Modeling the Link between Confidence and Truth Judgments in the Repetition-Based Truth Effect

Semih C. Aktepe (semih.aktepe@uni-marburg.de)[1] **& Daniel W. Heck (daniel.heck@uni-marburg.de)**[1]
[1]Department of Psychology, Marburg University, Gutenbergstraße 18, 35037 Marburg, Germany

Abstract

Repeated exposure to information increases its perceived truth, a phenomenon known as the illusory truth effect. Beyond influencing truth judgments, repetition also increases individuals' confidence judgments. Existing theoretical models generally assume that confidence is lowest when statements are maximally ambiguous and highest when their truth status is known. However, it is currently unclear how the relationship between perceived truth and confidence can be formally modeled and whether such models are consistent with empirical data. In this study, we developed a Bayesian model to characterize the link between confidence and truth judgments, with several model variants capturing alternative assumptions about this relationship. The model also accounts for the effect of repetition on both variables. Fitting these models to two datasets provided evidence for an asymmetric, V-shaped relationship between truth and confidence. Specifically, individuals showed higher confidence when judging highly true statements than when judging highly false ones, with confidence reaching a minimum for ambiguous statements.

Keywords: confidence; truth effect; repetition; psychometrics; Bayesian modeling

Introduction

The illusory truth effect is a cognitive phenomenon where individuals tend to judge repeated information as more likely to be true than novel information (Hasher et al., 1977). This effect is widely attributed to processing fluency, the subjective ease with which repeated information is processed, which is assumed to be misattributed to the perceived truth of a statement (Arkes et al., 1989; Unkelbach, 2007). While the truth effect is most pronounced for ambiguous statements where prior knowledge is absent (Dechêne et al., 2010; Newman et al., 2020), it also occurs for very implausible statements (Aktepe & Heck, 2025; Fazio et al., 2019).

Repetition-induced fluency has been shown to increase individuals' confidence regarding various tasks and responses (Fiechter & Kornell, 2021; Foster et al., 2012). Specifically, repeated exposure to a stimulus has been found to enhance confidence in truth judgments (Stump et al., 2022, 2024). These findings suggest that individuals not only judge repeated statements as more true than novel statements, but also report higher confidence in their judgments. Overall, repetition-induced fluency may boost both truth judgments and subjective confidence ratings simultaneously. However, open questions are how the relationship between truth judgments and confidence is best modeled, how exactly repetition

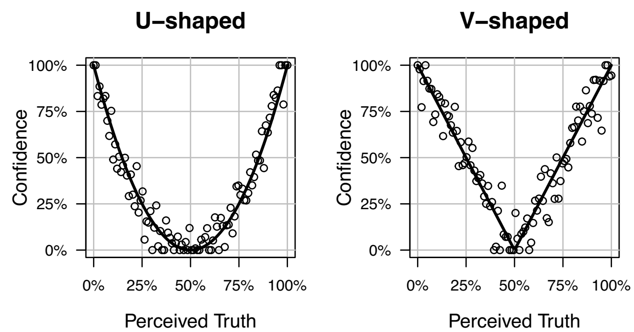


Figure 1: Theoretical link between perceived truth and confidence

affects the two judgment dimensions jointly, and whether such models are consistent with data.

Previous studies generally show that confidence is lowest at the point of maximum ambiguity of the presented statements, and highest when their actual truth is known (Bruine de Bruin et al., 2000). Such response-confidence relationships often show a quadratic relationship (Koriat, 2018; Lebreton et al., 2015). When adapted to truth judgments, for the statements whose actual truth status is known (i.e., easier statements), people are expected to be more confident when judging the truth of such statements. However, when a statement is ambiguous (i.e., difficult to know whether it is true or false), confidence judgments of individuals decrease (Stump et al., 2024). Theoretically, this truth-confidence link can be modeled either with a U-shaped or a V-shaped function, depending on how steeply confidence increases when people are more knowledgeable about the truth status of the presented statements compared to the ambiguous ones. As shown in Figure 1, in the U-shaped model, for statements near ambiguity, the growth in confidence is not steep, yet for those statements whose truth is known (near 0% or 100%), there is a faster quadratic confidence growth. In the V-shaped model, confidence grows linearly on the probability scale from ambiguity to known truth or falseness (regarding issues of scale transformations, see Aktepe & Heck, 2025).

Confidence-response relationships can generally be asymmetrical (Yi & La, 2003). Regarding truth judgments, Stump et al. (2024) found that participants are more confident when judging a statement as true than when judging it as false,

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

irrespective of its factual validity. This pattern suggests a potential asymmetry in confidence, such that judgments of high perceived truth are accompanied by greater confidence than those of high perceived falseness. Theoretically, such an asymmetry could manifest an unequal maxima of the confidence peaks for statements perceived as highly true and false, or a shift in the point of minimum confidence away from the objective midpoint (50% truth).

A final question concerns whether the effects of repetition on both the truth and the confidence dimension are driven by the same underlying mechanism. If a single fluency signal drives both effects, one would expect a strong (possibly perfect) correlation between the magnitude of the truth effect and the confidence increase at the item level. If these effects are uncorrelated, it would suggest that repetition influences truth and confidence through distinct pathways.

Present Research

This study aims to adjudicate between different functional shapes and investigate the coupling of item-level repetition effects across both the truth and the confidence dimensions. Classical statistical methods are ill-equipped to handle this problem as they cannot easily model the intercept of confidence as a non-linear function of the latent intercept of truth while simultaneously estimating the correlation between their respective slopes. To address this, we developed several variants of a Bayesian model formalizing these assumptions. The competing model versions allow us to assess the evidence for the U-shaped against the V-shaped mapping, to test for asymmetrical relationships, and to estimate the correlation between the repetition effects for confidence and truth across items. We fitted these models to reanalyze datasets from Stump et al. (2024) and Nikolaev (2025), thereby providing a comprehensive cross-study validation of the truth-confidence link.

Bayesian Modeling

Our Bayesian modeling approach formalizes the relationship between overall truth perception of a statement and its overall confidence, thereby allowing us to test how repetition affects both dimensions. At their core, all model versions share a common generative structure and account for judgments averaged across participants. The intercept for confidence (α_{conf}) is a direct function of the intercept for truth (α_{truth}), reflecting the theory that confidence is anchored in the perceived truth of a statement. The model incorporates a fundamental parameter, γ_{min}, which represents the minimum confidence at the point of maximum ambiguity regarding the truth of a statement. Furthermore, the model employs statement-level random slopes for the effect of repetition on both truth (ν_{truth}) and confidence (ν_{conf}). By estimating the correlation between these slopes across items, the model captures the variability in sensitivity of how strongly a change in perceived truth due to repetition is linked to a corresponding change in confidence.

We derived several model variants to answer our research questions. These models differ in two aspects. First, they assume that the functional relationship between perceived truth and confidence is either U-shaped or V-shaped. While both model versions assume that confidence is lowest at a specific point on the truth axis and increases toward the extremes (0% and 100% perceived truth), they differ in the sharpness of this transition. The U-shaped variants utilize a quadratic function, resulting in a smoother, curved transition that implies a slower increase in confidence near the point of maximum ambiguity, and a more rapid increase as truth approaches the extremes. In contrast, the V-shaped variants assume a linear relationship on the probability scale where confidence increases at a constant rate as perceived truth approaches each of the two extremes.

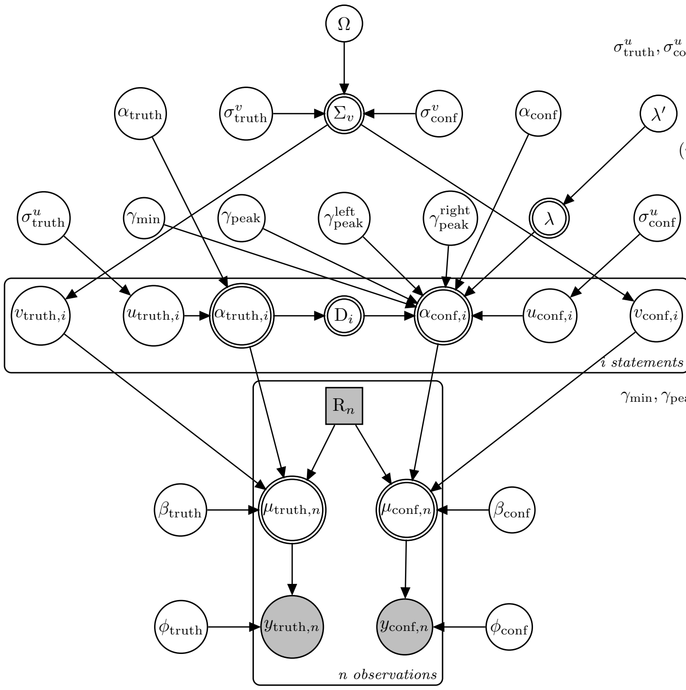
Second, the model variants differ in their assumptions regarding the assumed symmetry across the whole spectrum of perceived truth. Symmetry in this context is characterized by two aspects: first, the relative magnitudes of the confidence peaks at the extremes of perceived truth and falseness (i.e., statements that are considered to be true by all or none of the participants), and second, the location of the point of maximum ambiguity regarding perceived truth, where confidence reaches its minimum. To investigate these two aspects, we derived three more model versions.

Model 1: Model 1 assumes a symmetrical relationship where confidence levels at the extremes of perceived falseness and truth are identical, governed by a single γ_{peak} parameter. In its V-shaped version, the perceived truth of a statement i is mapped onto the corresponding confidence value with the function, $\alpha_{\text{conf},i} = \text{logit}((\gamma_{\text{peak}} - \gamma_{\text{min}}) \cdot \text{D}_i + \gamma_{\text{min}})$, where $\text{D}_i = 2 \times |\text{logit}^{-1}(\alpha_{\text{truth},i}) - 0.5|$. In the U-shaped version, $\text{D}_i = (2 \times \text{logit}^{-1}(\alpha_{\text{truth},i}) - 1)^2$, creating a parabola centered at 50% of perceived truth.

Model 2: Model 2 relaxes the symmetry assumption, allowing the maximum confidence for low truth statements to differ from that for high truth statements via the $\gamma_{\text{peak}}^{\text{left}}$ and $\gamma_{\text{peak}}^{\text{right}}$ parameters, respectively. In the V-shaped variant, for low truth statements (< 50%), $\text{D}_i = 1 - 2 \times \text{logit}^{-1}(\alpha_{\text{truth},i})$, while $\text{D}_i = 2 \times \text{logit}^{-1}(\alpha_{\text{truth},i}) - 1$ for high truth statements. In the U-shaped variant, the D_i term is squared.

Model 3: Model 3 introduces the most flexible structure by including a unique parameter λ, which estimates the location of the minimum confidence value on the truth axis rather than fixing it at 50% truth. In the V-shaped model, confidence increases linearly from λ towards the boundaries. Therefore, for statements with perceived truth lower than λ, $\text{D}_i = \frac{\lambda - \text{logit}^{-1}(\alpha_{\text{truth},i})}{\lambda}$, while for those with higher perceived truth than λ, $\text{D}_i = \frac{\lambda - \text{logit}^{-1}(\alpha_{\text{truth},i})}{1-\lambda}$. In the U-shaped version, D_i terms are squared similar to Model 2.

Null Model: Besides these three models, we constructed a null model to assess whether the theoretically assumed functional shapes are supported by the data or whether the relation-



$$
\begin{aligned}
\Omega &\sim \mathrm{LKJ}(4)\\
\sigma^u_{\mathrm{truth}}, \sigma^u_{\mathrm{conf}}, \sigma^v_{\mathrm{truth}}, \sigma^v_{\mathrm{conf}} &\sim \mathrm{Exponential}(1)\\
u_{\mathrm{truth},i} &\sim \mathrm{Normal}(0, \sigma^u_{\mathrm{truth}}) \qquad \forall i \in \{1, \ldots, S\}\\
u_{\mathrm{conf},i} &\sim \mathrm{Normal}(0, \sigma^u_{\mathrm{conf}}) \qquad \forall i \in \{1, \ldots, S\}\\
\Sigma_v &\leftarrow \mathrm{diag}(\sigma^v_{\mathrm{truth}})\Omega\mathrm{diag}(\sigma^v_{\mathrm{conf}})\\
(v_{\mathrm{truth},i}, v_{\mathrm{conf},i}) &\sim \mathrm{Normal}(0, \Sigma_v) \qquad \forall i \in \{1, \ldots, S\}\\
\lambda &\leftarrow \mathrm{logit}^{-1}(\lambda') \text{ where } \lambda' \sim \mathrm{Normal}(0, 0.5)\\
\alpha_{\mathrm{truth},i} &\leftarrow \alpha_{\mathrm{truth}} + u_{\mathrm{truth},i} \text{ where } \alpha_{\mathrm{truth}} \sim \mathrm{Normal}(0, 1.5)\\
p_i &\leftarrow \mathrm{logit}^{-1}(\alpha_{\mathrm{truth},i})\\
\mathrm{D}_i &\leftarrow \begin{cases} |2p_i - 1|, & \text{for } \mathcal{M}_{V1,V2}\\ (2p_i - 1)^2, & \text{for } \mathcal{M}_{U1,U2}\\ \dfrac{\lambda - p_i}{\lambda}\, \mathbf{1}_{\{p_i<\lambda\}} + \dfrac{\lambda - p_i}{1-\lambda}\, \mathbf{1}_{\{p_i \geq \lambda\}}, & \text{for } \mathcal{M}_{V3}\\ \left(\dfrac{\lambda - p_i}{\lambda}\, \mathbf{1}_{\{p_i<\lambda\}} + \dfrac{\lambda - p_i}{1-\lambda}\, \mathbf{1}_{\{p_i \geq \lambda\}}\right)^2, & \text{for } \mathcal{M}_{U3} \end{cases}\\
\gamma_{\mathrm{min}}, \gamma_{\mathrm{peak}}, \gamma^{\mathrm{left}}_{\mathrm{peak}}, \gamma^{\mathrm{right}}_{\mathrm{peak}} &\sim \mathrm{Beta}(5, 5)\\
\alpha_{\mathrm{conf},i} &\leftarrow \alpha_{\mathrm{conf}} + u_{\mathrm{conf},i} \text{ where } \alpha_{\mathrm{conf}} \sim \mathrm{Normal}(0, 1.5) \text{ for } \mathcal{M}_0\\
\alpha_{\mathrm{conf},i} &\leftarrow \begin{cases} \mathrm{logit}((\gamma_{\mathrm{peak}} - \gamma_{\mathrm{min}}) \cdot \mathrm{D}_i + \gamma_{\mathrm{min}}), & \text{for } \mathcal{M}_{U1,V1}\\ \mathrm{logit}((\gamma^{\mathrm{left}}_{\mathrm{peak}} - \gamma_{\mathrm{min}}) \cdot \mathrm{D}_i + \gamma_{\mathrm{min}}), & \text{if } p_i < 0.5, \text{ for } \mathcal{M}_{U2,V2}\\ \mathrm{logit}((\gamma^{\mathrm{right}}_{\mathrm{peak}} - \gamma_{\mathrm{min}}) \cdot \mathrm{D}_i + \gamma_{\mathrm{min}}), & \text{if } p_i \geq 0.5, \text{ for } \mathcal{M}_{U2,V2}\\ \mathrm{logit}((\gamma^{\mathrm{left}}_{\mathrm{peak}} - \gamma_{\mathrm{min}}) \cdot \mathrm{D}_i + \gamma_{\mathrm{min}}), & \text{if } p_i < \lambda, \text{ for } \mathcal{M}_{U3,V3}\\ \mathrm{logit}((\gamma^{\mathrm{right}}_{\mathrm{peak}} - \gamma_{\mathrm{min}}) \cdot \mathrm{D}_i + \gamma_{\mathrm{min}}), & \text{if } p_i \geq \lambda, \text{ for } \mathcal{M}_{U3,V3} \end{cases}\\
\beta_{\mathrm{truth}}, \beta_{\mathrm{conf}} &\sim \mathrm{Normal}(0, 0.5), \quad \phi_{\mathrm{truth}}, \phi_{\mathrm{conf}} \sim \mathrm{Gamma}(5, 3)\\
\mu_{\mathrm{truth},n} &\leftarrow \mathrm{logit}^{-1}(\alpha_{\mathrm{truth},i} + (\beta_{\mathrm{truth}} + v_{\mathrm{truth},i}) \cdot \mathrm{R}_n)\\
\mu_{\mathrm{conf},n} &\leftarrow \mathrm{logit}^{-1}(\alpha_{\mathrm{conf},i} + (\beta_{\mathrm{conf}} + v_{\mathrm{conf},i}) \cdot \mathrm{R}_n)\\
y_{\mathrm{truth},n} &\sim \mathrm{Beta}(\mu_{\mathrm{truth},n} \cdot \phi_{\mathrm{truth}}, (1 - \mu_{\mathrm{truth},n}) \cdot \phi_{\mathrm{truth}})\\
y_{\mathrm{conf},n} &\sim \mathrm{Beta}(\mu_{\mathrm{conf},n} \cdot \phi_{\mathrm{conf}}, (1 - \mu_{\mathrm{conf},n}) \cdot \phi_{\mathrm{conf}})
\end{aligned}
$$

Figure 2: Graphical representation of the Bayesian model. S is the numbers of statements. The data vector R is an indicator variable with values 1 and 0, denoting whether a statement was repeated or new, respectively. For new statements, the effect of repetition is eliminated as $\mathrm{R}_n = 0$.

ship between truth judgments and confidence is completely random. The null model shares the same structure as the other models but does not impose any assumptions about the relationship between the intercept terms for perceived truth and confidence.

The likelihood of the mean truth and confidence judgments, both scaled to the interval $[0, 1]$, is modeled using a reparameterized beta distribution. The so-called beta-proportion distribution is particularly suited for observations bounded between 0 and 1, as it independently models the mean (μ) and the precision (ϕ), where the latter is inversely related to the variance. For each of the n observed judgment means, the likelihood is defined as

$$
\begin{aligned}
y_{\mathrm{truth},n} &\sim \mathrm{Beta}(\mu_{\mathrm{truth},n} \cdot \phi_{\mathrm{truth}}, (1 - \mu_{\mathrm{truth},n}) \cdot \phi_{\mathrm{truth}})\\
y_{\mathrm{conf},n} &\sim \mathrm{Beta}(\mu_{\mathrm{conf},n} \cdot \phi_{\mathrm{conf}}, (1 - \mu_{\mathrm{conf},n}) \cdot \phi_{\mathrm{conf}}).
\end{aligned}
$$

In each beta distribution, the multiplicative components are defined using the complementary factors $\mu_{*,n}$ and $(1-\mu_{*,n})$ in such a way that the two beta shape parameters sum to the precision parameter (ϕ_*). Thereby, each distribution is centered on the mean (μ_*) while allocating the remaining probability mass symmetrically according to the specified precision.

For each of the n observed judgment means associated with statement i, the mean parameters are determined via the logit link function, incorporating the specific truth-confidence relationship defined by the model variant and the repetition effect,

$$
\begin{aligned}
\mu_{\mathrm{truth},n} &= \mathrm{logit}^{-1}(\alpha_{\mathrm{truth},i} + (\beta_{\mathrm{truth}} + v_{\mathrm{truth},i}) \cdot \mathrm{R}_n)\\
\mu_{\mathrm{conf},n} &= \mathrm{logit}^{-1}(\alpha_{\mathrm{conf},i} + (\beta_{\mathrm{conf}} + v_{\mathrm{conf},i}) \cdot \mathrm{R}_n),
\end{aligned}
$$

where R is the vector of repetition values (0 for new and 1 for repeated).

Priors

Priors of the models were selected to provide weak regularization, ensuring stable estimation while allowing the data to dominate the posterior. The population-level intercepts and slopes were assigned weakly informative normal priors (Normal$(0, 1.5)$ and Normal$(0, 0.5)$, respectively) to prevent extreme estimates on the logit scale. For the hierarchical structure, statement-level standard deviations (σ_u and σ_v) followed exponential distributions (Exponential(1)) to favor shrinkage toward the population mean, while the Cholesky factor (Ω) of the correlation matrix (Σ_v) for the random slopes was assigned an LKJ(4) prior. The mapping parameters γ_{min}, γ_{peak}, $\gamma^{\mathrm{left}}_{\mathrm{peak}}$, $\gamma^{\mathrm{right}}_{\mathrm{peak}}$ being bounded between 0 and 1, followed Beta$(5, 5)$ distributions, which center prior mass around 0.5. λ parameter in Model 3 was estimated $\lambda = \mathrm{logit}^{-1}(\lambda')$ where $\lambda' \sim \mathrm{Normal}(0, 0.5)$. Finally, the precision parameters (ϕ_{truth} and ϕ_{conf}) for the Beta-proportion likelihoods were assigned Gamma$(5, 3)$ priors, providing a robust expected range for the observed variance. Figure 2 demonstrates an overview of the parameters, prior distributions, and the likelihood function of the all model variants.

Implementation of the Models

Posterior distributions of the model parameters were estimated by fitting each model to the data using Hamiltonian Monte Carlo sampling in R (R Core Team, 2023) via the *rstan* package (Stan Development Team, 2025). Each model was fitted using eight chains, with 1,000 warm-up iterations per chain followed by 10,000 post–warm-up samples.

For model comparison, Bayes factors were computed using bridge sampling (Bennett, 1976; Meng & Wong, 1996) as implemented in the *bridgesampling* package (Gronau et al., 2020). In addition, predictive performance was assessed using leave-one-out cross-validation (LOOCV; Vehtari et al., 2017) via the *loo* package (Vehtari et al., 2024).

Reanalyses of Experimental Data

To get the parameter estimates and to select the best-performing model capturing the relationship between perceived truth and confidence, we fitted all models to two datasets.

Stump et al. (2024): This study investigated the cognitive processes underlying subjective confidence by extending a referential theoretical account and testing whether judging a statement as true enhances confidence. In this study, participants provided binary truth judgments and confidence ratings on a 6-point scale. A total of 75 participants completed the study, evaluating 120 statements. These statements consisted of difficult, affectively neutral trivia facts, balanced for factual truth. The main findings revealed that repetition significantly increased confidence ratings. Critically, participants reported higher confidence when they judged a statement as true rather than false, and a significant interaction showed that the boosting effect of repetition on confidence was most pronounced for statements perceived as true.

Nikolaev (2025): This study explored whether the elicitation of confidence ratings influences the magnitude of the truth effect. The study utilized an experiment with 131 participants assigned to four groups to manipulate the presence and timing of confidence ratings: 1) before truth judgment and 2) after truth judgment. The 64 presented statements were selected to be ambiguous trivia items to ensure a lack of certain prior knowledge, thereby facilitating the influence of repetition. Participants provided binary truth judgments and confidence judgments on a 6-point Likert scale. The results confirmed a repetition-based truth effect. Furthermore, repetition significantly increased confidence judgments. The findings also revealed that providing a confidence rating either before or after giving a truth judgment did not moderate the influence of repetition.

Data processing: In these two datasets, we aggregated binary truth ratings and 6-point Likert-scale confidence judgments at the item level separately for new and repeated statements. Truth judgments had values between 0 and 1. To match the scale of the different types of judgments, Likert scale confidence judgments between 1 and 6 were mapped onto 0 and 1. The final datasets included four means for each statement (i.e., average truth and confidence judgments under repeated and new presentation).

Results

Data of Stump et al. (2024)

To determine whether perceived truth serves as reliable predictor for baseline confidence, we first compared all informed models against the null model, which assumes no structural mapping between perceived truth and confidence. Bayes factors indicate substantial to strong evidence in favor of all proposed models over the null (all BF > 11). Specifically, the symmetrical V-shaped version of Model V1 showed the strongest evidence relative to the null model ($\mathrm{BF}_{V1/0} = 34.93$), suggesting that assuming a specific relationship between confidence and perceived truth provides a better account of the data than an unconstrained model.

Regarding the functional form of the relationship, Bayes factors favored the V-shaped (linear) models over their U-shaped (quadratic) counterparts across all variants. However, the evidence was only anecdotal. The Bayes factors were $\mathrm{BF}_{V/U} = 1.70$ for Model 1, 1.34 for Model 2, and 1.10 for Model 3. All Bayes factors favoring U-shaped models were weaker ($\mathrm{BF}_{U/V} < 1$).

As the V-shaped models were slightly favored, we continued with these versions to select the best model variant (i.e., symmetrical, asymmetrical, and undefined minimum). Among the V-shaped variants, the symmetrical Model V1 emerged as the best-fitting model (Figure 3). Assuming equal model prior odds, Model 1 had the highest posterior probability (P(Model V1 | Data) = .489). However, the corresponding Bayes factors indicated that the evidence for Model V1 was ambiguous ($\mathrm{BF}_{V1/V2} = 1.46$; $\mathrm{BF}_{V1/V3} = 3.03$). Therefore, it is not clear whether the additional parameters sufficiently improve the model fit to justify the added complexity. In line with these results, LOOIC results showed that Model V1 yielded the highest expected log point-wise predictive density among the other models (ELPD = 387.61).

As the slightly better model, the parameter estimates for the symmetrical Model V1 showed that the baseline confidence at the point of maximum ambiguity was $\widehat{\gamma}_{\text{min}} = .28$ (95% CrI: [.25, .32]). The estimated confidence at the extremes of falseness and truth was $\widehat{\gamma}_{\text{peak}} = .53$ (95% CrI: [.38, .68]). The results also showed that repetition has a positive effect for both truth ($\widehat{\beta} = 0.61$, 95% CrI: [0.47, 0.74]) and confidence ($\widehat{\beta} = 0.32$, 95% CrI: [0.19, 0.44]). However, the correlation between the truth and confidence effects across statements was estimated to be zero ($\widehat{\rho} = .00$, 95% CrI: [-.63, .63]). The very wide credible interval indicates that the data were not sufficiently informative for estimating the correlation.

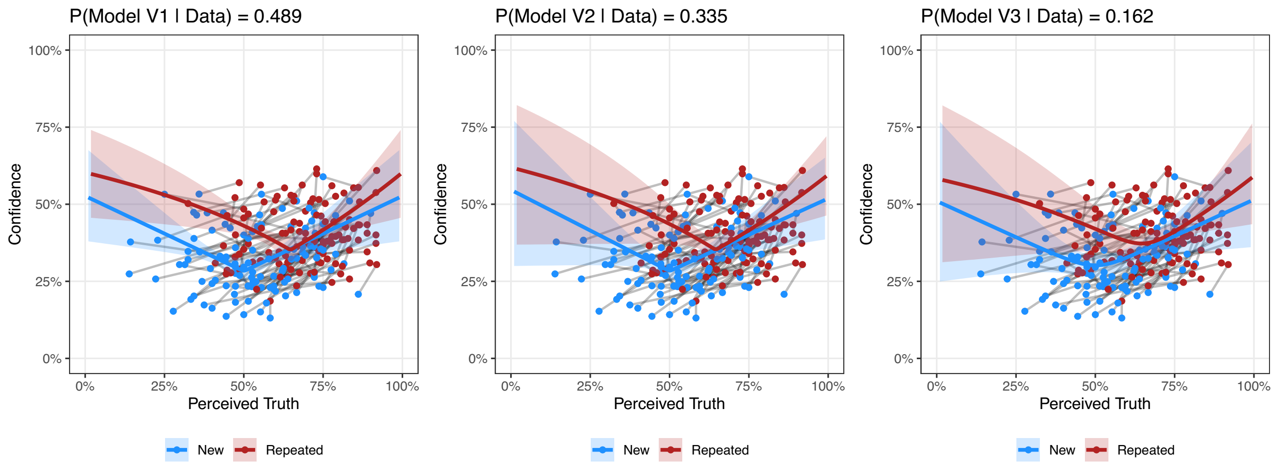


Figure 3: Fit of the three competing models for the data of Stump et al. (2024). Gray lines connect judgment means under new and repeated presentation of the same statement.

Data of Nikolaev (2025)

The comparison of the informed models against the null model revealed clear evidence in favor of a structural mapping between perceived truth and baseline confidence. Bayes factors relative to the null were substantially higher than in Stump et al. (2024), with the asymmetrical V-shaped model showing the strongest support ($\text{BF}_{V2/0} > 10^3$). Consistent with the findings from Stump et al. (2024), the functional form was better captured by the V-shaped models rather than the U-shaped models, but with stronger evidence. The Bayes factors were $\text{BF}_{V/U} = 7.53$ for Model 1, 115.69 for Model 2, and 7.48 for Model 3.

As the V-shaped models were again favored, we continued model selection with these versions. In terms of specific model variants, the asymmetrical Model V2 was clearly better than Model V1 ($\text{BF}_{V2/V1} = 53.56$) and relatively better than Model V3 ($\text{BF}_{V2/V3} = 1.31$). This model yielded the highest posterior probability (P(Model V2 | Data) = .561) and the best predictive performance according to LOOIC (ELPD = 185.59). While the symmetrical Model V1 was heavily discounted (P(Model V1 | Data) = .01; $\text{ELPD}_{\text{diff}} = -10.756$), the model with an undefined minimum (Model V3) was competitive (P(Model V3 | Data) = .428, $\text{ELPD}_{\text{diff}} = -0.538$). However, the higher parsimony of Model V2, which fixes the minimum confidence at 0.5 while allowing for asymmetrical confidence peaks, was slightly favored against Model V3 (Figure 4).

The parameter estimates for Model V2 showed that the baseline confidence at the point of maximum ambiguity was $\widehat{\gamma}_{\text{min}} = .26$ (95% CrI: [.21, .31]). A striking disparity was observed between the confidence peaks. While the peak confidence for statements perceived as certainly false was relatively low ($\widehat{\gamma}^{\text{left}}_{\text{peak}} = .33$, 95% CrI: [.15, .56]), the peak confidence for statements perceived as certainly true was much higher ($\widehat{\gamma}^{\text{right}}_{\text{peak}} = .76$, (95% CrI: [.59, .90]). This indicates that in Nikolaev (2025), the subjective perception of truth is associated with a much higher confidence peak than the perception of falseness. Finally, both perceived truth ($\widehat{\beta} = 0.75$, 95% CrI: [0.55, 0.95]) and confidence ($\widehat{\beta} = 0.62$, 95% CrI: [0.41, 0.83]) showed robust increases as a result of repetition. Nevertheless, similar to Stump et al. (2024), the estimated correlation between the statement-level random slopes was nearly zero ($\widehat{\rho} = .02$, 95% CrI: [-.62, .64]). Again, the wide credible interval indicates that the data were not sufficiently informative.

Discussion

This study aimed to formalize and adjudicate between different functional mappings of the relationship between perceived truth and confidence within the context of the repetition-based truth effect. By employing a Bayesian hierarchical framework, we tested structural assumptions about how fluency induced by repetition affects both dimensions jointly.

Our results provided consistent evidence against the null model, confirming that confidence is not an independent judgment dimension but rather structurally linked to the perceived truth of a statement. Across both datasets, the V-shaped (linear) models outperformed the U-shaped (quadratic) model variants, which are often suggested in the broader metacognition literature (Koriat, 2018). Our findings suggest that in the domain of truth judgments, confidence grows linearly on the probability scale. This indicates that the transition from uncertainty to certainty follows a constant rate of growth rather than an accelerating one. However, note that relying on nonlinear link functions (e.g., logit and probit transformations) may lead to different conclusions regarding the functional shape of the relationship (Aktepe & Heck, 2025; Loftus, 1978).

We observed notable differences in asymmetry between the two datasets. While the data of Stump et al. (2024) were relatively ambiguous regarding symmetry, the data of Nikolaev (2025) strongly supported an asymmetrical mapping. In the

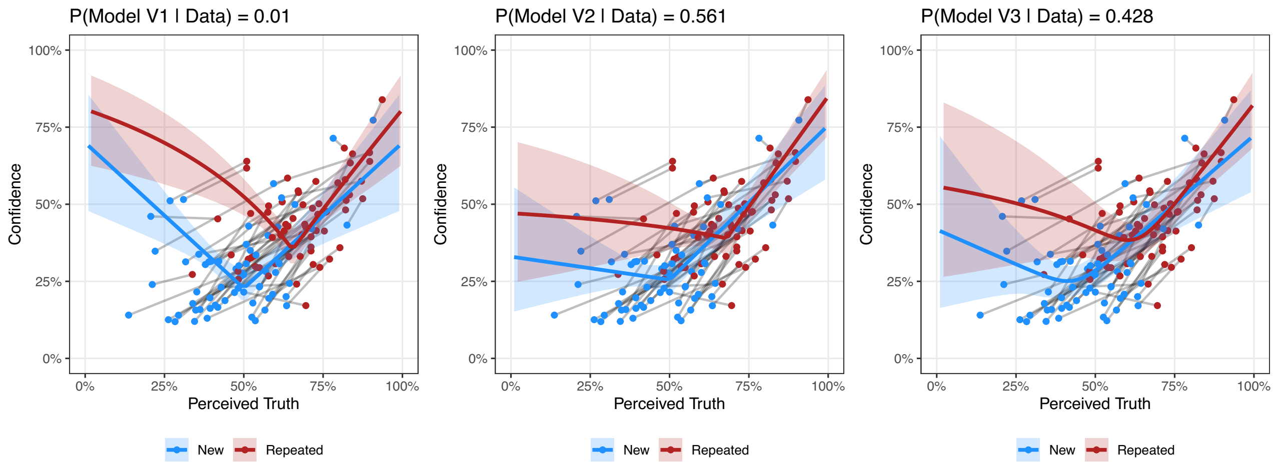


Figure 4: Fit of the three competing models for the data of Nikolaev (2025). Gray lines connect judgment means under new and repeated presentation of the same statement.

latter, the peak of confidence for the statements perceived as certainly true was substantially higher than for those perceived as certainly false. This suggests that the subjective perception of truth is accompanied by a higher meta-cognitive certainty than that of falseness, potentially reflecting a truth bias where participants are more prepared to be certain about the truth of a fact than its falseness (Brown & Nix, 1996). However, if our model were to utilized in fake news research, the patterns could be different. Individuals exhibit greater skepticism when evaluating the truthfulness of genuine news, while expressing higher confidence when spotting fake news as false (Pfänder & Altay, 2025). Thus, the truth–confidence asymmetry may be particularly pronounced in fake news contexts, and the direction of this asymmetry may be reversed relative to that in the data of Nikolaev (2025).

A central question was whether the repetition-induced increase in binary truth judgments and confidence ratings stems from a single underlying fluency signal (Stump et al., 2024; Unkelbach, 2007). If so, a strong positive correlation between the respective random slopes would be expected at the item level. However, our analyses were not informative as the estimated correlation was near zero, and the credible intervals were very wide. This suggests that while we can estimate the average effect of repetition both on perceived truth and confidence, the two reanalyzed datasets do not cover sufficiently many statements to determine whether the effects are driven by a single, underlying mechanism.

Finally, the Bayesian models developed here offer two major advantages for future research in cognitive psychology more broadly. First, truth and confidence judgments are traditionally analyzed using separate regressions (Nikolaev, 2025; Stump et al., 2022, 2024). Our approach enables the simultaneous estimation of these parameters, allowing the intercept of confidence to be a direct function of latent, perceived truth. Second, researchers often use a single joint response format to measure perceived truth and confidence (e.g, a Likert scale with response options from 'definitely false' to 'definitely true'). However, such a joint response format conflates plausibility (whether the statement is perceived as true) with meta-cognitive certainty (how sure the participant is). By collecting separate judgments for perceived truth and confidence, and then analyzing both dimensions via our models, researchers can maintain the psychometric properties of these two constructs in a joint analysis.

Despite the strengths of our modeling approach, two key limitations remain. First, as noted, the two datasets did not contain sufficiently many statements to provide a precise estimate of the correlation between repetition effects across items. Future studies should employ larger item pools specifically designed for testing this association. Second, our current models focus on mean truth and confidence judgments and do not fully capture subject-level variation at the trial level. Investigating the relationship of perceived truth and confidence at the subject-level is inherently difficult in typical experimental paradigms investigating the truth effect. Participants cannot judge the same statement twice once as repeated and once as new. When statements are repeatedly presented in different experimental blocks, their responses are correlated, which diminishes the size of the truth effect (Nadarevic & Erdfelder, 2014).

Conclusion

Our study formalized the relationship of perceived truth and confidence through a Bayesian hierarchical framework. We demonstrated that the relationship between perceived truth and confidence is best described by a linear V-shaped mapping rather than a quadratic curve, often characterized by an asymmetry. Nevertheless, our data did not allow us to answer the question of whether a single fluency signal drives the increase in both truth and confidence judgments. Overall, our Bayesian models offer a robust framework for future researchers to jointly model truth and confidence judgments.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 290878970-GRK2271, Project 4. We would like to thank to Daniel Nikolaev for the data, and Lara Orphal for useful discussion.

All data and analysis codes are publicly available on the Open Science Framework repository, accessible at https://osf.io/mvjp8/.

References

Aktepe, S. C., & Heck, D. W. (2025). Modeling the link between the plausibility of statements and the truth effect. *Psychonomic Bulletin & Review*, *32*, 1504–1520. https://doi.org/10.3758/s13423-025-02647-z

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81–94. https://doi.org/10.1002/bdm.3960020203

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*(2), 245–268. https://doi.org/10.1016/0021-9991(76)90078-4

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1088. https://doi.org/10.1037/0278-7393.22.5.1088

Bruine de Bruin, W., Fischhoff, B., Millstein, S. G., & Halpern-Felsher, B. L. (2000). Verbal and numerical expressions of probability: “It’s a fifty–fifty chance”. *Organizational Behavior and Human Decision Processes*, *81*(1), 115–131. https://doi.org/10.1006/obhd.1999.2868

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

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin & Review*, *26*(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Fiechter, J. L., & Kornell, N. (2021). Answering a factual question today increases one’s confidence in the same answer tomorrow–independent of fluency. *Psychonomic Bulletin & Review*, *28*(3), 962–968. https://doi.org/10.3758/s13423-021-01882-4

Foster, J. L., Huthwaite, T., Yesberg, J. A., Garry, M., & Loftus, E. F. (2012). Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. *Acta Psychologica*, *139*(2), 320–326. https://doi.org/10.1016/j.actpsy.2011.12.004

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, *92*(10), 1–29. https://doi.org/10.18637/jss.v092.i10

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

Koriat, A. (2018). When reality is out of focus: Can people tell whether their beliefs and judgments are correct or wrong? *Journal of Experimental Psychology: General*, *147*(5), 613. https://doi.org/10.1037/xge0000397

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167. https://doi.org/10.1038/nn.4064

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, *6*(3), 312–319. https://doi.org/10.3758/BF03197461

Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 831–860. https://www.jstor.org/stable/24306045

Nadarevic, L., & Erdfelder, E. (2014). Initial judgment task and delay of the final validity-rating task moderate the truth effect. *Consciousness and Cognition*, *23*, 74–84. https://doi.org/10.1016/j.concog.2013.12.002

Newman, E. J., Jalbert, M. C., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition*, *78*, 102866. https://doi.org/10.1016/j.concog.2019.102866

Nikolaev, D. (2025). *Certainty and truth–how does the elicitation of confidence ratings affect truth judgments?* [Master’s thesis]. Philipps-Universität Marburg.

Pfänder, J., & Altay, S. (2025). Spotting false news and doubting true news: A systematic review and meta-analysis of news judgements. *Nature Human Behaviour*, *9*, 688–699. https://doi.org/10.1038/s41562-024-02086-1

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Stan Development Team. (2025). RStan: The R interface to Stan [R package version 2.32.7]. 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*(1), 12–36. https://doi.org/10.1007/s00426-020-01459-1

Stump, A., Voss, A., & Rummel, J. (2024). The illusory certainty: Information repetition and impressions of truth enhance subjective confidence in validity judgments independently of the factual truth. *Psychological Research*, *88*(4), 1288–1297. https://doi.org/10.1007/s00426-024-01956-7

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219. https://doi.org/10.1037/0278-7393.33.1.219

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2024). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models [R package version 2.8.0]. 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

Yi, Y., & La, S. (2003). The moderating role of confidence in expectations and the asymmetric influence of disconfirmation on customer satisfaction. *The Service Industries Journal*, *23*(5), 20–47. https://doi.org/10.1080/02642060308565622