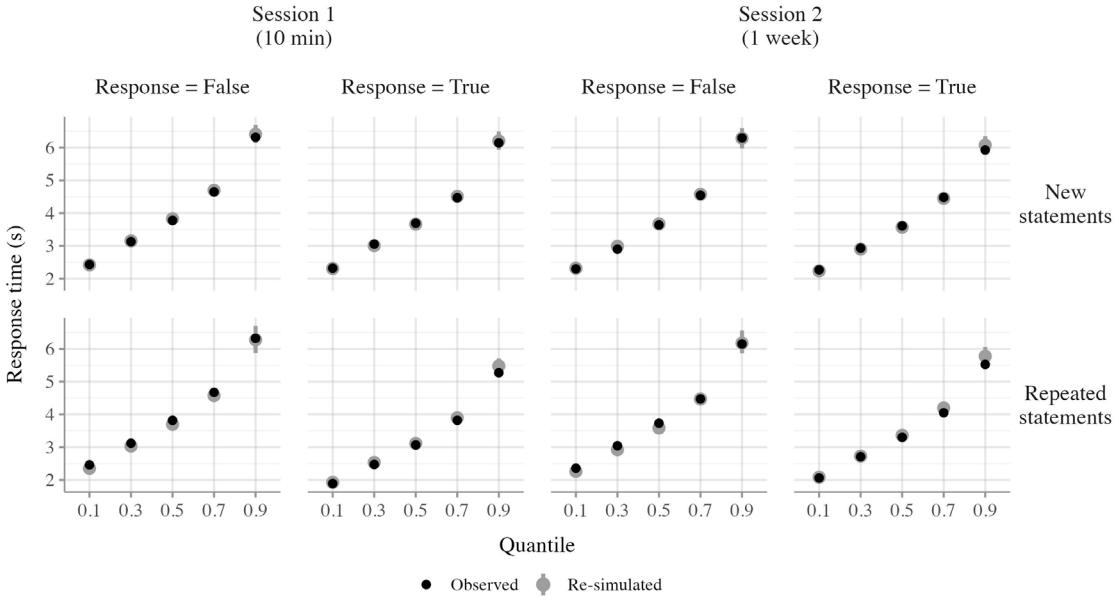
Prior and Posterior Distributions for the Group Level Parameters of the Hierarchical Logistic Regression Model for Experiment 1.



Note. Prior Parameter Distributions are given by dashed lines; The posterior distributions are plotted as histograms and kernel densities. BF = Bayes Factor.

Figure 2

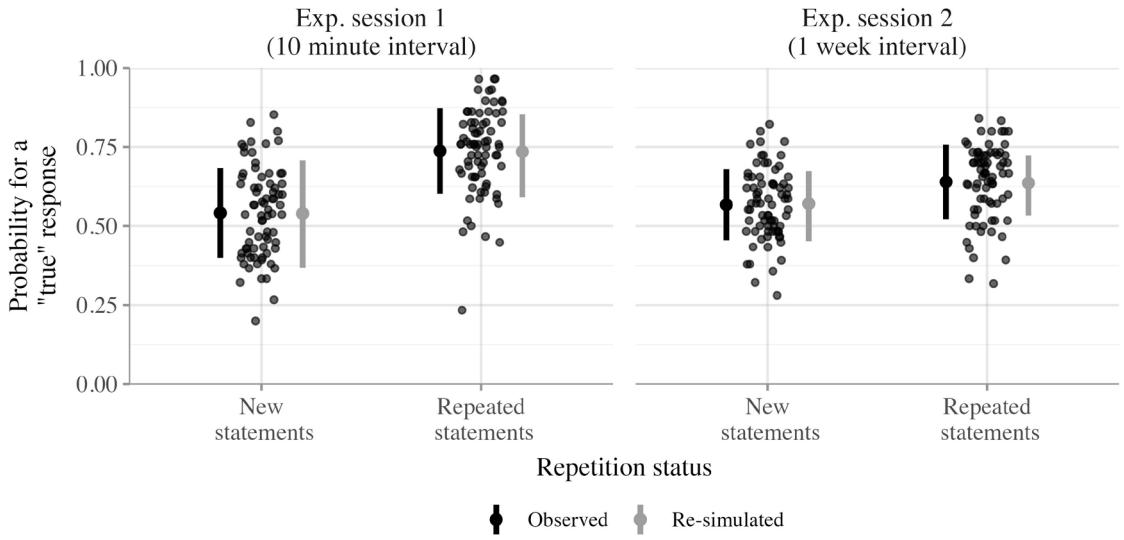
Simulations from the posterior predictive distribution for Experiment 1: response time quantiles



Note. Observed and re-simulated response times quantiles (0.1; 0.3; 0.5; 0.7; 0.9) for both sessions of Experiment 1, possible responses (false; true), and statement repetition status separately.

Figure 3

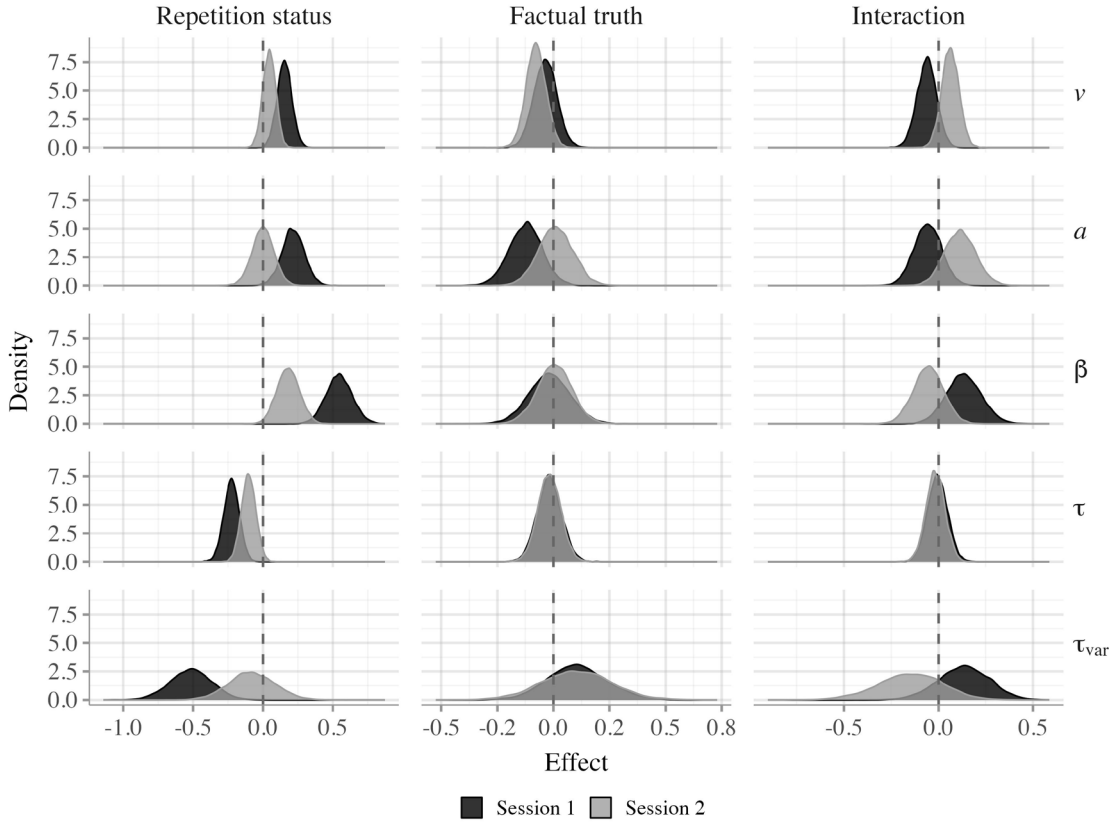
Simulations from the posterior predictive distribution for Experiment 1: response probabilities



Note. Observed and re-simulated probabilities for a "true" response, presented separately for each experimental session and statement repetition status. Individual response probabilities are shown by dots. The black dots and ranges represent the empirical grand mean and standard deviation, respectively. The gray dots and ranges illustrate grand mean and its corresponding 95% credibility interval re-simulated with our model.

Figure 4

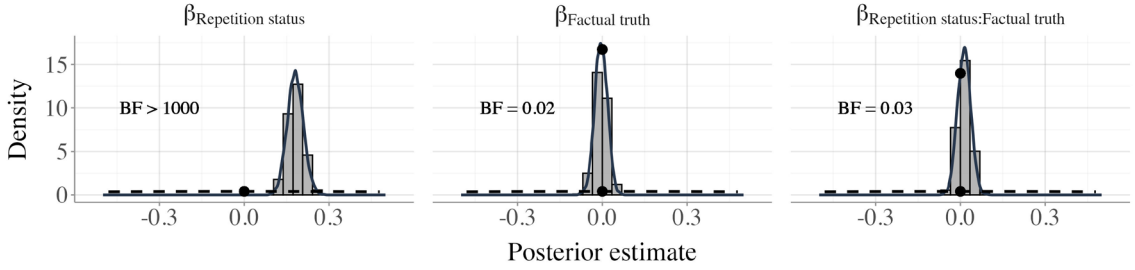
Posterior distributions of regression weights on DDM parameters from Experiment 1.



Note. Posterior distributions are shown for each DDM parameter, drift rate (v), boundary separation (a), starting point bias (β), non-decision time (τ), and non-decision time variability (τ_{var}), across both sessions of Experiment 1. Columns represent effects (statement truth, repetition, and their interaction), with Session 1 in black and Session 2 in gray.

Figure 5

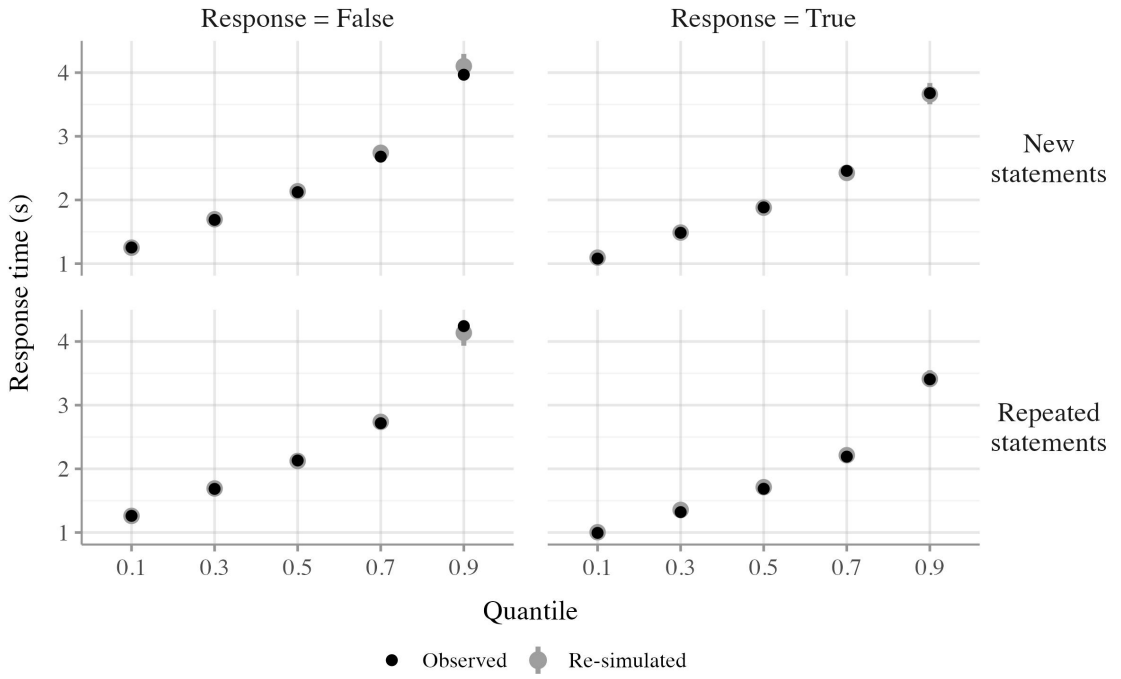
Prior and posterior distributions for the group-level parameters of the hierarchical logistic regression model for Experiment 2 (based on one-word answer reaction times).



Note. Prior Parameter Distributions are given by dashed lines; The posterior distributions are plotted as histograms and kernel densities. BF = Bayes Factor.

Figure 6

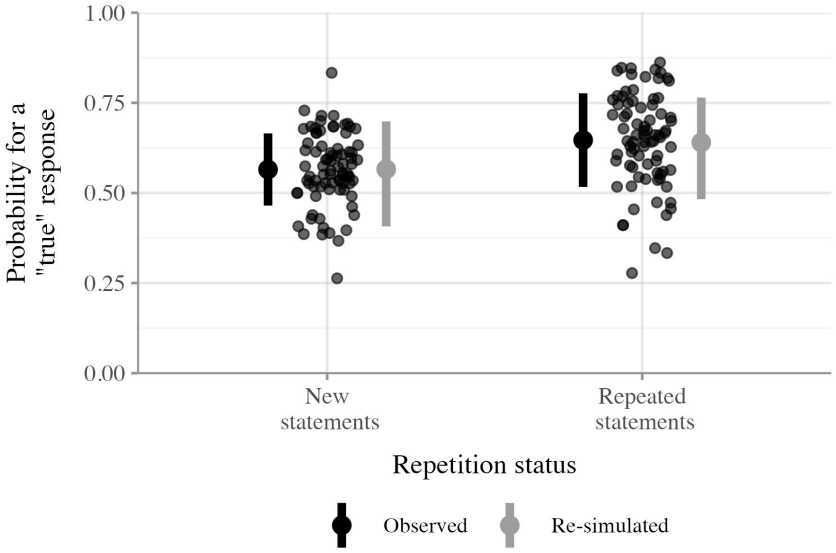
Simulations from the posterior predictive distribution for Experiment 2: response time quantiles (based on one-word answer reaction times).



Note. Observed and re-simulated response times quantiles (0.1; 0.3; 0.5; 0.7; 0.9) for each statement repetition status and possible response (false; true) separately.

Figure 7

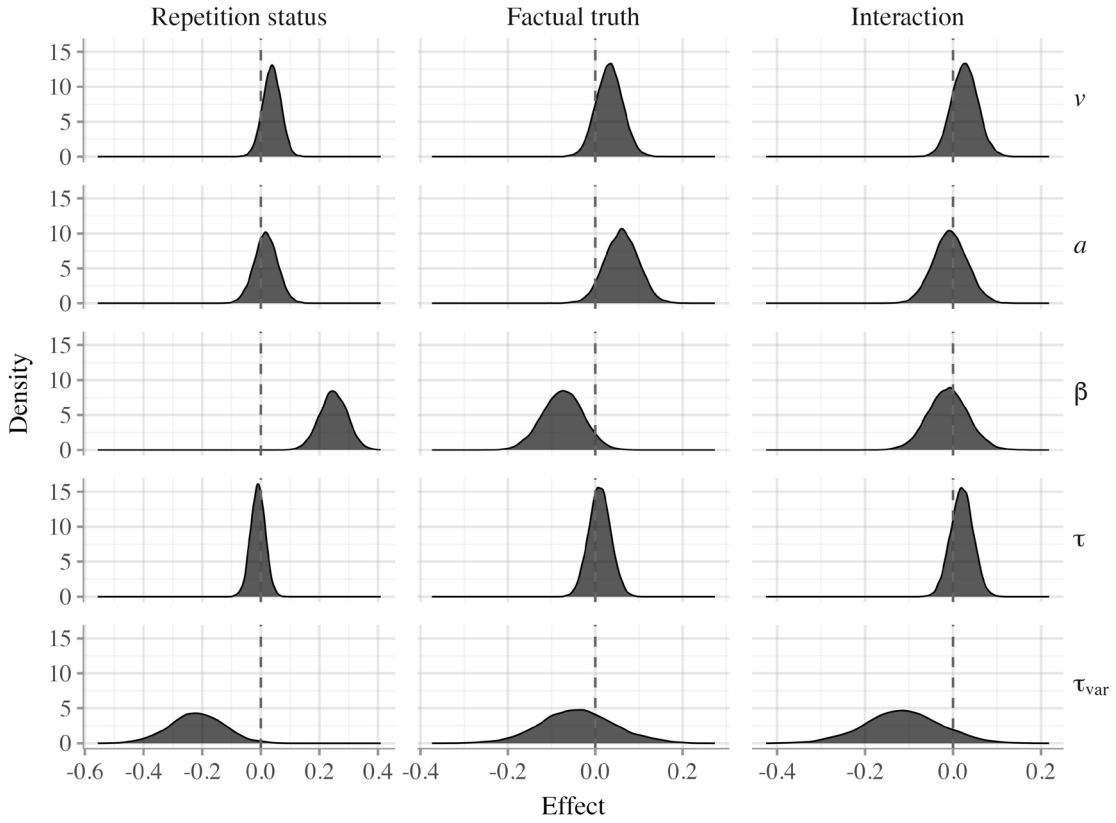
Simulations from the posterior predictive distribution for Experiment 2: response probabilities (based on one-word answer reaction times).



Note. Observed and re-simulated probabilities for a “true” response, presented separately for each statement repetition status. Individual response probabilities are shown by dots. The black dots and ranges represent the empirical grand mean and standard deviation, respectively. The gray dots and ranges illustrate grand mean and its corresponding 95% credibility interval re-simulated with our model.

Figure 8

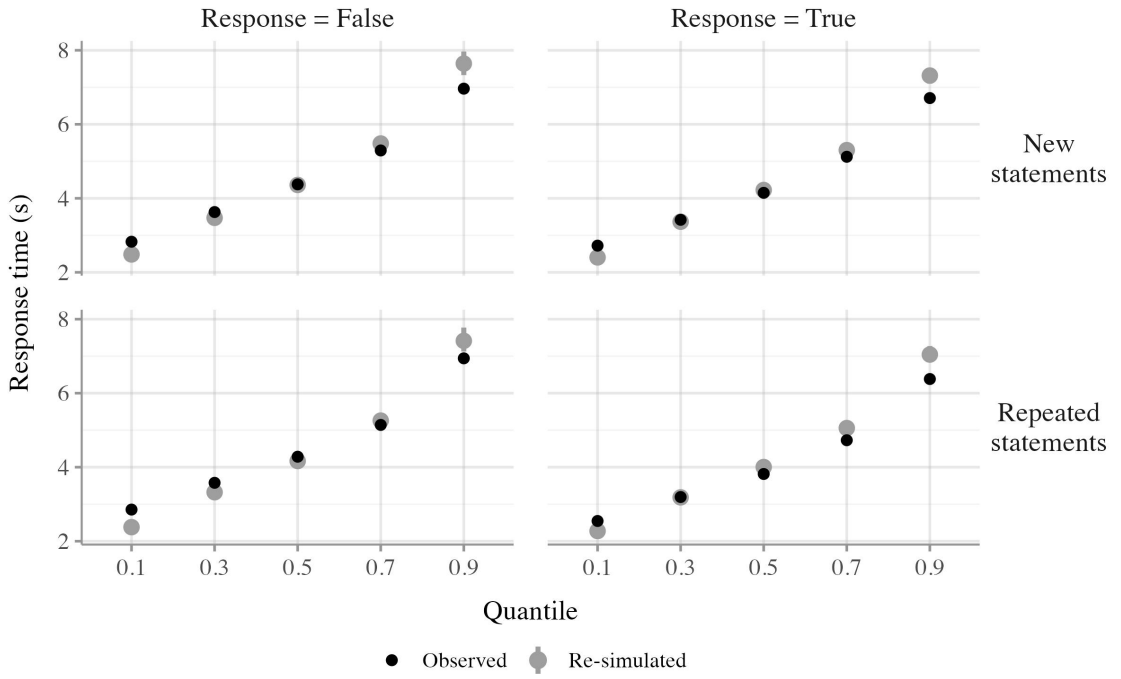
Posterior distributions of regression weights on DDM parameters from Experiment 2 (based on one-word answer reaction times).



Note. Posterior distributions are shown for each DDM parameter, drift rate (v), boundary separation (a), starting point bias (β), non-decision time (τ), and non-decision time variability (τ_{var}). Columns represent effects (statement truth, repetition, and their interaction).

Figure 9

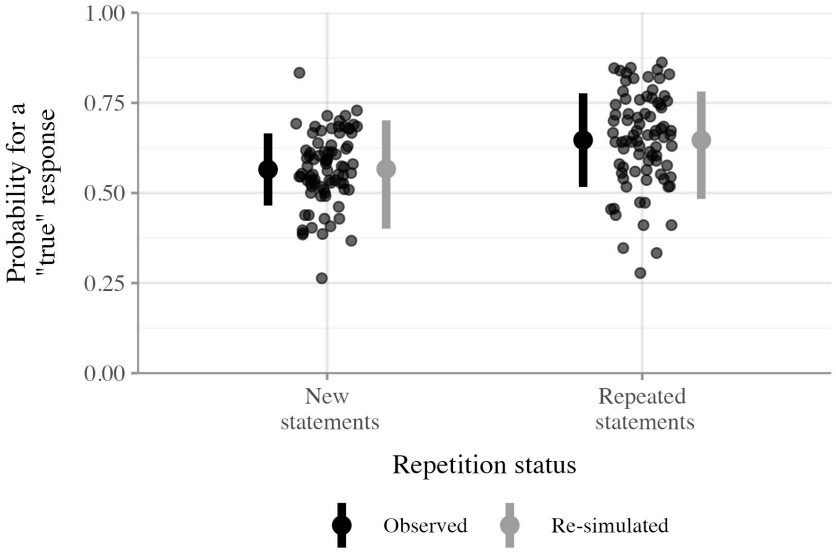
Simulations from the posterior predictive distribution for Experiment 2: response time quantiles (based on total reaction times).



Note. Observed and re-simulated response times quantiles (0.1; 0.3; 0.5; 0.7; 0.9) for each statement repetition status and possible response (false; true) separately.

Figure 10

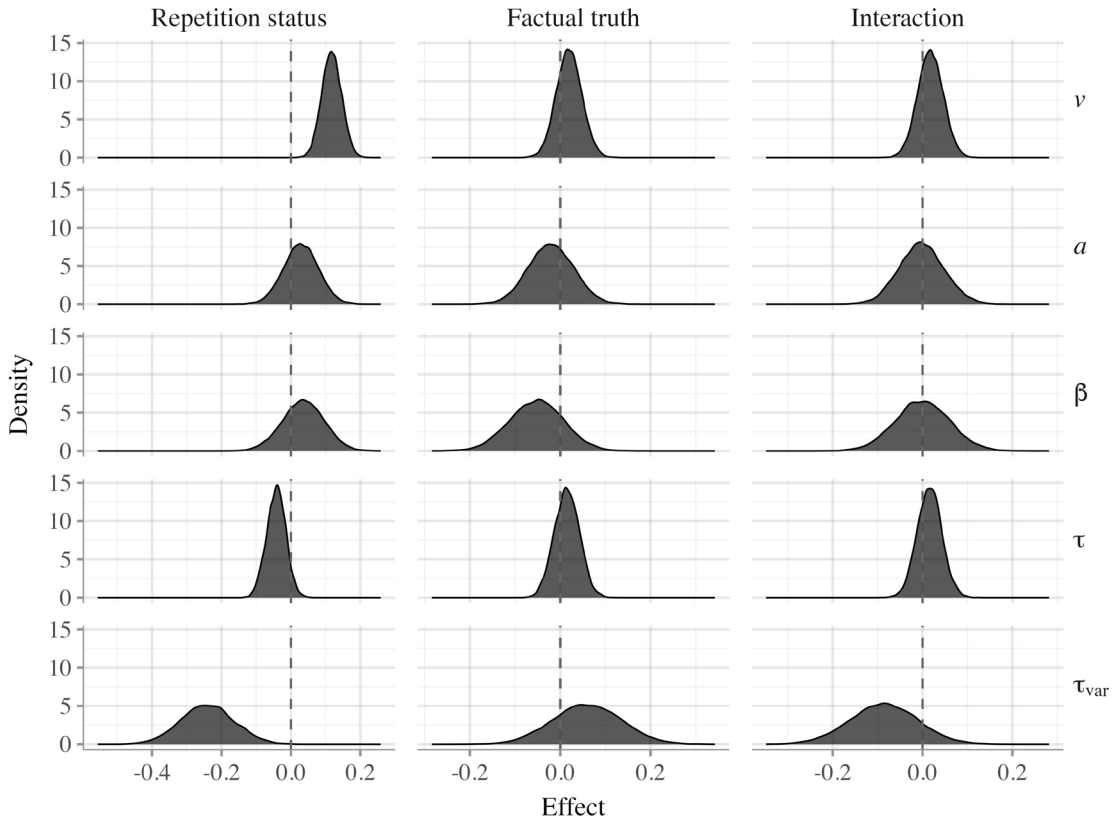
Simulations from the posterior predictive distribution for Experiment 2: response probabilities (based on total reaction times).



Note. Observed and re-simulated probabilities for a “true” response, presented separately for each statement repetition status. Individual response probabilities are shown by dots. The black dots and ranges represent the empirical grand mean and standard deviation, respectively. The gray dots and ranges illustrate grand mean and its corresponding 95% credibility interval re-simulated with our model.

Figure 11

Posterior distributions of regression weights on DDM parameters from Experiment 2 (based on total reaction times).



Note. Posterior distributions are shown for each DDM parameter, drift rate (v), boundary separation (a), starting point bias (β), non-decision time (τ), and non-decision time variability (τ_{var}). Columns represent effects (statement truth, repetition, and their interaction).

Appendix A

Prior distributions for the group-level parameters used in the hierarchical DDM:

$$v \sim \text{Normal}(0, 1)$$

$$\text{Softplus}(a) \sim \text{Normal}(2, 1)$$

$$\tau \sim \text{Normal}(1, 1)$$

$$\text{Inverse logit}(\beta) \sim \text{Normal}(0, 1)$$

$$\text{Softplus}(\tau_{\text{var}}) \sim \text{Normal}(1, 1)$$

$$\text{Softplus}(\sigma_{\text{group},i}) \sim \text{Normal}(0, 1).$$

Prior distribution for the effects on the group-level DDM parameters:

$$\gamma_i \sim \text{Normal}(0.\, 0.5).$$

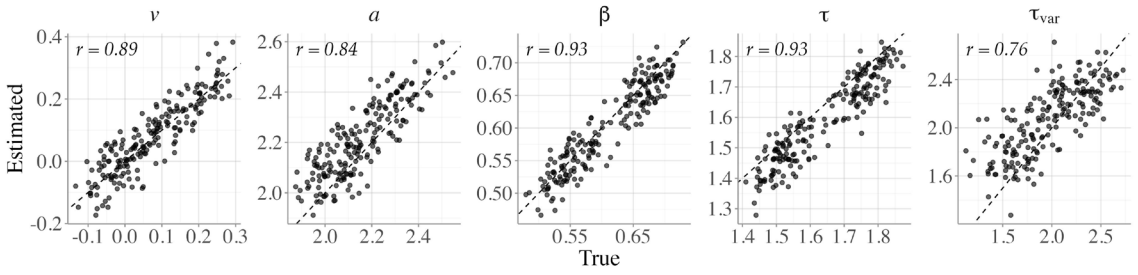
Note. $\sigma_{\text{group},i}$ refers to the standard deviation of any group-level parameter distribution i.

Appendix B

To ensure that our DDM's parameters are identifiable given the number of trials in our experiment, we conducted a parameter recovery study. Specifically, we randomly drew 50 individual level parameter sets from the posterior distributions inferred from Session 1 of Experiment 1. Using these parameter sets, we simulated new data sets under the same conditions as Session 1 of Experiment 1, that is, with 75 subjects and a total of 4290 trials (after exclusion). We then fit our model to each synthetic data sets and compared the true data-generating parameters with the inferred posterior medians. Figure B1 displays this comparison across all five DDM group-level parameters: drift rate, threshold separation, starting point bias, non-decision time, and non-decision time inter-trial variability. Each parameter was estimated separately for the four experimental conditions, yielding 20 group-level mean parameters per synthetic dataset. Given our balanced design, the four condition-specific estimates for each of the five DDM parameters are shown in the same panel. We observed excellent recovery of all group-level parameters, with correlations exceeding 0.76 in all cases. We also assessed parameter recovery at the individual level and found similarly strong performance, with all correlations above 0.72 (see Figure B2).

Figure B1

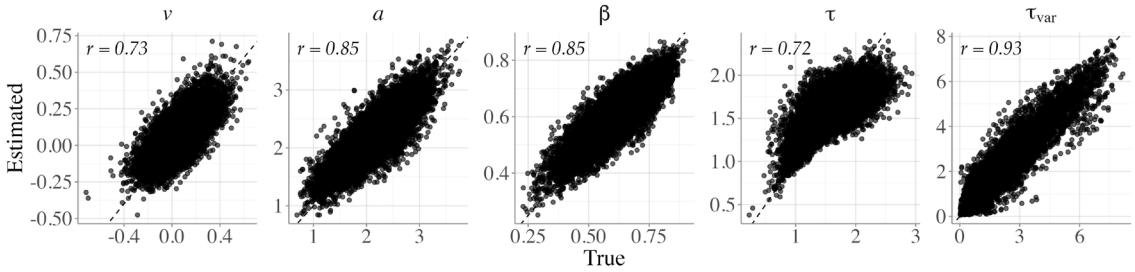
Parameter recovery of group-level DDM parameters.



Note. True data-generating parameters on the x-axis and estimated median posteriors on the y-axis for each group-level DDM parameter separately.

Figure B2

Parameter recovery of individual-level DDM parameters.



Note. True data-generating parameters on the x-axis and estimated median posteriors on the y-axis for each individual-level DDM parameter separately.

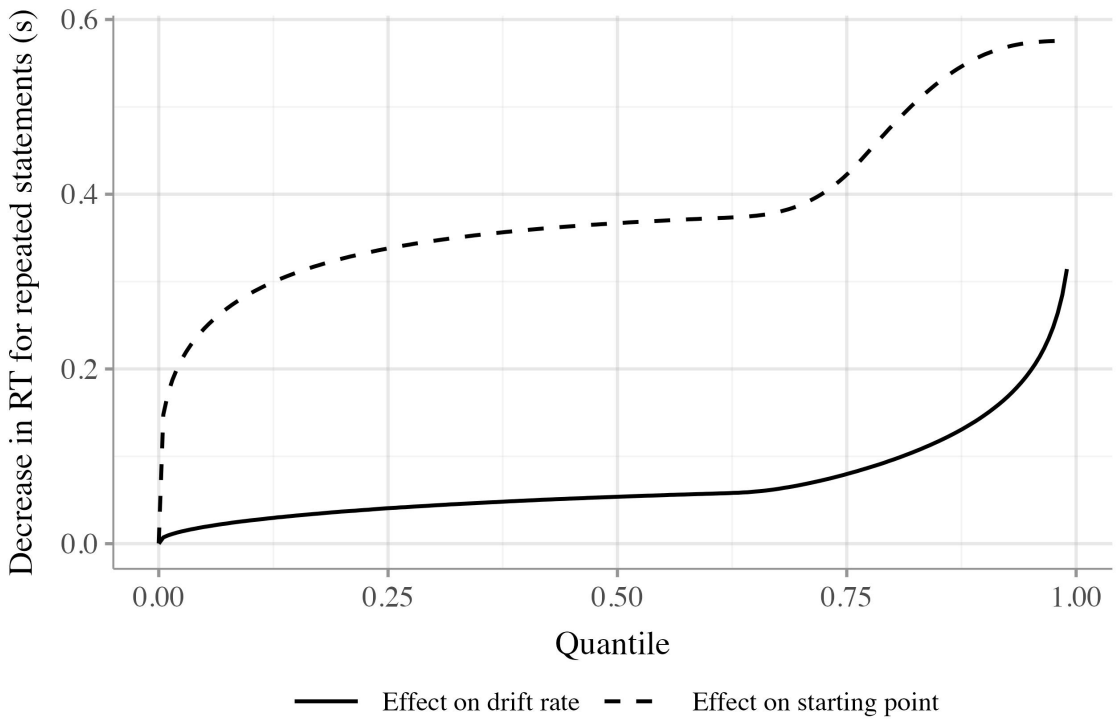
Appendix C

A challenge in our model concerns distinguishing whether repetition-related changes in the frequency and speed of responses are driven primarily by changes in drift rate or by changes in starting point. This distinction relies on differences in the predicted shapes of the RT distributions produced by the DDM. In particular, changes in starting point are expected to produce an approximately uniform shift of the RT distribution, affecting both the leading edge and the upper tail to a similar extent. In contrast, changes in drift rate should leave the leading edge largely unchanged while producing disproportionately large effects in the upper tail of the distribution. In our model, this inference is complicated by the fact that non-decision time and its inter-trial variability are estimated simultaneously and, in some cases, take on comparatively large values. To directly assess whether drift rate and starting point effects are distinguishable under these conditions, we conducted an additional qualitative analysis of model-predicted RT quantiles.

To this end, we predicted RT quantiles for "true" responses using parameter values close to the best-fitting group-level medians from Session 1 of Experiment 1 (true, new statements). Using these parameters, we generated two alternative sets of predictions for repeated statements. In the first, the repetition effect was implemented exclusively as a shift in starting point. In the second, repetition effect resulted exclusively from a change in drift rate. In both cases, the magnitude of the parameter change was chosen such that the predicted proportion of "true" responses increased from 0.53 for new statements to 0.75 for repeated statements, closely matching the empirical data. Thus, both mechanisms produced an identical change in response proportions and differed only in how they affected the RT distributions. For each mechanism, we computed the RT quantile function for true responses and plotted the difference between the predicted quantiles.

Figure C1 shows the resulting quantile-difference functions for the two mechanisms. The two predicted patterns are qualitatively different. When the repetition effect is produced by a change in starting point, the reduction in RT is already substantial at low quantiles, and thus, has a relatively large effect on the leading edge of the distribution. In contrast, when the repetition effect is implemented via a change in drift rate, the reduction in RT is minimal at early quantiles and becomes pronounced only in the upper tail of the distribution. As a consequence, the predicted effect on the median RT is small, despite an equally large increase in the proportion of "true" responses. Importantly, these differences emerge even under the large estimated variability in non-decision time used in this analysis. This demonstrates that, in principle, the model can use information of the RT distribution to discriminate between drift-rate and starting-point effect, even in the presence of substantial non-decision time variability.

Figure C1



Note. Difference between predicted response time for repeated and new statements for "true" responses as a function of quantile. Repetition effects were implemented either as a change in starting point (dashed line) or drift rate (solid line), calibrated to produce the same increase in "true" responses.