Behavior Research Methods (2025) 57:127
https://doi.org/10.3758/s13428-025-02650-1

ORIGINAL MANUSCRIPT

Consistency of perceptual response variability in size estimation and reproduction tasks

Kenny Yu[1] · Tzu-Yao Lin[1,2] · Jonas Zaman[1,3,4] · Francis Tuerlinckx[1] · Niels Vanhasbroeck[5]

Accepted: 5 March 2025
© The Psychonomic Society, Inc. 2025

Abstract

Measuring perceptual uncertainty is important for understanding how perception influences post-perception behavior, while it remains unknown whether measured perceptual responses in behavioral tasks reflect true perceptual uncertainty or methodological artifacts. This study compared two size perception approaches: a visual analog scale (VAS) estimation task and a reproduction task. We recruited 180 participants who completed both tasks by estimating circle diameters on a VAS and adjusting circle sizes to match specified diameters. Our analysis used two Bayesian multilevel models – a variance decomposition model computing intraclass correlation coefficients (ICCs) for overall response consistency, and a generative psychophysical model to characterize perceptual response patterns. Results revealed high overall response consistency across tasks, but detailed variance component analysis uncovered systematic method differences. Participant, stimulus, and interaction variances were consistently higher in the estimation task, indicating greater individual differences and more idiosyncratic responses. Psychophysical analyses further showed that the estimation task produced a steeper perceptual slope and a lower intercept compared to reproduction. Notably, while overall task-level perceptual uncertainty was nearly identical, the scaling of uncertainty with stimulus size was markedly stronger in estimation. These findings suggest that much of the observed variability reflects genuine perceptual uncertainty rather than measurement error, though distinct cognitive demands shape its expression. Our results confirm that both VAS and reproduction tasks yield consistent measures of perceptual variability, underscoring their value in behavioral research and the need for future studies to disentangle intrinsic perceptual processes from task-specific noise.

Keywords Consistency · Visual perception · ICC · Psychophysics · Cognitive measurement

Introduction

Much of human behavior is guided by perception, as we depend on the processing of sensory inputs to inform our actions. For example, when navigating through a forest, we use visual perception to identify potential threats, allowing us to avoid danger. The relationship between visual perception and behavior is a central focus in various domains of psychology and behavioral science, including categorization (Tanaka & Farah, 1993; Tarr & Gauthier, 2000; Palmeri & Gauthier, 2004), generalization (Zaman et al., 2019a, b, 2023), avoidance (Öhman & Soares, 1994; Phelps et al., 2006; Tamietto & De Gelder, 2010), social behavior (Adolphs, 2003; Haxby et al., 2000), motor behavior (Mazzoni & Krakauer, 2006; Goodale & Milner, 1992), and decision making (Heekeren et al., 2008; Ratcliff & Rouder, 1998). Understanding how perception influences behavior requires a precise measurement of perceptual patterns, as these patterns are fundamental to accurately capture the nuances of how sensory information is processed and acted upon.

In measuring visual perception and subsequent behavior, the alternative forced-choice (AFC) task is one of the most widely employed techniques. In this task, participants are sequentially or simultaneously presented with a set of stimuli and are required to either identify the target stimulus (i.e.,

✉ Kenny Yu
kenny.yu@kuleuven.be

1 KU Leuven, Leuven, Belgium
2 Maastricht University, Maastricht, Netherlands
3 University of Hasselt, Diepenbeek, Belgium
4 University of Duisburg-Essen, Essen, Germany
5 University of Amsterdam, Amsterdam, Netherlands

Published online: 28 March 2025

detection) (Nestor et al., 2008; Petursdottir & Aguilar, 2016; Liu et al., 2020; Watanabe et al., 2001; Zaman et al., 2023), determine which stimulus differs from the others (i.e., discrimination) (Neri et al., 1999; Peli et al., 2004; Carrasco et al., 2004; Corbetta et al., 1990), or both (Fritz et al., 2005; Watanabe et al., 2001; Carrasco & Barbot, 2019; Ling & Carrasco, 2006). A key advantage of the AFC task is that it minimizes response biases by requiring participants to choose between discrete options, thereby providing a clear measure to isolate specific perceptual aspects. Researchers use participant responses in these tasks to derive measures of perceptual sensitivity, which can then be analyzed in relation to the targeted behavior under investigation. However, the AFC task also has limitations. It may not capture subtle gradations in perception, can be less sensitive to small perceptual changes over time, and may not reflect naturalistic perceptual judgments in some contexts.

While the AFC task is widely used in visual perception research, alternative methods provide distinct advantages in other research contexts. This study highlights two such approaches: perceptual estimation using a visual analog scale (VAS) and perceptual reproduction. The VAS method requires participants to indicate their perceptual judgment of a stimulus feature on a continuous scale, allowing for a nuanced measurement of perceptual judgments. This makes the VAS suited to reveal subtle perceptual shifts over time. For instance, in fear generalization studies, the VAS has been used to measure temporal subjective perceptual estimations of stimuli, such as circle diameters (Yu et al., 2023, 2024b) and colors (Zaman et al., 2023), before assessing generalized response. Similarly, VAS has proven effective in pain research for gauging subjective intensity (Bijur et al., 2001; Zaman et al., 2017), and in color perception studies to capture a detailed continuum of responses, such as degrees of blueness or greenness, compared to categorical judgments (Douven et al., 2017). This approach enables researchers to track how perceptual changes evolve and influence subsequent behaviors.

Similarly, perceptual reproduction tasks have emerged as a powerful methodology for investigating perceptual information encoding, maintenance, and retrieval processes. Unlike recognition-based paradigms, these tasks require participants to actively reconstruct stimuli properties, providing continuous measurements that capture both systematic biases and precision in perceptual memory. This approach has proven particularly valuable in quantifying the fidelity of internal representations across individuals. In fear generalization research, studies employing reproduction tasks have demonstrated that variability in perceptual memory precision directly predicts the extent of fear generalization gradients, establishing a mechanistic link between memory uncertainty and defensive responding (Zenses et al., 2021; Yu et al., 2024a). Additionally, contemporary work in working memory using perceptual reproduction tasks has revealed that memory precision functions as a parametrically tunable property, systematically modulated by multiple interacting cognitive constraints. These include set size-dependent resource allocation mechanisms, temporal parameters of encoding and maintenance, and the distribution of attentional resources across different perceptual dimensions (Bays et al., 2024; Brady et al., 2024). These findings collectively illuminate how reproduction paradigms can quantify the computational architecture and constraints of perceptual memory systems, offering mechanistic insights into how representational uncertainty shapes behavior across contexts.

These continuous measurement approaches – VAS-based perceptual estimation and perceptual reproduction-mark a paradigm shift in how we conceptualize and study perception, moving beyond static perceptual behavior to uncover the dynamic, inferential nature of perceptual processes. Unlike traditional methods that reduce perception to discrete snapshots, these methods capture rich, continuous perceptual information. This makes them uniquely suited to reflect modern theoretical frameworks like predictive coding and Bayesian inference, which view perception as a probabilistic synthesis of incoming information and existing knowledge (Adams et al., 2004; Chopin & Mamassian, 2012; Aru et al., 2016; Guo et al., 2004). By providing continuous response formats, estimation, and reproduction tasks enable researchers to quantify perceptual uncertainty, evaluate representational precision, and examine how predictive mechanisms adapt perception across diverse contexts.

Yet, to date, limited research has evaluated the reliability of both estimation and reproduction approaches to measuring visual perception. The reliance on VAS in other fields has uncovered several potential sources of measurement error, including difficulties in translating subjective experiences into precise numerical values, which may introduce unwanted variability into the data (Wewers & Lowe, 1990), as well as challenges related to the physical act of making accurate markings on the scale (Bijur et al., 2001). Reproduction tasks are not without their own potential sources of error. These may include motor control limitations affecting the accuracy of reproductions, individual differences in spatial abilities, and biases introduced by the specific reproduction method (e.g., adjusting line length vs. drawing a shape). Additionally, anchoring effects, where individuals overly rely on an initial reference point or context when making judgments, can systematically skew reproduction responses (Allred et al., 2016; LeBoeuf & Shafir, 2006).

Consequently, the variability observed in data from both estimation and reproduction tasks may not reflect true perceptual differences but could instead be an artifact of inconsistencies in how participants engage with these measurement tools. This, in turn, could lead to misleading conclusions

regarding the nature and extent of perceptual variability. For instance, an overestimation of variability might suggest greater individual differences in visual perception than actually exist, while underestimation could obscure real differences. Moreover, the sources of error may differ between the two methods, potentially leading to systematic differences in results that are not reflective of underlying perceptual processes but rather of the measurement methods themselves.

This study takes an initial step toward evaluating the reliability of two methods for measuring visual perceptual variability: the VAS-based estimation and the reproduction tasks, with a focus on circle size perception. Circle size perception is a widely studied dimension in research exploring how perception influences generalization behavior (Lange et al., 2017; Lissek, 2012; Lissek et al., 2008, 2014; Zaman et al., 2019a, b; Yu et al., 2023, 2024b). We recruited 180 participants to complete two distinct perceptual tasks. In the size estimation task, participants viewed circles of varying sizes and estimated their diameters using a VAS, a continuous response scale. In the size reproduction task, participants were given specific diameter values and adjusted a circle to match, relying on mental representation and motor skills. Notably, these tasks differ in their response modalities – direct estimation versus active reproduction – which may introduce distinct measurement errors.

Our analysis proceeded in two interconnected stages, both grounded in a multilevel framework to capture the hierarchical structure of our data. First, we employed intraclass correlation coefficient (ICC) analysis – akin to generalizability theory – to assess the reliability of perceptual responses across the VAS-based estimation and the size reproduction tasks. The ICC decomposes total response variance into components attributable to participants, stimuli, and their interactions, thereby providing a measure of response consistency both within individuals (across stimuli) and across individuals (for identical stimuli). To deepen our understanding and further refine this analysis, we developed a multilevel Bayesian generative model that not only estimates these variance components but also captures specific perceptual response trends, such as how perceptual responses scale with stimulus size. This approach enabled us to quantify perceptual uncertainty while comparing task- and stimulus-specific patterns, offering mechanistic insights into the processes underlying perception that extend beyond traditional variance metrics. Our primary objective was to determine whether the variability observed in perceptual responses is consistent between the two tasks. If this variability primarily reflects true perceptual uncertainty – an intrinsic characteristic of the cognitive process – rather than task-specific measurement errors (i.e., methodological noise), we would expect to observe consistent patterns across both methods. Such findings would support the reliability of these approaches for capturing visual perceptual variability and lay a foundation for future research into the interplay between perception and behavior.

Methods

The study was pre-registered on the Open Science Framework (https://osf.io/c9mkz), which includes the experiment design and the analysis plan for the generative model. However, the intraclass correlation coefficients analysis was not pre-registered. All relevant materials, including experiment scripts, data, analysis scripts, and supplementary information for prior sensitivity analysis and posterior predictive checks – are available at https://osf.io/f97rz/. Ethical approval was obtained from KU Leuven's Social and Societal Ethics Committee (G-2024-8083).

Experiment

The experiment was conducted online, with 180 participants recruited through the Prolific platform. The sample size was determined based on a planned budget and simulations designed to ensure convergence and obtain narrow posterior distributions for the perceptual response variability parameters. Participants were excluded from the analysis if they had more than 20% missing data in either of the perceptual tasks or if over 50% of their responses in either task fell within a 10% range. After applying these exclusion criteria, the final sample included 167 participants (49% female, mean age = 28, SD = 6.4). Participants were compensated €6 for their participation. The experiment lasted for about 40 min.

Visual stimuli

In the experiment, participants were presented with five circles, each varying in diameter from 50.800 mm to 81.296 mm, with a consistent difference of 7.624 mm between consecutive circles. These circles, designated as S1 through S5, were displayed as white outlines on a black background. This specific range of circle sizes has been extensively utilized in the investigation of the relationship between perception and generalization behavior (Lange et al., 2017; Lissek, 2012; Lissek et al., 2008, 2014; Zaman et al., 2019a; Yu et al., 2023).

Procedure

At the start of the experiment, participants were instructed to align a physical credit card with a virtual one displayed on their screen and adjust the virtual card's size to match the physical card. This calibration process ensured that the circles were displayed at the correct sizes, regardless of

variations in screen sizes across participants. Following calibration, participants provided their consent to participate in the experiment. The experiment included two perceptual tasks, the order of which was fixed across participants. Before each task, participants were presented with interactive instructions. These instructions allowed participants to practice the tasks, ensuring they understood what was required before commencing the tasks.

The first perceptual task was a size estimation task. In each trial, participants were presented with a circle on the screen for 7 s, during which they were required to estimate the circle's diameter in millimeters using a visual analog scale (VAS). A fixation cross appeared during the 1.5-s intertrial interval (ITI) between trials. The task comprised 75 trials, divided into three blocks of 25 trials each, with each of the five circles being presented 5 times per block. A 20-s break separated each block. The size estimation task is illustrated in the left side of Panel A of Fig. 1.

The second perceptual task was a size reproduction task, where participants were tasked with constructing a circle to match a specific diameter provided as an instruction. These target diameters matched the circle sizes presented in the size estimation task, ensuring consistency across the two tasks. The trial structure mirrored that of the size estimation task: each trial began with participants seeing either a very small circle (1 mm) or a very large one (150 mm), which they then adjusted to the instructed size using the left button to decrease the diameter, the right button to increase it, and the down button to confirm their final adjustment. To keep the experiment efficient and prevent the use of external tools like rulers, participants had 12 s to complete each reproduction. As in the size estimation task, a fixation cross appeared during the 1.5-s inter-trial interval (ITI). Participants were debriefed at the experiment's end. This task is illustrated on the right side of Panel A of Fig. 1.

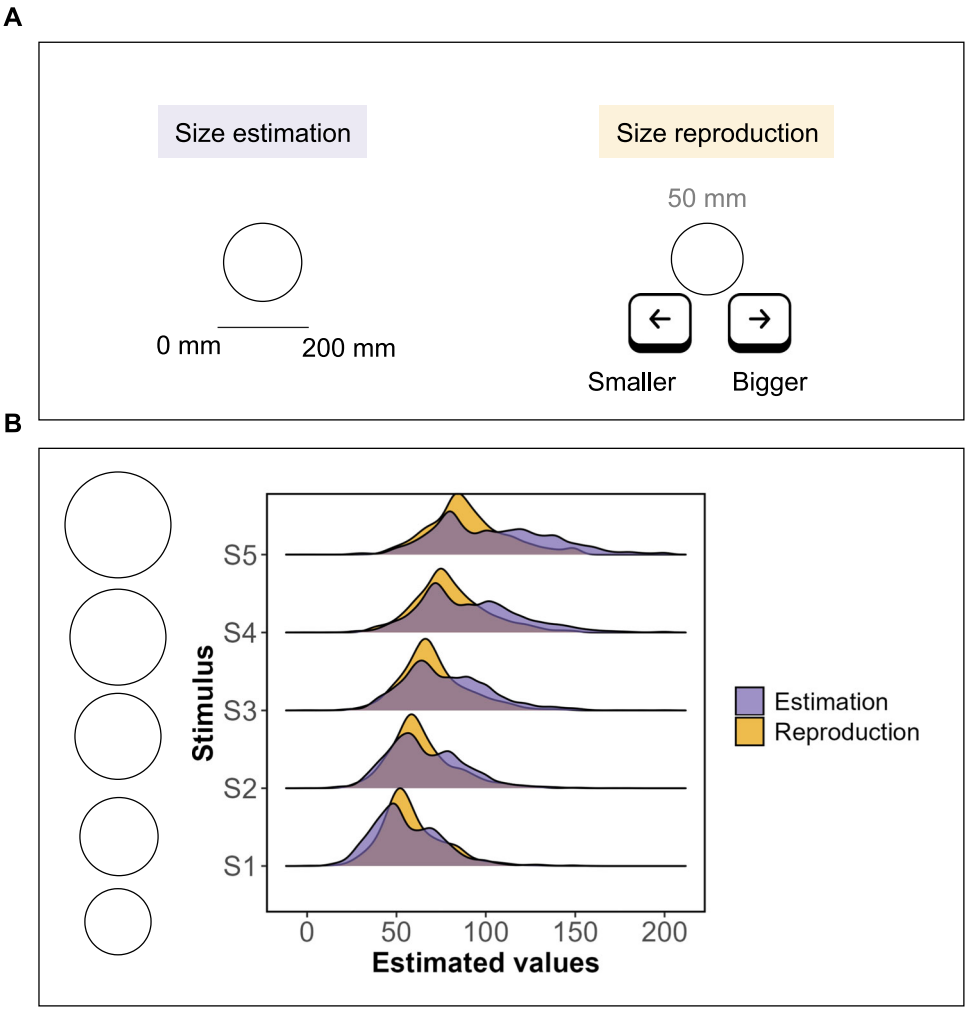


Fig. 1 The size estimation and reproduction tasks and the observed perceptual response patterns

Analysis

We conducted two complementary analyses using a multilevel Bayesian framework to evaluate perceptual response variability in size estimation and reproduction tasks. First, we employed a variance decomposition model to compute intraclass correlation coefficients (ICCs) and assess the consistency of perceptual measurements. This model systematically partitioned response variability into participant-specific effects, stimulus-induced variations, participant–stimulus interactions, and residual noise, providing a structured framework for evaluating response consistency across repeated trials. Second, we applied a generative psychophysical model to characterize response patterns and quantify perceptual uncertainty at multiple levels, including individual, stimulus-specific, and group-level variations. This approach revealed systematic differences in perceptual variability between tasks, complementing the insights from the ICC analysis and offering a more comprehensive understanding of how perceptual uncertainty manifests across different measurement methods.

Intraclass correlation coefficient

We denoted the task type as $t \in T = \{\text{size estimation}, \text{size reproduction}\}$, where size estimation involved participants viewing a stimulus circle and reporting its estimated diameter using a visual analog scale (VAS), and size reproduction required participants to construct a circle by adjusting its size to match a specific diameter provided as an instruction.

For participant $i = 1, \ldots, N$ responding to stimulus $s = 1, \ldots, S$ at replication $k = 1, \ldots, K$ in task t, the observed response $y_{t,i,s,k}$ was modeled as:

$$y_{t,i,s,k} = \mu_t + P_{t,i} + S_{t,s} + PS_{t,i,s} + \varepsilon_{t,i,s,k}, \tag{1}$$

where μ_t was the task-specific overall mean, $P_{t,i}$ represented the participant-specific effect for participant i in task t, $S_{t,s}$ denoted the stimulus effect for stimulus s in task t, $PS_{t,i,s}$ captured the interaction between participant i and stimulus s in task t, and $\varepsilon_{t,i,s,k}$ was the residual measurement noise. These random effects followed normal distributions with zero mean and task-specific variances:

$$P_{t,i} \sim \mathcal{N}(0, \sigma^2_{Pt}) \tag{2}$$

$$S_{t,s} \sim \mathcal{N}(0, \sigma^2_{St}) \tag{3}$$

$$PS_{t,i,s} \sim \mathcal{N}(0, \sigma^2_{PSt}) \tag{4}$$

$$\varepsilon_{t,i,s,k} \sim \mathcal{N}(0, \sigma^2_{\varepsilon t}) \tag{5}$$

The total variance of responses for task t was decomposed as:

$$\sigma^2_{yt} = \sigma^2_{Pt} + \sigma^2_{St} + \sigma^2_{PSt} + \sigma^2_{\varepsilon t}. \tag{6}$$

The ICC for task t, denoted ICC_t, measured the correlation between repeated responses within the same participant–stimulus combination, reflecting response consistency. It was calculated as the ratio of systematic variance (due to participant, stimulus, and their interaction) to total variance:

$$ICC_t = \frac{\sigma^2_{Pt} + \sigma^2_{St} + \sigma^2_{PSt}}{\sigma^2_{Pt} + \sigma^2_{St} + \sigma^2_{PSt} + \sigma^2_{\varepsilon t}}. \tag{7}$$

An ICC_t near 1 indicated high consistency (most variance was systematic), whereas a value close to 0 suggested low consistency (dominated by unexplained variability).

We specified non-informative to weakly informative priors for the ICC model parameters to ensure that the data primarily drove the estimation, given our limited prior knowledge about the variance components. These priors were detailed in Table 1.

To compare response consistency between tasks, we calculated the difference in ICCs:

$$ICC_{\text{diff}} = ICC_{\text{reproduction}} - ICC_{\text{estimation}}. \tag{8}$$

A positive ICC_{diff} indicated greater response consistency in the size reproduction task, while a negative value suggested higher consistency in size estimation, highlighting task-specific differences in the reliability of perceptual responses.

By decomposing total variance into participant, stimulus, interaction, and residual components, the ICC analysis clarified the relative contributions of these sources to perceptual response variability. However, the residual measurement noise component should not be strictly interpreted as pure measurement error; instead, it encompasses all unaccounted sources of variability, including trial-specific fluctuations, cognitive effects, and other uncontrolled influences. This distinction is crucial for correctly interpreting ICC values and

Table 1 Prior specifications for the ICC model parameters

Parameter	Prior Distribution	Description
μ_t	$\mathcal{N}(0, 1000)$	Task-specific overall mean
σ_{Pt}	Uniform$(10^{-9}, 50)$	SD of participant effects
σ_{St}	Uniform$(10^{-9}, 50)$	SD of stimulus effects
σ_{PSt}	Uniform$(10^{-9}, 50)$	SD of interaction effects
$\sigma_{\varepsilon t}$	Uniform$(10^{-9}, 50)$	SD of residual variance (unmodeled noise and variability)

assessing the reliability of perceptual measurements across tasks.

Perceptual response variability

We analyzed data from $i = 1, \ldots, 180$ participants and $s = 1, \ldots, 5$ stimuli, where each stimulus s corresponded to a circle with diameter X_s ranging from $X_1 = 50.80$ mm to $X_5 = 81.30$ mm, with equal intervals of 7.624 mm. The observed perceptual response $y_{t,i,s}$ for participant i to stimulus s in task t was modeled as:

$$y_{t,i,s} \sim \mathcal{N}(\mu_{t,i,s}, \sigma^2_{t,i,s}) \tag{9}$$

Here, $\mu_{t,i,s}$ represents the expected perceptual response, and $\sigma^2_{t,i,s}$ is the variance of responses, our key measure of perceptual variability (Panel A, Fig. 2). To reflect the nested nature of our data – responses varying across tasks, participants, and stimuli – we used a multilevel structure for $\sigma_{t,i,s}$, the standard deviation of responses, broken down into stimulus-level and group-level components:

$$\sigma_{t,i,s} \sim \text{LogNormal}(\mu_{\sigma_{t,s}}, \sigma^2_{\sigma_{t,s}}) \tag{10}$$

$$\mu_{\sigma_{t,s}} \sim \mathcal{N}(\mu_{\sigma_t}, \sigma^2_{\mu_{\sigma_t}}) \tag{11}$$

$$\sigma_{\sigma_{t,s}} \sim \text{Gamma}(a_{\sigma_t}, b_{\sigma_t}) \tag{12}$$

This multilevel setup works as follows: $\sigma_{t,i,s}$ varies for each participant i within a task t and stimulus s, drawn from a log-normal distribution with a mean $\mu_{\sigma_{t,s}}$ specific to that task and stimulus. Then, $\mu_{\sigma_{t,s}}$ itself varies across stimuli within each task, following a normal distribution centered at μ_{σ_t}, the task-wide average uncertainty. Finally, $\sigma_{\sigma_{t,s}}$, the spread of variability at the stimulus level, follows a gamma distribution, allowing flexibility across tasks. This structure captures individual differences while linking them to broader task and stimulus effects.

We modeled the expected perceptual response $\mu_{t,i,s}$ using a sigmoid function:

$$\mu_{t,i,s} = \frac{200}{1 + e^{-(\beta_{0,t,i} + \beta_{1,t,i} X_s)}} \tag{13}$$

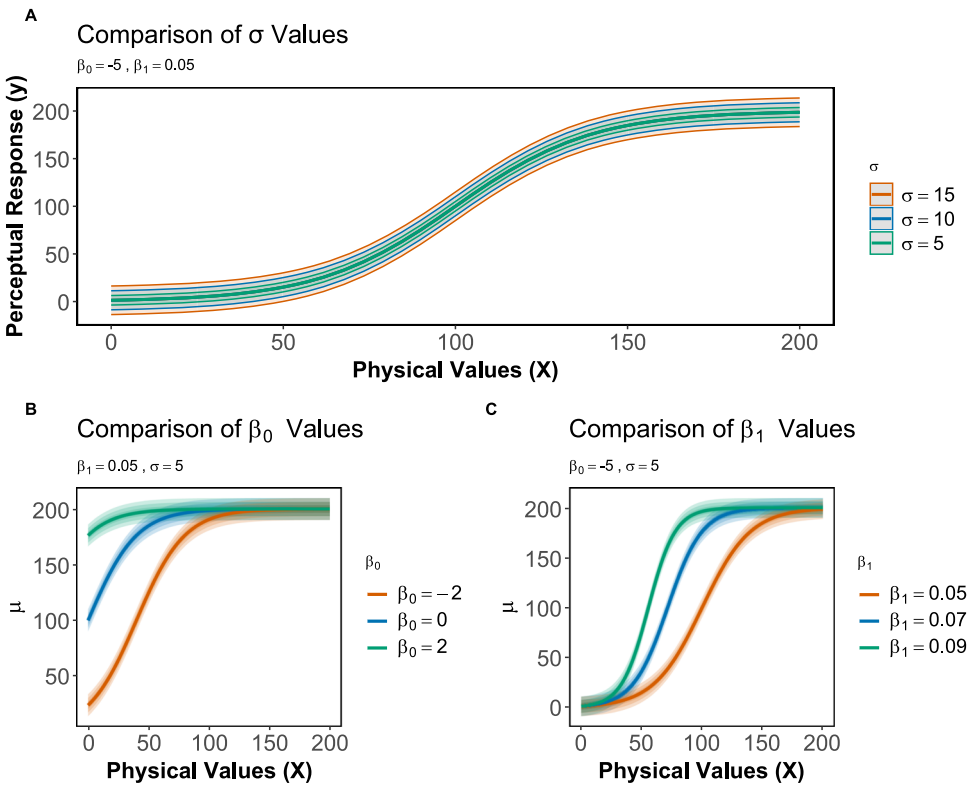


Fig. 2 Parameter effects. **Panel A**: The effect of $\sigma_{t,i,s}$, our primary parameter of interest, illustrating how it captured response variability while keeping other parameters fixed. **Panel B**: The effect of $\beta_{0,t,i}$. **Panel C**: The effect of $\beta_{1,t,i}$

where X_s represents the physical stimulus size. Traditional psychophysical functions, like Stevens' power law (Stevens, 1957), aim to explain group-averaged data by enforcing rigid scaling relationships between physical and perceived magnitudes, often overlooking individual variations. In contrast, we opted for a flexible sigmoid function, with parameters $\beta_{0,t,i}$ and $\beta_{1,t,i}$ controlling the curve's shift and slope, respectively. This allowed our model to adapt to diverse individual response patterns across tasks without assuming a uniform scaling rule. By leveraging the flexibility of $\beta_{0,t,i}$ and $\beta_{1,t,i}$, we captured a broad spectrum of perceptual sensitivities among participants, including deviations from classical psychophysical norms. Previous studies have effectively used this approach to explore human visual perception of circle sizes (Yu et al., 2023, 2024a, b). The sigmoid function has also been effective in other perceptual domains using VAS data (Yoshida et al., 2013; Zaman et al., 2017). Accounting for individual differences in response patterns enabled us to isolate the remaining variability, captured by $\sigma_{t,i,s}$, as true perceptual uncertainty. These parameters adhered to a multilevel structure:

$$\beta_{0,t,i} \sim \mathcal{N}(\mu_{\beta_{0,t}}, \sigma^2_{\beta_{0,t}}) \tag{14}$$

$$\beta_{1,t,i} \sim \text{LogNormal}(\mu_{\beta_{1,t}}, \sigma^2_{\beta_{1,t}}) \tag{15}$$

To compare perceptual response variability between tasks, we defined:

$$\delta_\sigma = \mu_{\sigma_{\text{estimation}}} - \mu_{\sigma_{\text{reproduction}}} \quad \text{(group-level difference)} \tag{16}$$

$$\delta_{\sigma_s} = \mu_{\sigma_{\text{estimation},s}} - \mu_{\sigma_{\text{reproduction},s}} \quad \text{(stimulus-level difference)} \tag{17}$$

We specified non-informative to weakly informative priors for all model parameters to ensure that the data primarily drove the parameter estimation, given our limited prior knowledge about the perceptual phenomena. These priors were detailed in Table 2.

Additionally, we extended the model to assess whether perceptual uncertainty scaled with stimulus magnitude, a phenomenon known as scalar variability in psychophysics (Petzschner et al., 2015; Gibbon, 1977; Loomis et al., 1992; Whalen et al., 1999). This relationship was expressed as:

$$\sigma_{t,i,s} = \sigma_{0,t,i} + \eta_{t,i} \ln(X_s) \tag{18}$$

Here, $\sigma_{0,t,i}$ represented the baseline perceptual uncertainty for participant i in task t (i.e., the uncertainty for the smallest stimulus), and $\eta_{t,i}$ was an individual- and task-specific scaling factor that determined how rapidly perceptual uncertainty increased with stimulus magnitude. The logarithmic transformation $\ln(X_s)$ reflected the principle of scalar variability, where uncertainty grows with stimulus magnitude at a diminishing rate, consistent with psychophysical evidence that perceptual sensitivity decreases as stimulus size increases (Petzschner et al., 2015). These parameters also followed a multilevel structure:

Table 2 Prior specifications for the perceptual scaling model parameters

Parameter	Prior Distribution	Description
$\mu_{\beta_{1,t}}$	$\mathcal{N}(-3, 2^2)$	Mean of log-slope parameter
$\sigma_{\beta_{1,t}}$	Gamma$(2, 1)T(10^{-9}, \infty)$	SD of log-slope parameter
$\mu_{\beta_{0,t}}$	$\mathcal{N}(0, 5^2)$	Mean of intercept parameter
$\sigma_{\beta_{0,t}}$	Gamma$(2, 1)T(10^{-9}, \infty)$	SD of intercept parameter
μ_{σ_t}	$\mathcal{N}(0, 5^2)$	Group-level mean of log-SD
$\sigma_{\mu_{\sigma_t}}$	Gamma$(2, 1)T(10^{-9}, \infty)$	Group-level SD of log-SD
$a_{\sigma_t}, b_{\sigma_t}$	Gamma$(2, 1)T(10^{-9}, \infty)$	Shape and rate for stimulus-level SD

$$\sigma_{0,t,i} \sim \text{LogNormal}(\mu_{\sigma_t}, \sigma^2_{\sigma_t}) \tag{19}$$

$$\eta_{t,i} \sim \mathcal{N}(\mu_{\eta_t}, \sigma^2_{\eta_t}) \tag{20}$$

For this extension, we introduced additional priors for $\sigma_{0,t,i}$ and $\eta_{t,i}$, specified as:

$$\mu_{\sigma_t} \sim \mathcal{N}(0, 5^2) \tag{21}$$

$$\sigma_{\sigma_t} \sim \text{Gamma}(2, 1)T(10^{-9}, \infty) \tag{22}$$

$$\mu_{\eta_t} \sim \mathcal{N}(0, 10^2) \tag{23}$$

$$\sigma_{\eta_t} \sim \text{Gamma}(2, 1)T(10^{-9}, \infty) \tag{24}$$

Statistical inference

The estimation of the models was performed using Gibbs sampling via JAGS (Plummer, 2003), utilizing the Markov chain Monte Carlo (MCMC) method. The analysis involved running four MCMC chains, each with a total of 50,000 iterations. To ensure stability in the sampling process, we incorporated a burn-in period of 25,000 iterations and applied a thinning factor of 10. This resulted in 10,000 samples being retained per parameter ($\frac{50000-25000}{10} \times 4$).

Convergence of parameter sampling was evaluated through visual inspection for irregular chain patterns and by checking the $\hat{R}$ value, which was required to be below 1.1, based on Gelman and Rubin diagnostics (Gelman & Rubin, 1992; Brooks & Gelman, 1998). For all post-sampling analyses, the

statistical computing language R (R Core Team, 2021) was utilized, along with the R package *jagsUI* (Kellner, 2021) to interface with JAGS in R.

To assess whether differences in perceptual response variability from the first analysis and ICC from the second analysis significantly differ from 0, we calculated Bayes factors (BF) using the Savage-Dickey method (Dickey, 1971; Wagenmakers et al., 2010). We compared these against the null hypothesis of no difference (i.e., difference = 0). Following the scale proposed by Kass and Raftery (1995), BF values between 10^{-1} and $10^{-0.5}$ indicate substantial evidence in favor of no difference in perceptual uncertainty, while BF values between 10^{-2} and 10^{-1} indicate strong evidence for this conclusion. On the other hand, BF values between $10^{0.5}$ and 10^{1}, and between 10^{1} and 10^{2}, provide substantial and strong evidence, respectively, supporting the conclusion that differences in perceptual uncertainty differ from 0.

Table 3 Comparison of variance components and ICC across tasks, derived from 10,000 MCMC posterior samples

Parameter	Condition	Mean (SD)	Median [95% CI]
μ	Estimation	68.35 (12.90)	70.05 [38.16, 89.63]
	Reproduction	67.04 (9.94)	68.47 [43.18, 82.85]
σ_P	Estimation	21.49 (1.24)	21.44 [19.23, 24.11]
	Reproduction	16.59 (0.93)	16.54 [14.89, 18.52]
σ_S	Estimation	26.57 (9.68)	24.75 [12.36, 47.42]
	Reproduction	19.10 (9.05)	16.58 [8.17, 42.89]
σ_{PS}	Estimation	8.06 (0.25)	8.05 [7.57, 8.57]
	Reproduction	4.37 (0.17)	4.37 [4.05, 4.71]
σ_E	Estimation	11.73 (0.08)	11.73 [11.58, 11.88]
	Reproduction	10.13 (0.07)	10.13 [10.00, 10.26]
ICC	Estimation	0.89 (0.04)	0.89 [0.83, 0.95]
	Reproduction	0.85 (0.05)	0.85 [0.77, 0.95]

Results

We present the results of our Bayesian multilevel analyses of the variance decomposition model and the psychophysical generative model, which assessed the consistency of perceptual responses and elucidated the psychophysical mechanisms underlying variability in size estimation and reproduction tasks. In both models, we employed non-informative to weakly informative priors to account for our limited prior knowledge of the perceptual phenomena under investigation. To evaluate the influence of these priors, we conducted sensitivity analyses using alternative weakly informative priors, confirming the robustness of our results (see Supplementary Figures 1–2). These analyses demonstrated that our findings remained consistent across prior specifications, indicating minimal dependence on prior assumptions. Additionally, posterior predictive checks assessed the generative model's fit and predictive accuracy (see Supplementary Figures 3–4), showing strong alignment with empirical data.

Intraclass correlation coefficient

The ICC analysis revealed high response consistency in both tasks (Table 3). For the estimation task, the ICC was 0.89 (95% CI [0.83, 0.95]), and for the reproduction task, it was 0.85 (95% CI [0.77, 0.95]), indicating that 89% and 85% of the total variance, respectively, stem from systematic differences between participants and stimuli. In this model, ICC is defined as the proportion of variance due to participant effects (σ_P^2), stimulus effects (σ_S^2), and their interaction (σ_{PS}^2) relative to the total variance, which includes measurement error (σ_E^2). While these ICC values suggest strong consistency, no standardized benchmarks exist for perceptual tasks, necessitating further research to establish interpretive guidelines.

In the estimation task, an ICC of 0.89 implies that only 11% of the variance reflects measurement error or inconsistency within participant–stimulus pairs. Similarly, the reproduction task's ICC of 0.85 indicates 15% unexplained variance. A Bayes factor of 0.205 supports the null hypothesis of no meaningful difference between the ICCs, suggesting comparable consistency across tasks.

Variance components highlight task-specific differences (Table 3). Participant variance (σ_P) was higher in estimation (median = 21.44, 95% CI [19.23, 24.11]) than in reproduction (median = 16.54, 95% CI [14.89, 18.52]), reflecting greater individual variability in size judgments. Stimulus variance (σ_S) followed a similar pattern, with estimation (median = 24.75, 95% CI [12.36, 47.42]) exceeding reproduction (median = 16.58, 95% CI [8.17, 42.89]), indicating more diverse stimulus responses in estimation. The participant–stimulus interaction (σ_{PS}) was also elevated in estimation (median = 8.05, 95% CI [7.57, 8.57]) compared to reproduction (median = 4.37, 95% CI [4.05, 4.71]), suggesting greater stimulus-specific idiosyncrasies. Finally, error variance (σ_E) was slightly higher in estimation (median = 11.73, 95% CI [11.58, 11.88]) than in reproduction (median = 10.13, 95% CI [10.00, 10.26]), pointing to increased trial-to-trial variability in estimation.

These results indicate that while both tasks exhibit high consistency, the estimation task shows greater variability across all components, particularly in individual and stimulus effects. The reproduction task, though slightly less consistent

overall, demonstrates more stable response patterns, with reduced variance across participants and stimuli.

Perceptual uncertainty

The multilevel generative model showed that perceptual response variability patterns were largely consistent across size estimation and reproduction tasks (Panels A and B, Fig. 3). We modeled perceptual variability, expressed as $\ln(\sigma)$, at individual, stimulus, and task levels, estimating all parameters on a logarithmic scale (see Table 2). At the task level, $\ln(\sigma)$ values were nearly identical for size estimation (2.118, 95% CI [1.938, 2.292]) and size reproduction (2.081, 95% CI [2.018, 2.137]), with a minimal difference ($\delta_\sigma = 0.038$, 95% CI [-0.152, 0.221]). A Bayes factor of 0.22 reinforced the lack of a significant task-related difference in log variability.

At the stimulus level, the largest differences in perceptual variability, measured as $\ln(\sigma)$, occurred at the extremes: the smallest stimulus (S1: -0.099, 95% CI [-0.207, 0.021]) and the largest (S5: 0.205, 95% CI [0.069, 0.322]). Intermediate stimuli showed smaller differences (S2: -0.036, 95% CI [-0.143, 0.063]; S3: 0.039, 95% CI [-0.064, 0.145]; S4: 0.091, 95% CI [-0.014, 0.206]). Bayes factors for these differences were 0.62 (S1), 0.16 (S2), 0.18 (S3), 0.60 (S4), and 12.61 (S5). Except for S5, where evidence supported a reliable difference, these Bayes factors indicated that the perceptual response variability was generally consistent across tasks.

Analysis of the group-level perceptual scaling parameters (β_0 and β_1 in Eq. 13) revealed distinct differences between the size estimation and size reproduction tasks. The group-level perceptual slope ($\mu_{\beta_1 t}$) was higher for size estimation (0.032, 95% CI [0.031, 0.034]) than for size reproduction (0.022, 95% CI [0.021, 0.023]), indicating greater sensitivity to changes in stimulus size in the estimation task. In contrast, the group-level perceptual intercept ($\mu_{\beta_0 t}$) was lower for size estimation (-2.715, 95% CI [-2.832, -2.603]) compared to size reproduction (-2.118, 95% CI [-2.194, -2.039]), suggesting a shift in the sigmoid function. These findings highlighted systematic task differences: size estimation exhibited a steeper perceptual response curve with a lower baseline compared to size reproduction.

As shown in Fig. 4, the higher $\mu_{\beta_1 t}$ for size estimation indicated a sharper increase in perceptual response per unit of stimulus size (X_s) compared to size reproduction. The more negative $\mu_{\beta_0 t}$ in size estimation shifted the sigmoid

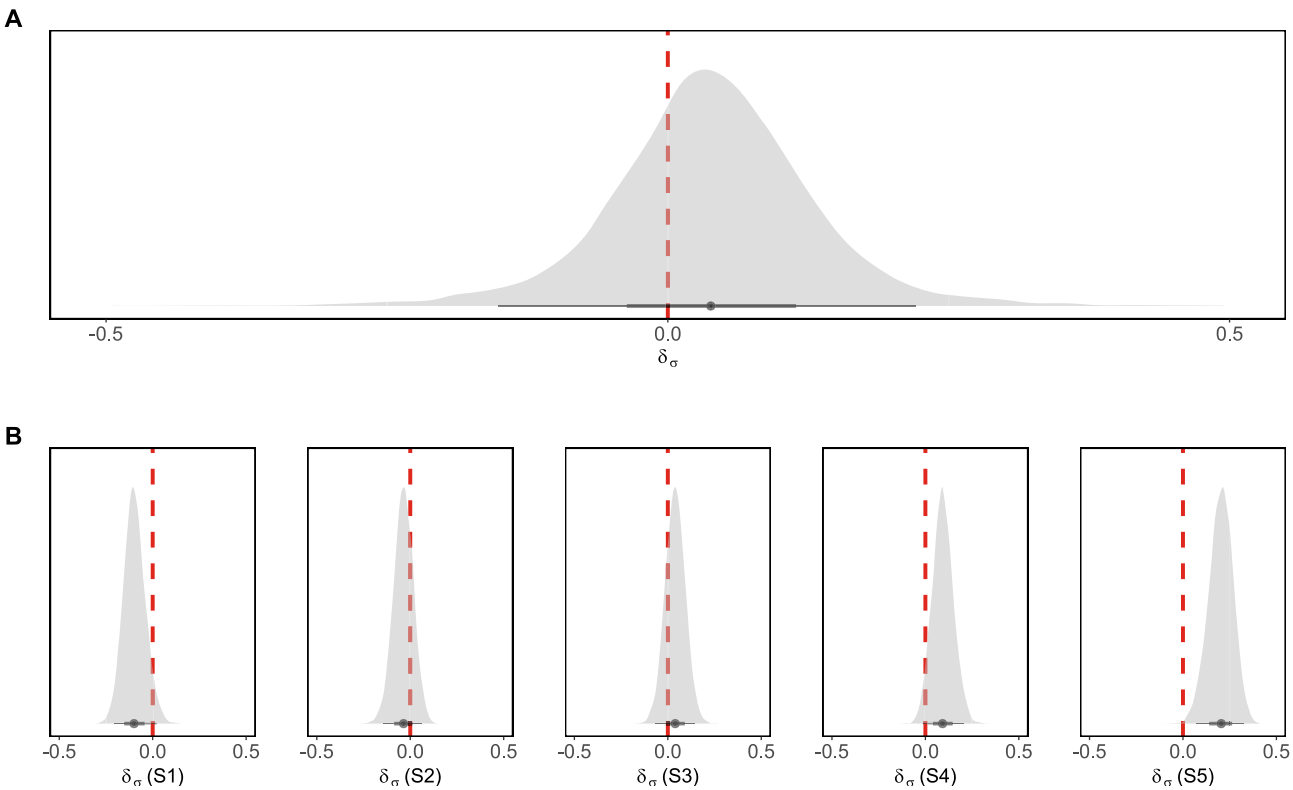


Fig. 3 Difference in perceptual response variability between size estimation and size reproduction tasks. **Panel A**: Posterior distributions based on 10,000 MCMC samples for the difference in mean task-level perceptual response variability. **Panel B**: Posterior distributions based on 10,000 MCMC samples for the difference in mean stimulus-level perceptual response variability. All parameters are presented on a logarithmic scale

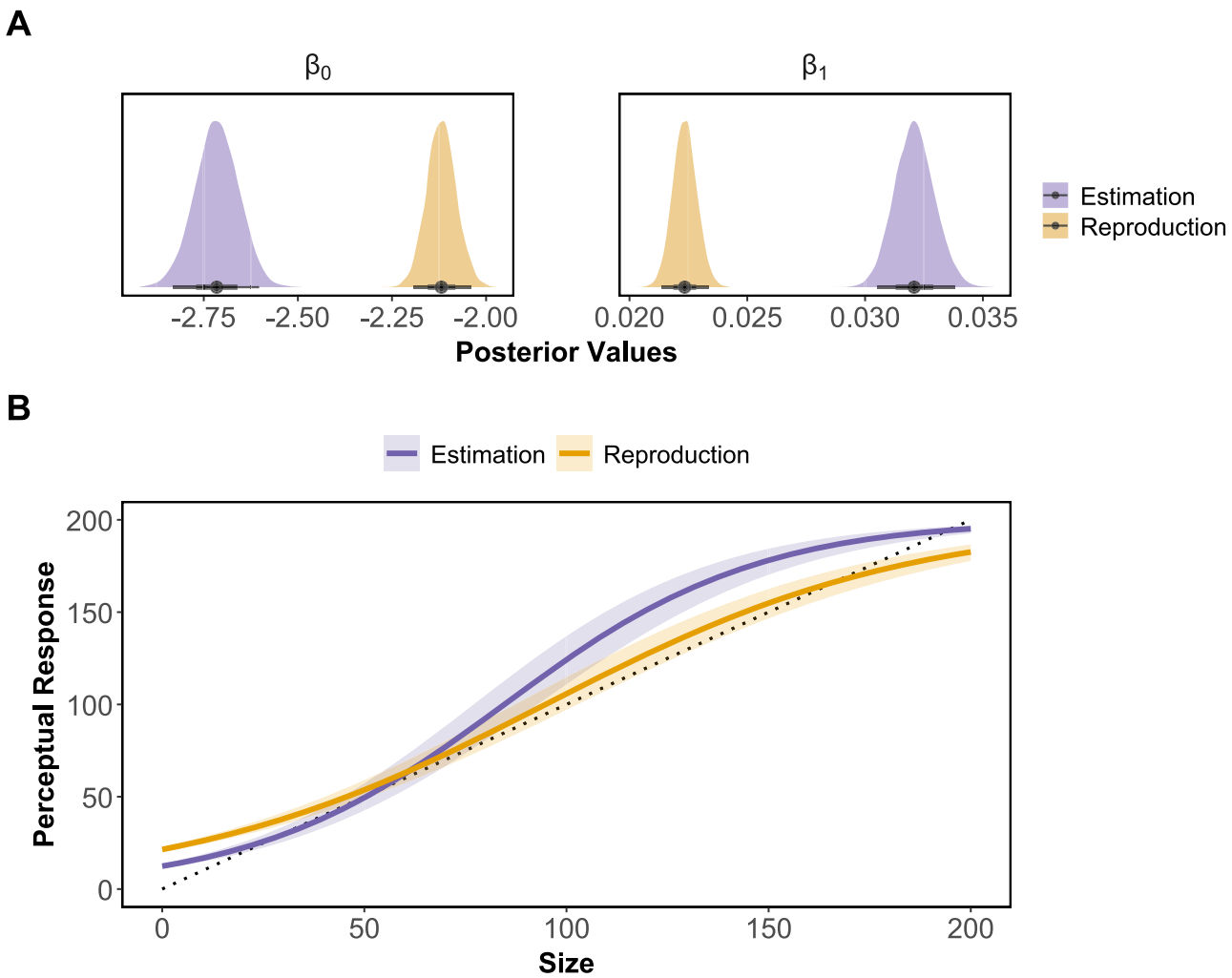


Fig. 4 The estimates of the task-level perceptual scaling parameters and a simulation of psychophysical curves for the size estimation and reproduction tasks. **Panel A**: Posterior distributions based on 10,000 MCMC samples for the task-level perceptual intercept (β_0) and perceptual slope (β_1) parameters. **Panel B**: Simulated psychophysical curves for stimulus sizes ranging from 0 to 200, with the main curve generated using the median estimates of the task-level β_0 and β_1, and the shaded region representing the 2.5th and 97.5th quantiles of the estimates

curve's midpoint to a larger stimulus size relative to size reproduction. Consequently, participants displayed greater sensitivity to smaller stimulus sizes in the reproduction task, with this difference diminishing at larger stimulus sizes.

The scalar effect model tested whether perceptual variability increased proportionally with stimulus magnitude in the size estimation and size reproduction tasks. The group-level parameter $\mu_{\eta t}$, which measures the rate at which variability scales with the logarithm of stimulus size ($\ln(X_s)$), was positive and credibly different from zero for both tasks: size estimation ($\mu_{\eta t} = 1.946$, 95% CI [1.483, 2.421]) and size reproduction ($\mu_{\eta t} = 0.781$, 95% CI [0.352, 1.207]). These estimates confirmed that perceptual variability (σ_{tis}) grew with increasing stimulus size in both tasks, supporting the principle of scalar variability wherein larger magnitudes exhibit greater variability. The effect was stronger in size estimation than in size reproduction, as illustrated in Fig. 5.

Discussion

In this study, we investigated perceptual response patterns in two distinct tasks: size estimation using a visual analog scale (VAS) and size reproduction, each employing a fundamentally different measurement approach. Our primary aim was to evaluate the consistency of response variability across these tasks. We found that individuals displayed highly consistent variability patterns in both, suggesting that task-specific measurement noise, while present, is unlikely to be the primary driver of the observed differences. Instead, a

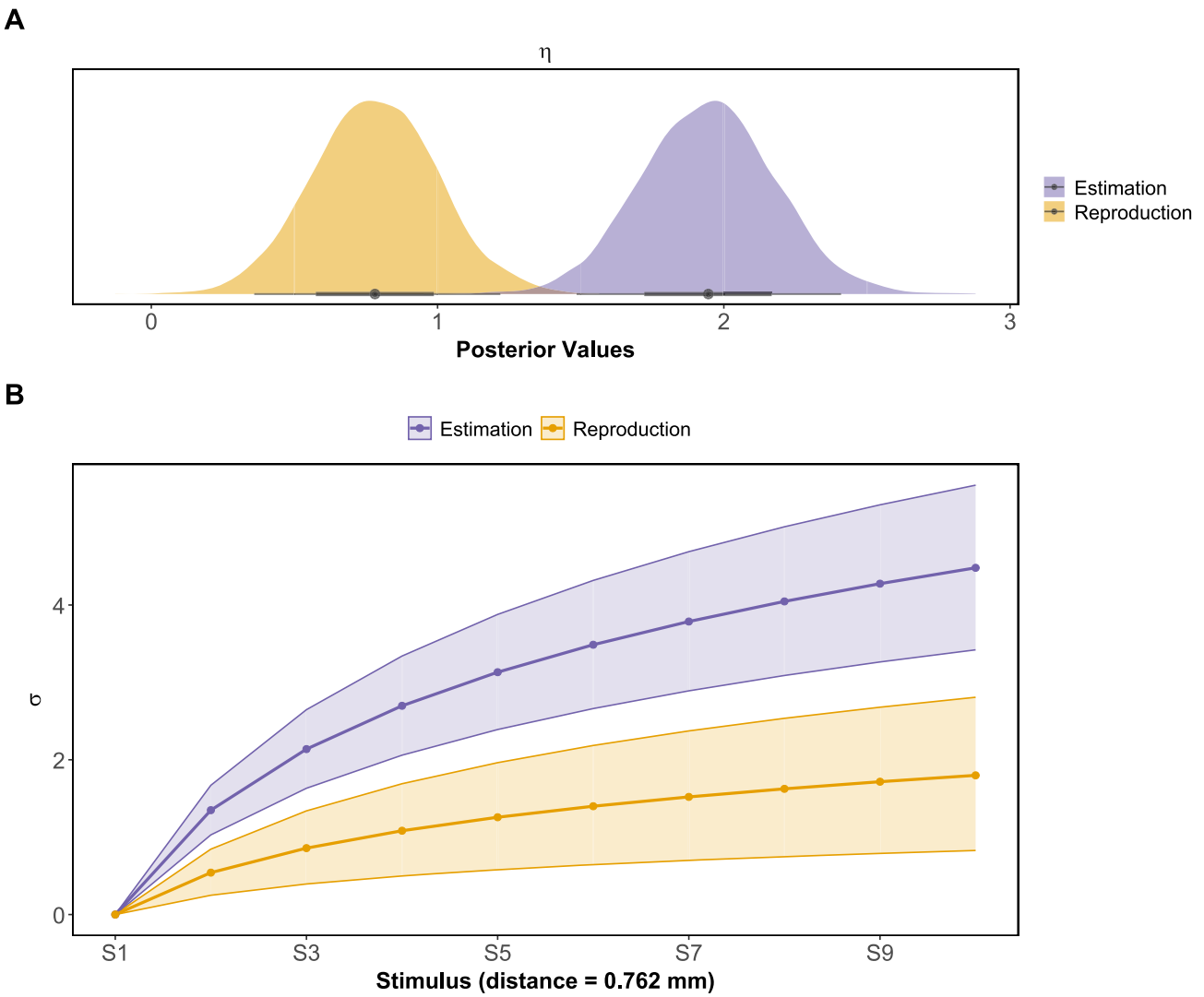


Fig. 5 The estimates of the group-level parameter η, which determines the effect of scalar variability, and a simulation of scalar variability for the size estimation and reproduction tasks. **Panel A**: Posterior distributions based on 10,000 MCMC samples for the group-level η. **Panel B**: Simulated scalar variability effect for ten stimuli (increment: 7.624 mm), with the *main curve* generated using the median estimates of η, and the *shaded region* representing the 2.5th and 97.5th quantiles of the estimates

substantial portion of this variability likely arises from genuine, stimulus-specific perceptual processes shared across both tasks. However, as our analysis did not directly distinguish perceptual from measurement errors, this interpretation remains preliminary. Nevertheless, the consistent variability patterns observed across these disparate tasks bolster the validity of both VAS-based estimation and size reproduction as reliable methods in perceptual research.

Methodologically, our study advances the understanding of perceptual uncertainty by employing intraclass correlation coefficient (ICC) analysis alongside multilevel Bayesian generative models. ICC analysis serves as a foundational step in evaluating the reliability of perceptual responses, systematically decomposing response variance into components attributable to participants, stimuli, and their interactions. This decomposition provides a quantitative measure of response consistency by partitioning variance into components attributable to systematic differences between participants, systematic differences between stimuli, their interactions, and unexplained variance. This allowed us to assess the relative contribution of these different sources to overall measurement variability. However, it is important to note that the residual noise component in ICC does not directly represent true measurement noise. Rather, it reflects the residual variance that remains unexplained by the specified components – participant, stimulus, and their interaction. As such, this component may contain a mix of actual measurement errors and unmodeled systematic influences rather than purely random

noise. Building on this foundation, we employed multilevel Bayesian generative models to further elucidate the latent perceptual processes underlying these tasks. These models offer a mechanistic perspective by explicitly characterizing how perceptual responses are shaped by both stimulus properties and task demands. By modeling psychophysical functions – including parameters such as slope, intercept, and uncertainty – our approach captures systematic differences in perceptual scaling and response variability. Together, ICC analysis and generative modeling provide a complementary framework for assessing perceptual reliability, demonstrating their utility for future research across diverse perceptual and cognitive domains.

Our exploration of perceptual response variability in size estimation and reproduction tasks reveals an intricate blend of consistency and divergence. The ICC analysis demonstrated high response consistency in both tasks (estimation: ICC = 0.89; reproduction: ICC = 0.85). However, an examination of the individual variance components revealed systematic differences in how variability manifested across tasks. All variance components were consistently higher in estimation compared to reproduction. Participant variance (σ_P) was greater in estimation (median = 21.44) than in reproduction (median = 16.54), indicating more pronounced individual differences in size judgments. Similarly, stimulus variance (σ_S) was higher in estimation (median = 24.75) than in reproduction (median = 16.58), reflecting more diverse responses to different stimulus sizes. Notably, the participant–stimulus interaction variance (σ_{PS}) was nearly twice as high in estimation (median = 8.05) as in reproduction (median = 4.37), suggesting that participants exhibited more idiosyncratic responses to specific stimuli in the estimation task. These findings indicate that while both tasks exhibit similar overall consistency, the estimation task elicits greater variability across all measured components.

Similarly, at the task level, the generative model showed that perceptual uncertainty estimates were nearly identical between size estimation (median $\ln(\sigma) = 2.118$) and reproduction (median $\ln(\sigma) = 2.081$). However, beneath this surface similarity, key distinctions emerged in how uncertainty scaled with stimulus size. Both tasks exhibited scalar variability (Petzschner et al., 2015; Gibbon, 1977; Loomis et al., 1992; Whalen et al., 1999), where uncertainty increased with stimulus magnitude. This effect was stronger in size estimation (median $\mu_{\eta t} = 1.946$) than in reproduction (median $\mu_{\eta t} = 0.781$) (Fig. 5). Psychophysical analysis further revealed systematic differences in how participants processed stimulus magnitudes across tasks. Size estimation exhibited greater sensitivity to size changes, as reflected in a steeper perceptual slope (median $\mu_{\beta_1 t} = 0.032$) compared to reproduction (median $\mu_{\beta_1 t} = 0.022$) (Fig. 4). Additionally, estimation showed a lower perceptual intercept, suggesting a fundamental difference in how stimulus magnitudes were mapped in each task.

These findings highlight the dynamic interplay between task-specific cognitive processes and stimulus properties in shaping perceptual variability. Size estimation, which requires mapping stimuli to a numerical scale, appears more susceptible to individual differences, stimulus-specific effects, and heightened scalar variability. In contrast, size reproduction, which integrates visual feedback, motor planning, and comparative judgment, exhibits reduced variance components and a tempered scalar effect, suggesting stabilizing mechanisms that limit variability growth – aligning with findings in visual mental imagery research (Liao et al., 2023; Thompson et al., 2011; Pylyshyn, 2002). This distinction underscores that perceptual variability arises not only from external stimuli but also from the unique cognitive demands of each task. Psychophysical curves for size estimation and reproduction revealed subtle but significant differences, pointing to distinct perceptual mapping processes. While size estimation likely involves transforming physical stimuli into mental representations, reproduction entails converting mental representations back into physical dimensions. Both tasks are widely used to study visual perception (Yu et al., 2023; Zaman et al., 2023) and perceptual memory (Zenses et al., 2021; Yu et al., 2024a) in perception-guided behavior. Our results showed a steeper curve for size estimation than reproduction, with a lower intercept for estimation. This pattern suggests that similar response profiles across tasks do not necessarily indicate identical perceptual or memory processes but rather stem from task-specific cognitive mechanisms. This interpretation is supported by studies demonstrating varied psychophysical curves across perceptual contexts (Franz, 2003; Gyles et al., 2023; Thomson et al., 2016). Consequently, caution is warranted when using these tasks to infer relationships among perception, memory, and behavior, as their distinct underlying processes may lead to divergent conclusions.

This investigation establishes a foundation for quantifying perceptual uncertainty through both VAS-based size estimation and size reproduction methodologies. A central finding of this study – the good consistency in perceptual variability patterns across distinct measurement approaches – bolsters confidence in using continuous methods to accurately measure perceptual uncertainty, paving the way for deeper insights into post-perception behavior. While this consistency suggests that both methods may be capturing a shared underlying perceptual signal, it does not fully disentangle intrinsic perceptual uncertainty from measurement-related artifacts. Specifically, the observed consistency could, in part, reflect common sources of measurement error inherent to

both tasks rather than purely perceptual variability. Consequently, future research must prioritize the development of methods to isolate task-specific measurement errors, potentially through multi-method approaches or by incorporating known error structures into analytical models. This critical distinction is necessary to determine whether the measured variability genuinely reflects fundamental perceptual noise or is confounded by methodological constraints. Extending this analytical framework to encompass diverse perceptual paradigms – such as identification tasks (Nestor et al., 2008; Petursdottir & Aguilar, 2016; Liu et al., 2020) and discrimination procedures (Neri et al., 1999; Peli et al., 2004) – via a comprehensive joint modeling approach could help unify these insights. Such integration could yield a consistent, cross-contextual measure of perceptual uncertainty while accounting for task-specific sources of variability.

While generative models provide a powerful lens for uncovering the intricacies of perceptual processes, offering detailed insights into how sensory inputs are transformed into perceptual judgments, their effectiveness rests critically on the empirical grounding of the psychophysical functions they incorporate. These functions, which define the relationship between physical stimuli and perceived responses, must accurately reflect the underlying perceptual mechanisms to ensure the models' predictions hold across diverse scenarios. Further research is essential to rigorously validate and refine these functions, testing their robustness across varied experimental contexts – such as different sensory modalities or task demands – and perceptual domains, from simple detection to complex discrimination. Future studies should prioritize a systematic approach: validating the assumed psychophysical relationships against empirical data, exploring alternative functional forms to capture a broader range of perceptual behaviors, and assessing how well these models generalize across diverse perceptual tasks, such as size estimation, motion perception, or texture discrimination. By strengthening these empirical foundations through such comprehensive efforts, researchers can not only bolster the reliability of generative modeling but also enhance its applicability, making it a more versatile and trustworthy tool for advancing perceptual science.

Limitations

The experiment was conducted in an online setting, and while we calibrated the screen sizes for participants to ensure the uniform presentation of circle sizes, several inherent limitations associated with online studies persisted. A primary challenge was the inability to control certain variables that could potentially affect the perceptual experiences of the participants. For example, variations in head position and viewing distance may have influenced the perceived size of the stimuli, thereby introducing inconsistencies into the data. Furthermore, environmental factors, including ambient lighting conditions and screen resolution, likely contributed to additional perceptual bias that was beyond our control.

To prevent the experiment from becoming excessively long and to minimize the likelihood of participants using measurement tools during the tasks, we imposed a maximum response time for both perceptual tasks. The time limits for each task were determined based on pilot studies and previous similar experiments, ensuring that most participants could respond within the allotted time. However, it is important to note that the time required for these perceptual responses can vary significantly between individuals, particularly in the size reproduction task. In the study, we observed that perceptual response variability was greater for the smallest and largest stimuli. A potential explanation for this finding is that adjusting the circle size to match the smallest or largest target circle may require more effort and time, especially when the initial circle size is at the opposite end of the spectrum (e.g., starting with a very large circle when reproducing the smallest target, or vice versa). This increased effort and time demand could contribute to the observed variability in responses for these extreme stimulus sizes.

The findings of the current study are specific to the dimension of size magnitudes. Since the VAS requires participants to translate their subjective estimations into numerical values, it is plausible that the measurement errors associated with the VAS may vary depending on how familiar participants are with converting the target perceptual constructs into numeric form. Future research should explore the use of the VAS across different perceptual dimensions to better understand how familiarity with the perceptual domain influences measurement accuracy.

Author Contributions Conceptualization, K.Y., T-Y.L., J.Z., F.T., N.V.; Methodology, K.Y., T-Y.L., N.V.; Software, K.Y.; Formal Analysis, K.Y., T-Y.L.; Investigation, K.Y.; Resources, K.Y.; Data Curation, K.Y.; Writing - Original Draft, K.Y.; Writing - Review & Editing, K.Y., T-Y.L., J.Z., F.T., N.V.; Visualization, K.Y.; Funding Acquisition, F.T. and J.Z.

Funding K.Y. is supported by an FWO research project (co-PI: J.Z., G079520N). J.Z. is a Postdoctoral Research Fellow of the Research Foundation Flanders (FWO, 12P8623N) and received funding from the Alexander von Humboldt Stiftung. K.Y., T-Y.L., and F.T. are also supported in part by the Research Fund of KU Leuven (C14/23/062). The resources and services used in this work were also provided by the VSC (Flemish Supercomputer Center), funded by FWO and the Flemish government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability The raw and processed data for the experiment in this study can be accessed at the following Open Science Framework (OSF) repository: https://osf.io/f97rz/.

Code Availability The code for the computational model and analysis can be found at the same repository as the data: https://osf.io/f97rz/. The Bayesian sampling is conducted with JAGS (version 4.3.1), and the post-sampling analysis and visualization are conducted with R (version 4.3.0).

Declarations

Ethical Approval Ethical approval for this research was obtained from KU Leuven's Social and Societal Ethics Committee (G-2024-8083).

Conflicts of Interest The authors declare no conflicts of interest.

Consent to Participate Informed consent to participate in the study was obtained from all participants.

Consent for Publication All participants consented to the publication of their data.

References

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience,7*, 1057–1058. https://doi.org/10.1038/nn1312

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4*, 165–178. https://doi.org/10.1038/nrn1056

Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review, 23*, 1825–1831. https://doi.org/10.3758/s13423-016-1039-0

Aru, J., Rutiku, R., Wibral, M., Singer, W. & Melloni, L. (2016). Early effects of previous experience on conscious perception. *Neuroscience of Consciousness*, niw004. https://doi.org/10.1093/nc/niw004

Bays, P. M., Schneegans, S., Ma, W. J., & Brady, T. F. (2024). Representation and computation in visual working memory. *Nature Human Behaviour, 8*, 1016–1034. https://doi.org/10.1038/s41562-024-01871-2

Bijur, P. E., Silver, W., & Gallagher, E. J. (2001). Reliability of the Visual Analog Scale for Measurement of Acute Pain. *Academic Emergency Medicine, 8*, 1153–1157. https://doi.org/10.1111/j.1553-2712.2001.tb01132.x

Brady, T. F., Robinson, M. M., & Williams, J. R. (2024). Noisy and hierarchical visual memory across timescales. *Nature Reviews Psychology, 3*, 147–163. https://doi.org/10.1038/s44159-024-00276-2

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*, 434–455. https://doi.org/10.1080/10618600.1998.10474787

Carrasco, M., & Barbot, A. (2019). Spatial attention alters visual appearance. *Current Opinion in Psychology, 29*, 56–64. https://doi.org/10.1016/j.copsyc.2018.10.010

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7*, 308–313. https://doi.org/10.1038/nn1194

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology, 22*, 622–626. https://doi.org/10.1016/j.cub.2012.02.021

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional Modulation of Neural Processing of Shape, Color, and Velocity in Humans. *Science, 248*, 1556–1559. https://doi.org/10.1126/science.2360050

Dickey, J. M. (1971). The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics, 42*, 204–223. https://doi.org/10.1214/aoms/1177693507

Douven, I., Wenmackers, S., Jraissati, Y., & Decock, L. (2017). Measuring Graded Membership: The Case of Color. *Cognitive Science, 41*, 686–722. https://doi.org/10.1111/cogs.12359

Franz, V. H. (2003). Manual size estimation: a neuropsychological measure of perception? *Experimental Brain Research, 151*, 471–477. https://doi.org/10.1007/s00221-003-1477-6

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005). Differential Dynamic Plasticity of A1 Receptive Fields during Multiple Spectral Tasks. *The Journal of Neuroscience, 25*, 7623–7635. https://doi.org/10.1523/JNEUROSCI.1318-05.2005

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science,7*. https://doi.org/10.1214/ss/1177011136

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Goodale, M. A., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*, 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Guo, K., et al. (2004). Effects on orientation perception of manipulating the spatio–temporal prior probability of stimuli. *Vision Research, 44*, 2349–2358. https://doi.org/10.1016/j.visres.2004.04.014

Gyles, S. P., McCarley, J. S., & Yamani, Y. (2023). Psychometric curves reveal changes in bias, lapse rate, and guess rate in an online vigilance task. *Attention, Perception, & Psychophysics, 85*, 2879–2893. https://doi.org/10.3758/s13414-023-02652-1

Haxby, J. V., Hoffman, E. A., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*, 467–479. https://doi.org/10.1038/nrn2374

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*, 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kellner, K. (2021). jagsUI: A Wrapper Around "rjags" to Streamline "JAGS" Analyses. *R package version,1*(5), 2.

Lange, I., et al. (2017). Behavioral pattern separation and its link to the neural mechanisms of fear generalization. *Social Cognitive and Affective Neuroscience, 12*, 1720–1729. https://doi.org/10.1093/scan/nsx104

LeBoeuf, R. A., & Shafir, E. (2006). The long and short of it: physical anchoring effects. *Journal of Behavioral Decision Making, 19*, 393–406. https://doi.org/10.1002/bdm.535

Liao, M.-R., Grindell, J. D., & Anderson, B. A. (2023). A comparison of mental imagery and perceptual cueing across domains of attention. *Attention, Perception, & Psychophysics, 85*, 1834–1845. https://doi.org/10.3758/s13414-023-02747-9

Ling, S., & Carrasco, M. (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research, 46*, 1210–1220. https://doi.org/10.1016/j.visres.2005.05.008

Lissek, S., et al. (2008). Generalization of conditioned fear-potentiated startle in humans: Experimental validation and clinical relevance. *Behaviour Research and Therapy, 46*, 678–687. https://doi.org/10.1016/j.brat.2008.02.005

Lissek, S. (2012). Toward an account of clinical anxiety predicated on basic, neurally-mapped mechanisms of pavlovian fear-learning: The case for conditioned overgeneralization. *Depression and anxiety, 29*, 257–263. https://doi.org/10.1002/da.21922

Lissek, S., et al. (2014). Neural substrates of classically conditioned fear-generalization in humans: A parametric fMRI study. *Social Cognitive and Affective Neuroscience, 9*, 1134–1142. https://doi.org/10.1093/scan/nst096

Liu, J., Lu, Z.-L., & Dosher, B. (2020). Similar perceptual learning in 10-alternative letter identification in external noise with and without feedback supervision. *Journal of Vision, 20*, 1237. https://doi.org/10.1167/jov.20.11.1237

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 906–921. https://doi.org/10.1037/0096-1523.18.4.906

Mazzoni, P., & Krakauer, J. W. (2006). An Implicit Plan Overrides an Explicit Strategy during Visuomotor Adaptation. *The Journal of Neuroscience, 26*, 3642–3645. https://doi.org/10.1523/JNEUROSCI.5317-05.2006

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*, 695–698. https://doi.org/10.1038/44409

Nestor, A., Vettel, J. M., & Tarr, M. J. (2008). Task-Specific Codes for Face Recognition: How they Shape the Neural Representation of Features for Detection and Individuation. *PLoS ONE, 3*, e3978. https://doi.org/10.1371/journal.pone.0003978

Öhman, A., & Soares, J. J. F. (1994). "Unconscious anxiety": Phobic responses to masked stimuli. *Journal of Abnormal Psychology,103*, 231–240. https://doi.org/10.1037/0021-843X.103.2.231

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*, 291–303. https://doi.org/10.1038/nrn1364

Peli, E., Garcia-Perez, M. A., Giorgi, R., & Woods, R. L. (2004). Spatial or temporal 2AFC may give different results depending on context. *Journal of Vision, 4*, 779–779. https://doi.org/10.1167/4.8.779

Petursdottir, A. I., & Aguilar, G. (2016). Order of stimulus presentation influences children's acquisition in receptive identification tasks. *Journal of Applied Behavior Analysis, 49*, 58–68. https://doi.org/10.1002/jaba.264

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences, 19*, 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion Facilitates Perception and Potentiates the Perceptual Benefits of Attention. *Psychological Science, 17*, 292–299. https://doi.org/10.1111/j.1467-9280.2006.01701.x

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Working Papers*

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157–182. https://doi.org/10.1017/S0140525X02000043

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science, 9*, 347–356. https://doi.org/10.1111/1467-9280.00067

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181. https://doi.org/10.1037/h0046162

Tamietto, M., & De Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience, 11*, 697–709. https://doi.org/10.1038/nrn2889

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*, 225–245. https://doi.org/10.1080/14640749308401045

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3*, 764–769. https://doi.org/10.1038/77666

Thompson, W. L., Hsiao, Y., & Kosslyn, S. M. (2011). Dissociation between visual attention and visual mental imagery. *Journal of Cognitive Psychology, 23*, 256–263. https://doi.org/10.1080/20445911.2011.477810

Thomson, D. R., Besner, D., & Smilek, D. (2016). A critical examination of the evidence for sensitivity loss in modern vigilance tasks. *Psychological Review, 123*, 70–83. https://doi.org/10.1037/rev0000021

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*, 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844–848. https://doi.org/10.1038/35101601

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health, 13*, 227–236. https://doi.org/10.1002/nur.4770130405

Whalen, J., Gallistel, C., & Gelman, R. (1999). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science, 10*, 130–137. https://doi.org/10.1111/1467-9280.00120

Yoshida, W., Seymour, B., Koltzenburg, M., & Dolan, R. J. (2013). Uncertainty Increases Pain: Evidence for a Novel Mechanism of Pain Modulation Involving the Periaqueductal Gray. *The Journal of Neuroscience, 33*, 5638–5646. https://doi.org/10.1523/JNEUROSCI.4984-12.2013

Yu, K., Vanpaemel, W., Tuerlinckx, F. & Zaman, J. (2024a). The representational instability in the generalization of fear learning. *npj Science of Learning*, *9*, 78. https://doi.org/10.1038/s41539-024-00287-x

Yu, K., Beckers, T., Tuerlinckx, F., Vanpaemel, W., & Zaman, J. (2024b). The assessment of gender differences in perceptual fear generalization and related processes. *Behaviour Research and Therapy, 183*, 104640. https://doi.org/10.1016/j.brat.2024.104640

Yu, K., Tuerlinckx, F., Vanpaemel, W., & Zaman, J. (2023). Humans display interindividual differences in the latent mechanisms underlying fear generalization behaviour. *Communications Psychology, 1*, 5. https://doi.org/10.1038/s44271-023-00005-0

Zaman, J., Ceulemans, E., Hermans, D., & Beckers, T. (2019). Direct and indirect effects of perception on generalization gradients. *Behaviour Research and Therapy, 114*, 44–50. https://doi.org/10.1016/j.brat.2019.01.006

Zaman, J., Struyf, D., Ceulemans, E., Beckers, T., & Vervliet, B. (2019). Probing the role of perception in fear generalization. *Scientific Reports, 9*, 10026. https://doi.org/10.1038/s41598-019-46176-x

Zaman, J., Vanpaemel, W., Aelbrecht, C., Tuerlinckx, F., & Vlaeyen, J. (2017). Biased pain reports through vicarious information: A computational approach to investigate the role of uncertainty. *Cognition, 169*, 54–60. https://doi.org/10.1016/j.cognition.2017.07.009

Zaman, J., Yu, K., Andreatta, M., Wieser, M. J., & Stegmann, Y. (2023). Examining the impact of cue similarity and fear learning on perceptual tuning. *Scientific Reports, 13*, 13009. https://doi.org/10.1038/s41598-023-40166-w

Zaman, J., Yu, K., & Verheyen, S. (2023). The idiosyncratic nature of how individuals perceive, represent, and remember their surroundings and its impact on learning-based generalization. *Journal*

of Experimental Psychology: General. https://doi.org/10.1037/xge0001403

Zenses, A.-K., Lee, J. C., Plaisance, V., & Zaman, J. (2021). Differences in perceptual memory determine generalization patterns. *Behaviour Research and Therapy, 136*, 103777. https://doi.org/10.1016/j.brat.2020.103777

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.