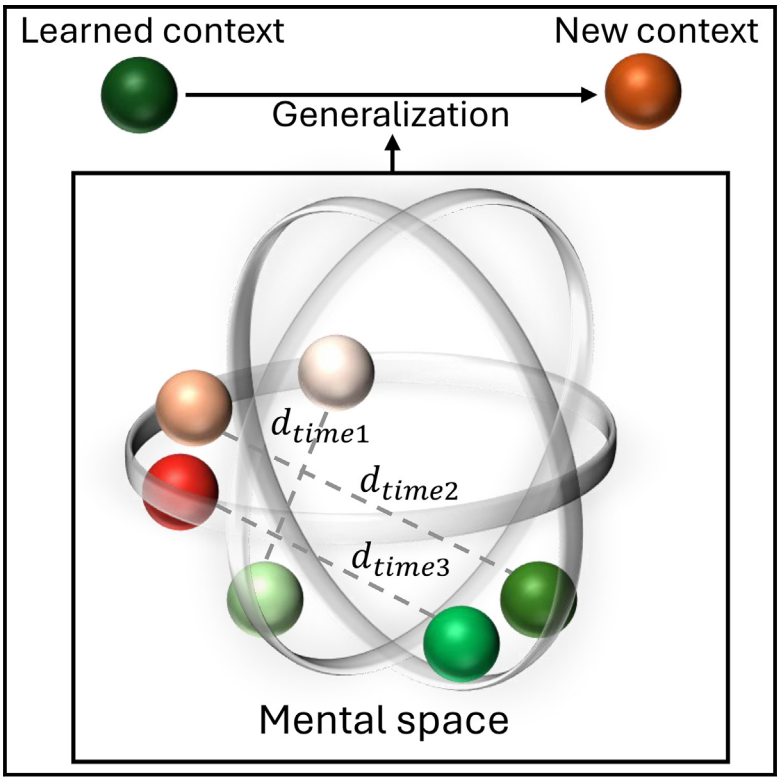
iScience

Article

The probabilistic and dynamic nature of perception in human generalization behavior

Graphical abstract



Authors

Kenny Yu, Wolf Vanpaemel, Francis Tuerlinckx, Jonas Zaman

Correspondence

kenny.yu@kuleuven.be

In brief

Social sciences; Psychology

Highlights

- Human perception varies between individuals and within individuals over time
- Probabilistic perception modulates how past fear experiences generalize to new situations
- Cognitive–perceptual interplay drives fear generalization
- Computational modeling connects generalization and perceptual theories

Yu et al., 2025, iScience *28*, 112228
April 18, 2025 © 2025 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2025.112228

CellPress

iScience

CellPress
OPEN ACCESS

Article

The probabilistic and dynamic nature of perception in human generalization behavior

Kenny Yu,[1,6,*] Wolf Vanpaemel,[1] Francis Tuerlinckx,[1,5] and Jonas Zaman[2,3,4,5]
[1]Quantitative Psychology and Individual Differences, KU Leuven, 3000 Leuven, Belgium
[2]REVAL Rehabilitation Research, Faculty of Rehabilitation Sciences, UHasselt, 3590 Diepenbeek, Belgium
[3]Centre for Learning and Experimental Psychopathology, KU Leuven, 3000 Leuven, Belgium
[4]Center for Translational Neuro- and Behavioral Sciences, University of Duisburg-Essen, 47057 Duisburg, Germany
[5]These authors contributed equally
[6]Lead contact
*Correspondence: kenny.yu@kuleuven.be
https://doi.org/10.1016/j.isci.2025.112228

SUMMARY

Generalization theories traditionally overlook how our mental representations dynamically change in the process of transferring learned knowledge to new contexts. We integrated perceptual and generalization theories into a computational model using data from 80 participants who underwent Pavlovian fear conditioning experiments. The model analyzed continuous measures of perception and fear generalization to understand their relationship. Our findings revealed large individual variations in perceptual processes that directly influence generalization patterns. By examining how perceptual and generalization mechanisms work together, we uncovered their combined role in producing generalization behavior. This research illuminates the probabilistic perceptual foundations underlying individual differences in generalization, emphasizing the crucial integration between perceptual and generalization processes. Understanding this relationship enhances our knowledge of generalization behavior and has potential implications for various cognitive domains including categorization, motor learning, language processing, and face recognition—all of which rely on generalization as a fundamental cognitive process.

INTRODUCTION

Humans possess a remarkable ability to extrapolate past learning to new situations, a cognitive process known as generalization.[1] This adaptive mechanism allows individuals to transfer knowledge efficiently, avoiding the need to relearn in similar contexts. Generalization relies heavily on the concept of similarity: the more alike two situations are, the more readily knowledge is transferred between them.[2] However, humans do not interact with a physical reality directly. Instead, we perceive the world through mental representations shaped by our sensory systems. These mental representations form the foundation for assessing similarity and guiding generalization. Crucially, these representations vary both between individuals and within the same individual across different instances, leading to idiosyncratic patterns in how similarity is assessed and generalization occurs.[3]

However, a prevalent trend in generalization research is the tendency to conflate the mental with the physical, often treating these dimensions as synonymous entities.[1,4] While some researchers have acknowledged the inherent noise in the perceptual process[5] or recognized that the mental representation may deviate from the physical reality (Figure 1A),[2] they often overlooked three critical aspects of perception: the variability across individuals, the changes within individuals over time, and how learning processes themselves can modify perception.[6]

As organisms, we perceive the physical world indirectly through sensory inputs across various modalities.[7] These inputs undergo complex mental processing to form perceptual judgments.[8,9] The predictive coding framework[10,11] within Bayesian brain theory[12–14] provides a compelling explanation for this process: the brain functions as an inference system that generates predictions based on environmental expectations. When encountering new information, perception involves integrating these new experiences with past predictions. The degree of updating depends on the relative uncertainty – greater updating occurs when sensory evidence is more certain or prior knowledge is uncertain. While debate continues about whether human perception optimally follows Bayes' rule,[15,16] substantial evidence supports that past experiences shape current perception.[17–20] Consequently, mental representations of stimuli exist not as fixed points but as evolving probability distributions, with individual differences in how sensory input and expectations combine.[21]

In generalization research, probability distributions have been used to explain how overlapping features of stimuli influence generalization.[22–24] For instance, when two stimuli share similar features, their representational distributions overlap, leading to a higher likelihood of generalization. However, all traditional theories assume the mental representation of a stimulus has a fixed shape and spread that remains constant over time and is

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

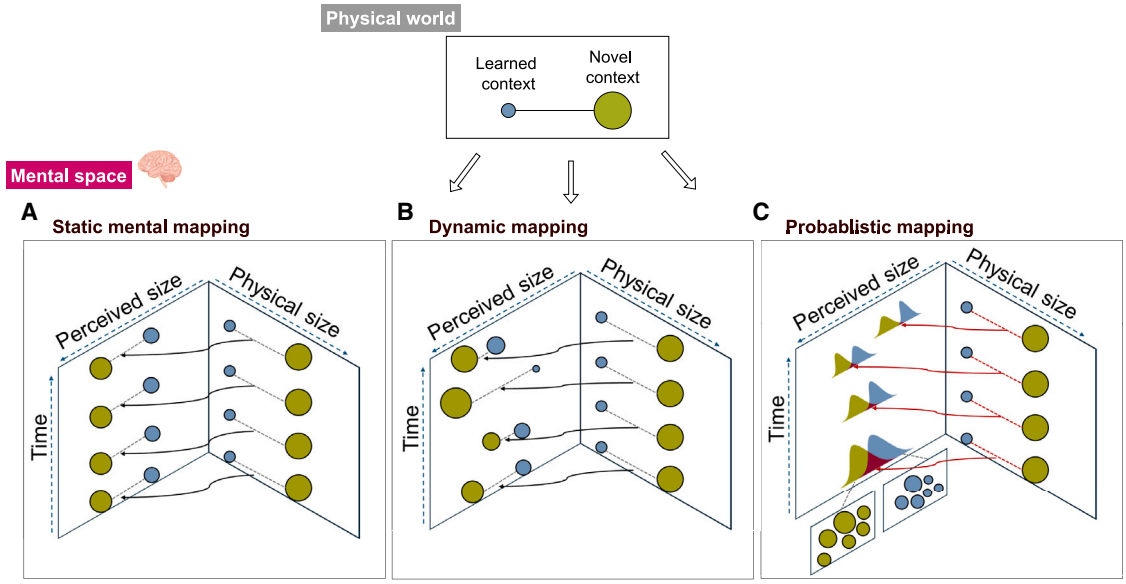


Figure 1. From physical world to mental space

(A) The mental distance between two contexts is determined by a one-to-one mapping between physical to perceptual quantity. (B) The mental distance has a temporal dynamic that shifts over time. (C) The mental distance has a dynamic and stochastic nature.

identical across different people (Figure 1A). This traditional approach faces a fundamental limitation: by constructing distributions solely from generalization behavior, it becomes impossible to disentangle the distinct cognitive components that shape generalization. This creates a critical theoretical inference problem, as researchers cannot determine whether observed generalization patterns stem from perceptual or memory processing differences. Given that overgeneralization has been consistently linked to anxiety-related disorders,[25–27] this theoretical limitation has profound clinical implications. Without the ability to isolate specific cognitive components driving generalization, interventions designed to modify generalization patterns may fail if they target the wrong cognitive component. For instance, a treatment approach focused on modifying response tendencies might prove ineffective if the patient's anxiety-related overgeneralization primarily stems from altered perceptual processing[28]. Previous research provided evidence for this perceptual account by showing that anxiety patients exhibited overgeneralization partly due to altered sensory representations.[29]

Recent empirical studies have begun addressing these limitations by revealing two key mechanisms underlying generalization behavior: perceptual processing[30–34] and memory operations.[35–38] These findings demonstrate the inadequacy of traditional generalization theories by showing how individuals may perceive the same physical features differently, and how their memories of these perceptions can shift over time (Figures 1B and 1C). For example, Zaman et al. (2023)[36] found large individual differences in how people perceive colors and identify previously learned stimuli, with these idiosyncratic patterns strongly predicting their generalization behavior. This dynamic variability suggests that to truly understand generalization and develop effective treatments, we need methodologies that can separately measure and model the distinct contributions of perceptual processing and memory mechanisms, rather than conflating them within a single distribution derived from behavioral outcomes alone.

Recently, we introduced a computational model[3] that made a first attempt to explicitly include perception into the generative process underlying fear generalization (Figure 2). It departs, like most generalization theories, from an error-driven learning process[39] that captures how organisms adapt their behavior based on experience. When an outcome differs from what was expected, learners update their predictions to minimize future errors – larger surprises lead to bigger adjustments in their expectations. Generalization of these learned expectations follows an exponential decay principle:[2] the more different two situations are, the less likely learning will transfer between them. This decay occurs within a mental space where stimuli are represented based on their perceptual features. In traditional theories, this mental space is assumed to be static and uniform – stimuli occupy fixed positions that remain constant over time and are identical across individuals. Our model departs from this assumption by using trial-by-trial perceptual data to track how individuals' mental representations of stimuli change over time.[3] This allows stimulus positions in mental space to shift dynamically and differ between individuals. Consequently, the similarity between any two stimuli, which is derived from their distances in this space, becomes dynamic rather than fixed. This enables the model to account for variations in generalization patterns based on how individuals uniquely represent stimuli in their mental space. Crucially, unlike previous approaches that inferred these representations from generalization behavior itself, our model derives them independently from trial-by-trial perceptual data, avoiding the circularity inherent in earlier theories.

While this model successfully accounts for individual differences in generalization behavior,[3] it nevertheless implements a

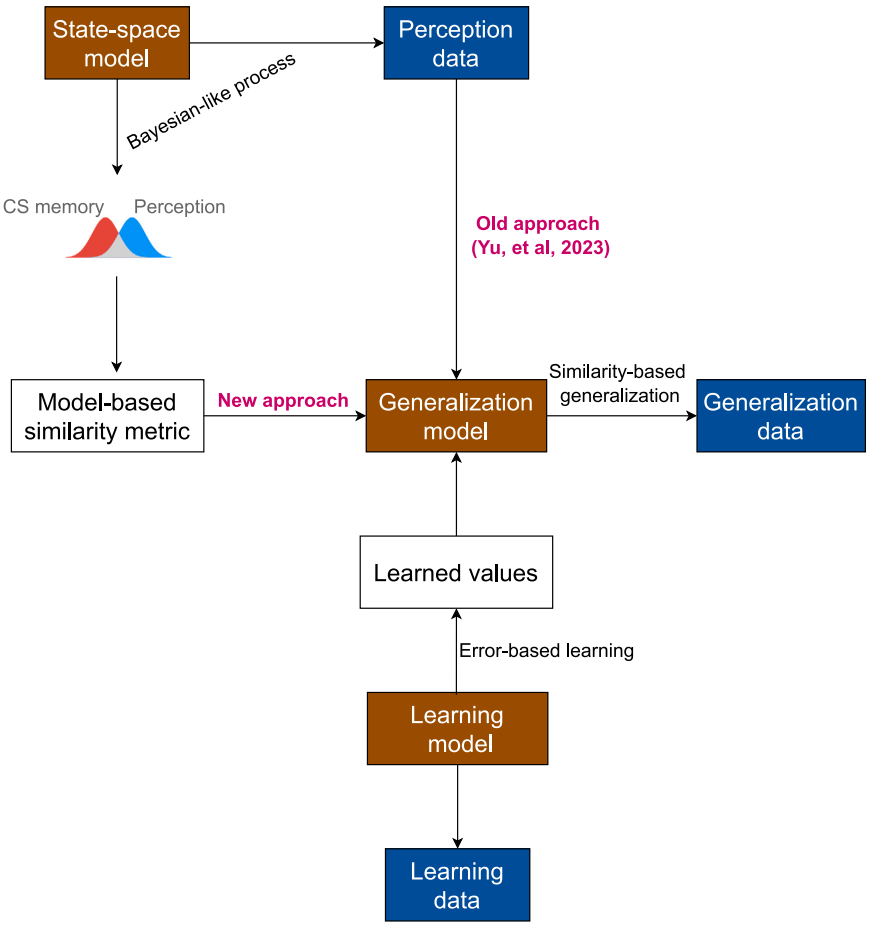


Figure 2. Model framework overview

The computational framework integrates a state-space model that processes perceptual data in a Bayesian-like manner, generating perceptual and CS memory distributions. These distributions inform a model-based similarity metric, which feeds into the generalization model alongside learned values from the learning model. This new approach contrasts with the previous point-based perceptual representation for similarity-based generalization.

reductionist approach to perception that warrants further refinement. The model's representation of perception as discrete points in mental space fails to capture two critical aspects of perceptual processing. First, it overlooks the inherent uncertainty in sensory processing, where each percept exists not as a singular value but as a probability distribution reflecting the brain's confidence in its interpretations.[13,14,20] Second, it does not account for the cumulative nature of perceptual experience, whereby current perceptions are actively shaped by the integration of past perceptual encounters with incoming sensory information.[17–20] This simplified conceptualization of perception, while enabling initial insights into individual differences in generalization, ultimately constrains our ability to fully elucidate the fundamental perceptual mechanisms that underlie generalization behavior and their dynamic interaction with learning processes.

In this study, we introduce a fundamental shift in understanding how perception shapes generalization by resolving a critical limitation in existing theories. While early theories assumed direct access to physical stimulus properties (Figure 1A), later approaches acknowledged uncertainty through probability distributions but could only infer these distributions from generalization behavior itself - creating a circular explanation. Our previous work[3] began addressing this circularity by incorporating direct perceptual measurements, but treated each perception as a precise point rather than an uncertain estimate (Figure 1B). We now present a framework that captures two fundamental aspects of human perception: its inherent uncertainty and its continuous evolution through experience (Figure 1C). By modeling perceptual representations as dynamic probability distributions derived from trial-by-trial perceptual judgments (Figure 2), we allow similarity to emerge naturally from the overlap between uncertain perceptual representations rather than inferring it from behavior or physical properties. To ensure rigorous comparison between probabilistic and traditional approaches, we implement a two-step analysis: first modeling perceptual uncertainty using only perceptual data, then using these pre-computed similarities in our generalization model. This separation prevents our probabilistic assumptions from biasing model comparison while still capturing how both immediate perception and accumulated experience shape generalization behavior.

To investigate these new assumptions, we re-analyzed data from two published fear conditioning experiments (total $N = 80$)[3] where participants rated the size of circular stimuli and their expectancy of receiving an electrocutaneous stimulus. In Experiment 1, participants underwent simple conditioning with one reinforced circle (CS+) paired with shock. In Experiment 2, differential conditioning was used with both a reinforced (CS+) and non-reinforced (CS−) circle. During generalization testing, participants rated their shock expectancy for new circles of varying sizes. Both experiments provided continuous measures of perception (size ratings) and fear learning (shock expectancy) throughout conditioning and generalization phases.

RESULTS

We contrasted two computational approaches that differed on their operationalization of the mental representation of stimuli. The first approach presumed stimuli as points with their coordinates directly derived from the perceptual judgment data (Figure 1B), replicating the methodology established in the previous

study.[3] The second approach conceptualized stimuli as probabilistic perceptual distributions that are the result of a Bayesian perception process that integrates past perceptual experiences and sensory input (Figure 1C).

In the following sections, we will first explore the inter- and intra-individual patterns observed in perceptual judgment data using the Bayesian perceptual model. Subsequently, we will explore how these perceptual patterns can be effectively used to predict generalization patterns.

Perception

Our investigation into perception centered on two key aspects: (1) understanding how individuals uniquely transform physical stimuli into subjective perceptual experiences through personalized sensory mappings (perceptual likelihood), and (2) examining the role of individual differences in accumulated perceptual history (perceptual prior) in shaping subsequent judgments. To model human perception, we employed a state-space framework using a Kalman filter—a mathematical model that captures how individuals integrate prior experiences with incoming sensory information. In essence, this approach reflects how the brain not only processes raw sensory input but dynamically combines it with past perceptions to form coherent judgments. The Kalman filter quantitatively accounts for the relative reliability of new sensory evidence and pre-existing perceptual beliefs, providing a principled mechanism to model this interplay (for detailed explanations, refer to section perceptual model).

Before delving into modeling the data collected from actual experiments, we initiated a simulation study where we generated perceptual responses for 300 synthetic participants. This simulation allowed us to explore parameter recovery and determine if the model could accurately identify parameter values. Encouragingly, the simulation results indicated successful recovery of all critical parameters within the models (Method S1: Parameter Recovery and Figure S1).

The architecture of the model centers on several key parameters that capture individual differences and temporal dynamics in perception. At the core are three participant-specific parameters: the scaling parameters ($\beta_{0,i}$ and $\beta_{1,i}$) define how physical stimulus properties are transformed into mental representations through a sigmoid function, with $\beta_{0,i}$ determining the baseline perceptual magnitude and $\beta_{1,i}$ controlling the steepness of the transformation. This initial transformation includes inherent sensory mapping uncertainty ($\sigma_{S,i}$), reflecting how consistently an individual perceives the same physical stimulus. The resulting perceptual distributions are characterized by their mean ($\mu_{\psi,ij}$) and uncertainty ($\sigma_{\psi,ij}$), which evolve over time. The rate of this evolution is controlled by the process noise decay parameter (ω_i), which determines how quickly perceptual uncertainty diminishes with repeated exposure to stimuli. A higher value of ω_i indicates a rapid decrease in uncertainty, leading to more precise perception over time, while a lower value maintains higher perceptual uncertainty throughout the experimental session. Together, these parameters provide a comprehensive account of both stable individual differences in perception and how perceptual precision dynamically evolves through experience (see Figure 3).

Sensory mapping

In Figure 4, we compared the perceptual responses observed in the experimental data with the replicated data generated from the posterior distribution of the parameter space. This comparison aimed to gauge the extent to which the perceptual model could accurately mimic the observed perceptual responses. As shown in Figure 4, the replicated data closely follows the patterns of actual perceptual responses, exhibiting consistency across different statistical quantiles.

With a multilevel structure, the model is capable of estimating these three perceptual parameters both at the group and individual levels, as illustrated in Figure 5. The parameters μ_{β_0} (mean of population-level β_0) and μ_{β_1} (mean of population-level β_1) encapsulate the group-level information regarding sensory mapping. In Experiment 1 and 2, μ_{β_0} displayed a 95% credible interval of [−3.31, −2.79] and [-2.99, −2.61], along with median values of −3.05 and −2.80. On the other hand, μ_{β_1}, in two experiments, displayed a 95% credible interval of [0.021, 0.026] and [0.018, 0.022], and the median values of 0.023 and 0.020. This parameter governs the overall steepness of the sensory mapping function. A larger value leads to a more pronounced and rapid change in the mapping process, making the function steeper around the midpoint of the logistic curve.

In the panel b of Figure 6, a diverse range of inter-individual variations becomes evident through the broad spectrum of $\beta_{0,i}$ and $\beta_{1,i}$ estimations. These variations give rise to unique patterns in how participants map physical stimuli to mental representations, as shown in Figure 6. Across both experiments, certain participants exhibit lower sensitivity to alterations in physical quantity, resulting in perceptual responses that are generally quite similar across different stimuli. Consequently, these participants are more likely to encounter confusion between different stimuli, given that minor perpetual uncertainties can already lead to overlapping mental representations. Conversely, another group of participants demonstrates heightened sensitivity to shifts in physical quantity, leading them to respond more distinctly to different stimuli and necessitating a greater degree of perceptual uncertainty to induce overlapping perceptual distributions.

Dynamic perception

Within each trial, prior perceptual expectations are combined with new sensory input to form the current perception. This integration is governed by the Kalman gain (α_ϕ), which determines the relative importance of prior expectations versus new sensory information. In essence, the Kalman gain adjusts the balance between relying on past experiences and incorporating fresh evidence. A low Kalman gain means prior expectations dominate perception when new sensory input is less reliable. Conversely, a high Kalman gain gives more weight to new input, shaping perception accordingly. This dynamic adjustment balances past and present information based on their relative reliability.

The model captures how perceptual uncertainty evolves dynamically over time through two key parameters. The first is a decay rate ω_i that governs how quickly an individual's perceptual uncertainty changes with repeated stimulus exposure. At the population level, this decay rate μ_ω showed similar estimates across both experiments, with 95% credible intervals of

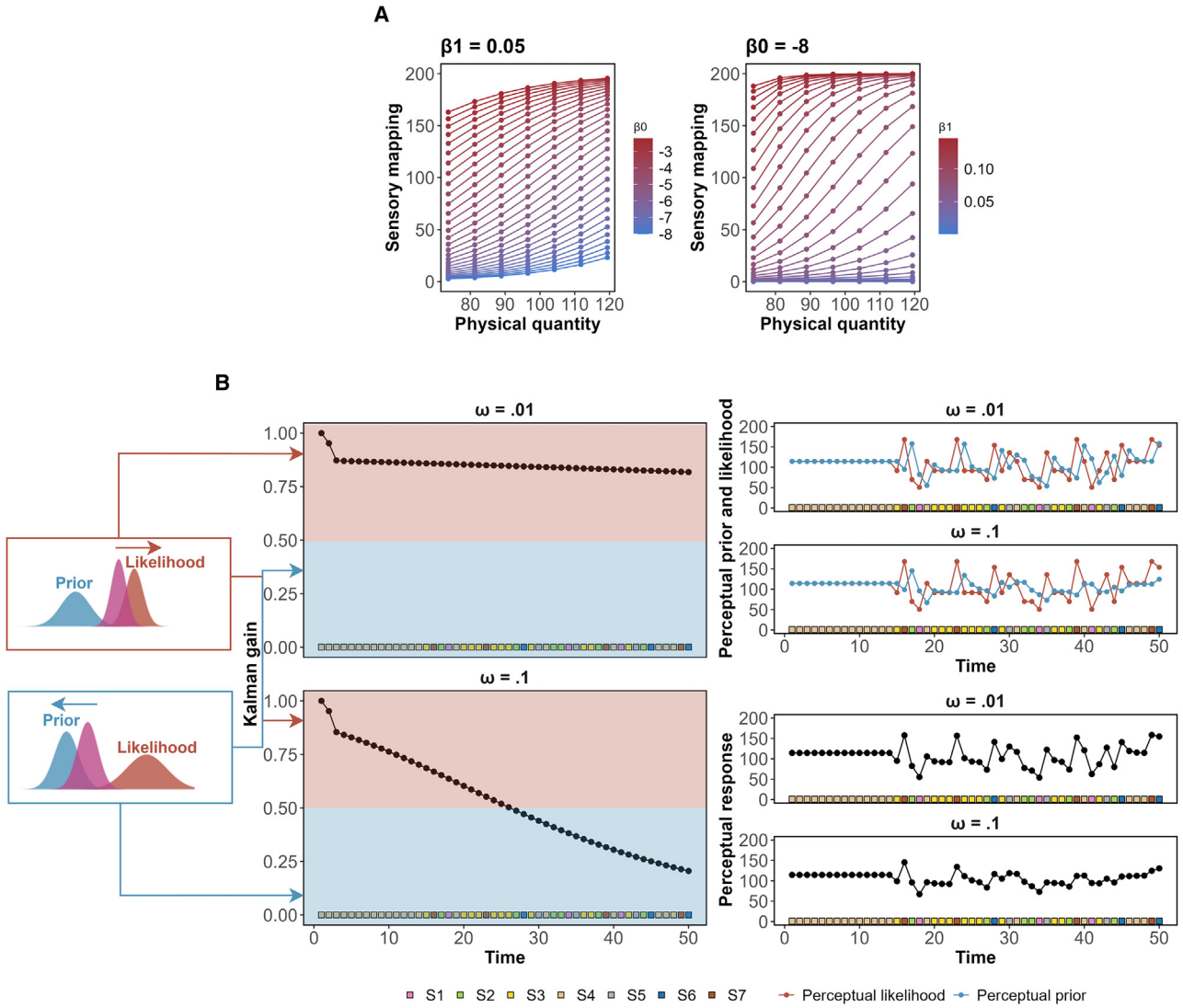


Figure 3. Patterns in perceptual model

(A) The combination of the two scaling parameters $\beta_{0,i}$ and $\beta_{1,i}$ in the sigmoid function can generate different patterns of perceptual mappings. (B) The determination of the perceptual mean at each time point involves a weighted combination of the mean values from the perceptual prior and the mean values from the perceptual likelihood (sensory mapping). The perception at a given time point is more influenced by the perceptual prior when the Kalman gain remains below 0.5, while it tends to be more heavily shaped by the perceptual likelihood as the Kalman gain surpasses 0.5. The effectiveness of this transition is intricately tied to the process noise forgetting rate denoted as ω, which essentially determines the pace at which the Kalman gain decreases with increasing perceptual instances.

[0.001,0.003] and [0.001,0.004] in Experiments 1 and 2 respectively, both having median values of 0.002. The second parameter η represents the initial magnitude of perceptual uncertainty at the group level, with 95% credible intervals of [10.63, 11.46] and [11.03, 12.34] for the two experiments, and median values of 11.04 and 11.68 respectively. Figure 7 illustrates how this perceptual uncertainty evolves throughout the experiments. Most participants showed a gradual decrease in uncertainty over time, indicating that repeated exposure to stimuli leads to more precise and stable perceptual representations.

The model also includes a sensory mapping uncertainty parameter $\sigma_{S,i}$ that captures individual differences in how consistently people perceive physical stimuli. This parameter reflects the stability of an individual's perceptual system - a lower value indicates more consistent perception of the same physical stimulus across repeated presentations. The majority of participants in both experiments demonstrated relatively stable perception with $\sigma_{S,i}$ values below 1. This perceptual stability influenced how individuals weighted new sensory information versus their prior perceptual experience, as measured by the Kalman gain which remained above 0.5 for most participants throughout the experiments (Method S2: Kalman Gain Analysis and Figure S2). The consistently high Kalman gain suggests that participants maintained more sensory-based perception of the stimuli,

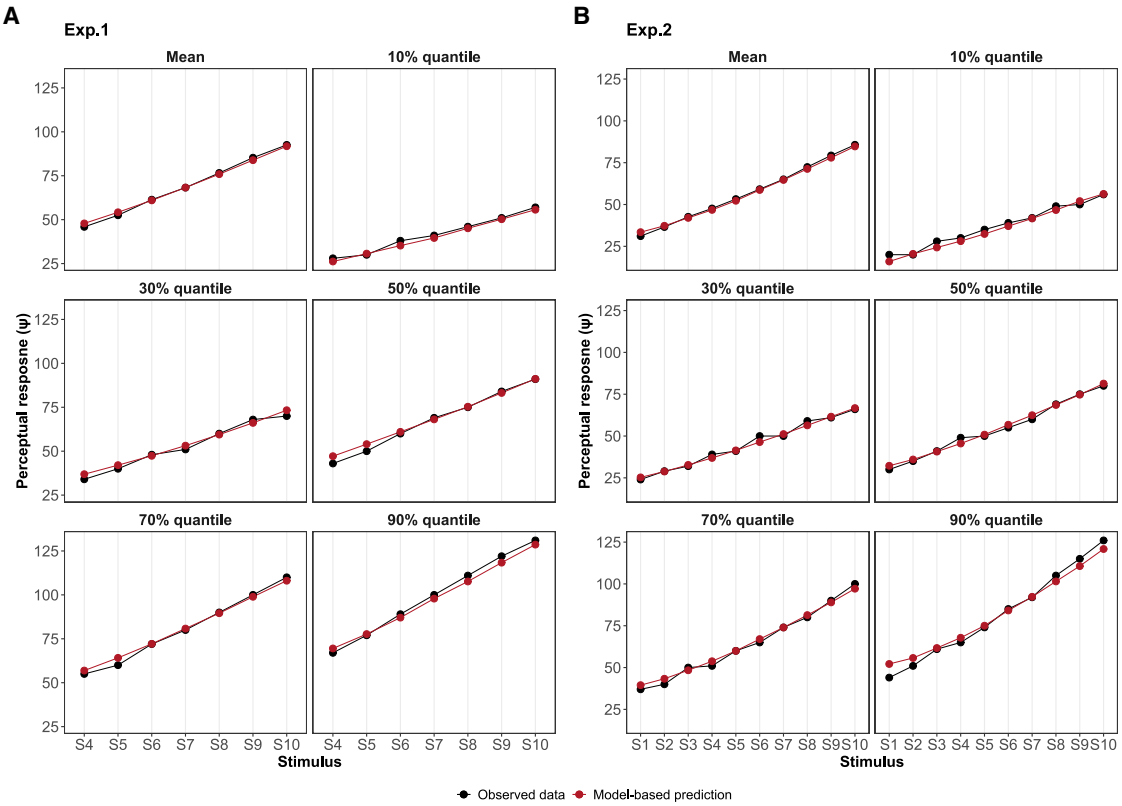


Figure 4. Posterior predictive checks of model-based perception

Comparisons are drawn between posterior predictive samples and actual perceptual response data in both simple conditioning (A) and differential conditioning (B) experiments. The black curve represents the mean as well as the quantiles at 10%, 30%, 50%, 70%, and 90% across diverse stimuli for observed perceptual responses. In comparison, the orange curve illustrates the mean along with corresponding quantiles at 10%, 30%, 50%, 70%, and 90% for 5000 replicated data across various stimuli.

integrating some influence from past perceptual experiences without converging on them too much.

Generalization

Upon scrutinizing the perceptual patterns evident in both experiments using the state-space perceptual model, our focus shifted to assessing how these different mental stimulus representations contributed to the observed generalization behavior. To accomplish this, we ran the computational model of generalization[3] twice. One run utilized a point-based approach where stimuli are considered as points in a mental space with their coordinates directly derived from perceptual data (Figure 1B). Here, stimulus similarity is derived by the Euclidean distance between coordinates, thereby ignoring the influence of perceptual uncertainty. In the new model, stimuli are represented as probability distributions that emerged from a Bayesian perception process, with now stimulus similarity being reflected in the overlap between two stimulus distributions (Figure 1C).

The generalization model integrates error-driven learning[39] and similarity-based generalization processes.[2] Through repeated exposures to the learned stimulus, stimulus-outcome associations are updated and decay exponentially with decreasing stimulus similarity. Depending on the model, this similarity is derived from point distances or by the overlap in perceptual distributions. To capture distinct patterns of generalization behavior, the model employs a mixture structure with four pathways: Non-Learners and Overgeneralizers represent maladaptive generalization patterns characterized by learning deficits or excessive generalization, respectively. Physical Generalizers and Perceptual Generalizers differentiate whether perceptual variability impacts generalization patterns (Perceptual Generalizers; Figures 1B and 1C) or not (Physical Generalizers; Figure 1A).

Model comparison

To assess the relative data fit, we integrated the two perceptual assumptions into a single model (referred to as a super model). Within this framework, we estimated the model selection parameter β_M, which provides information on which sub-assumption (model) better fits the data, and computed the Bayes factor based on the estimations of β_M.

Examining the posterior samples of β_M as depicted in Figure 8 and computed Bayes factors, we observe a distinct preference direction in Experiment 1. The data strongly favored the perceptual distributions model. In Experiment 1, the posterior distribution of β_M exhibited a 95% confidence

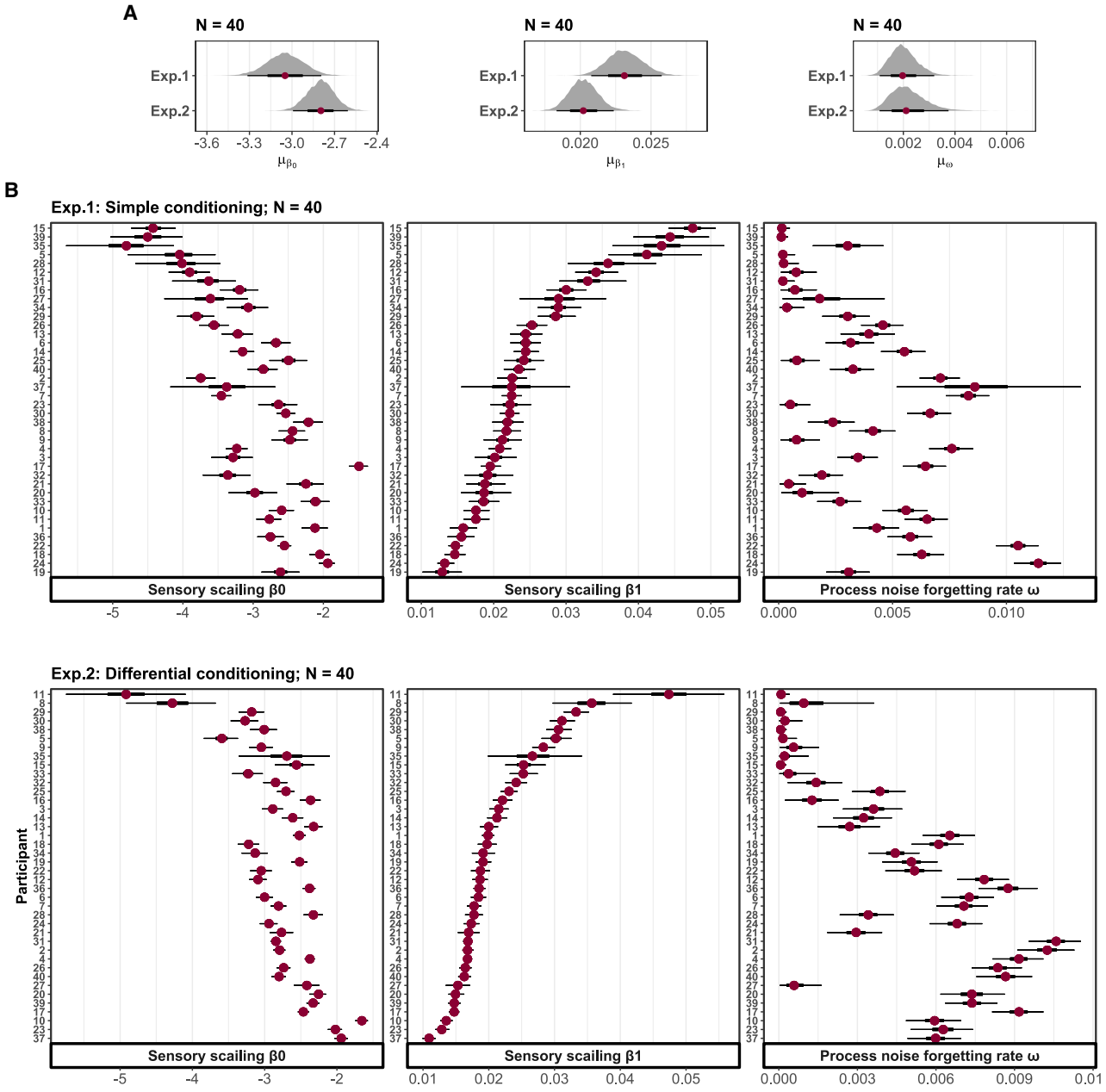


Figure 5. Perceptual model parameter estimates

Parameter estimates with 95% credible interval of the three important parameters in the perceptual model. (A) Posterior distributions of the group mean for the scaling parameter β_0, β_1, and ω. (B) Posterior distributions of the individual scaling parameter $\beta_{0,i}$, $\beta_{1,i}$, and ω_i (ordered by the 50% quantile of the perceptual slope parameter $\beta_{1,i}$).

interval of [0.59, 0.87], with a median value of 0.74 and a BF = 18.833. On the other hand, in Experiment 2, the data did not show a clear preference for either perceptual assumption, suggesting a more balanced performance between the model with two distinct perceptual assumptions in that context (i.e., the posterior distribution of β_M had a 95% confidence interval of [0.35, 0.65], with a median value of 0.50 and a BF = 0.197).

Latent group allocation

Next, we compared if the latent group allocation was affected by the different ways of stimulus representations. In Figure 9, it is shown that the use of stimulus distribution resulted in higher probability of having Perceptual Generalizers in both experiments (Experiment 1: 95% CI[0.28,0.63], median = 0.45; Experiment 2: 95% CI[0.51,0.80], median = 0.66) compared to the adoption of perceptual distance based on point estimates

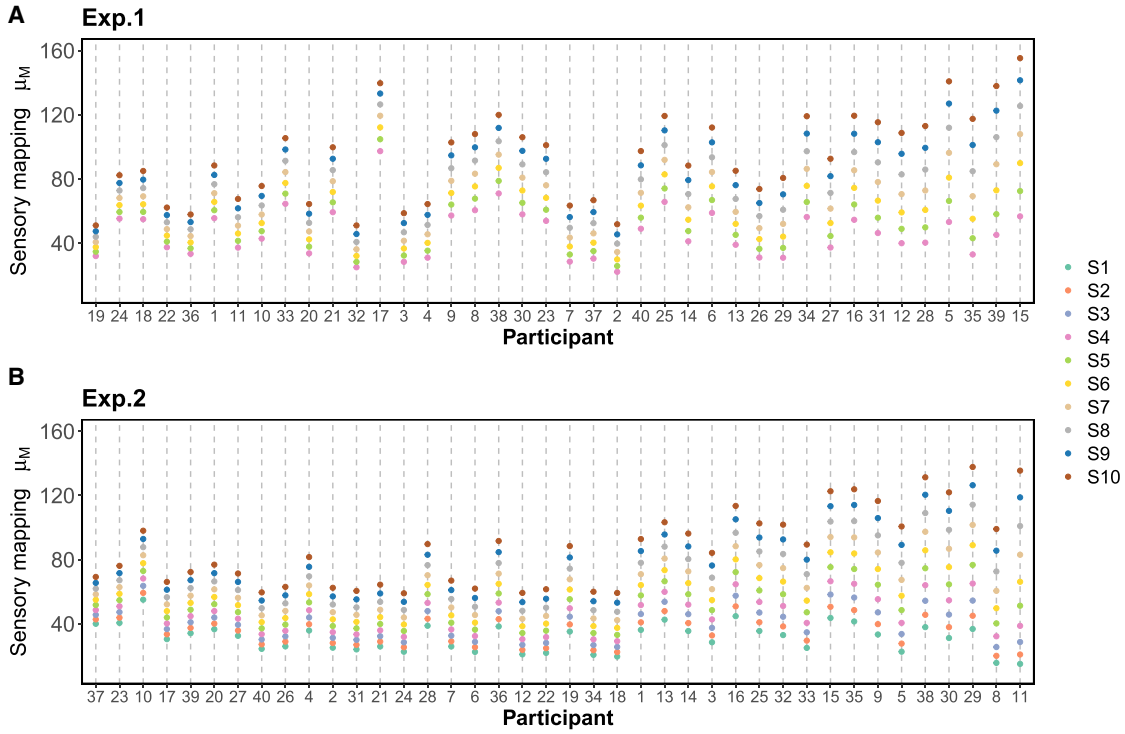


Figure 6. The sensory mapping patterns

For each physical stimulus, individual participants demonstrate a distinct mapping from physical to perceptual quantities. The mental representation of each experimental stimulus is calculated with the sensory mapping scaling parameters β_0 and β_1 (50th quantile from the MCMC samples, ordered by the 50% quantile of the perceptual slope parameter $\beta_{1,j}$). (A) and (B) display the mapping patterns for Experiment 1 (simple conditioning) and Experiment 2 (differential conditioning), respectively.

(Experiment 1: 95% CI[0.07,0.33], median = 0.17; Experiment 2: 95% CI[0.32,0.62], median = 0.47). An opposite pattern can be observed for the probability of Physical Generalizers allocation, with lower posterior probability in the new model (Experiment 1: 95% CI[0.02,0.24], median = 0.09; Experiment 2: 95% CI [0.03,0.23], median = 0.11) compared with the point-based approach (Experiment 1: 95% CI[0.23,0.56], median = 0.39; Experiment 2: 95% CI[0.18,0.46], median = 0.31). As to the probability of Non-Learners and Overgeneralizers allocation, as expected, it remains nearly identical regardless of the particular perceptual assumption (Method S3: Group Allocation Patterns and Figures S3 and S4).

DISCUSSION

The current study presents a framework for comparing two distinct approaches to modeling how humans mentally represent stimuli and its impact on fear generalization responding. The first approach, established in previous research,[3] acknowledges that perception deviates from physical reality but represents these perceptual differences as discrete points in mental space. Our new approach extends this by conceptualizing perception as probability distributions that emerge from the continuous interaction between sensory evidence and prior expectations. This shift from point-based (Figure 1B) to distribution-based (Figure 1C) representations captures additional complexities of human perception: the inherent uncertainty in perceptual processing, the individual difference in perceptual processing, and the dynamic evolution of distributional perception over time. By comparing these approaches, we demonstrate how incorporating probabilistic representations provides new insights into how perceptual processes shape fear generalization behavior.

Through a comparative analysis of the same generalization model utilizing distinct stimulus representations, we contrasted two fundamental approaches to quantifying perceptual similarity. Our model comparison revealed distinct patterns across the two experiments. In Experiment 1, we found strong evidence favoring the distribution-based over point-based perceptual similarity ($BF = 18.833$). Importantly, both approaches substantially outperformed models assuming veridical stimulus perception (Physical Generalizers). In Experiment 2, while the model comparison showed no clear preference between point-based and distribution-based approaches ($BF = 0.197$), both again proved superior to veridical perception models. The distribution-based approach led to a marked increase in the proportion of participants classified as Perceptual Generalizers in both experiments - from a median of 17%–45% in Experiment 1, and from 47% to 66% in Experiment 2. Correspondingly, we observed a decrease in Physical Generalizers (Experiment 1: from 39% to 9%; Experiment 2: from 31% to 11%). This consistent shift in group allocation across experiments, even when

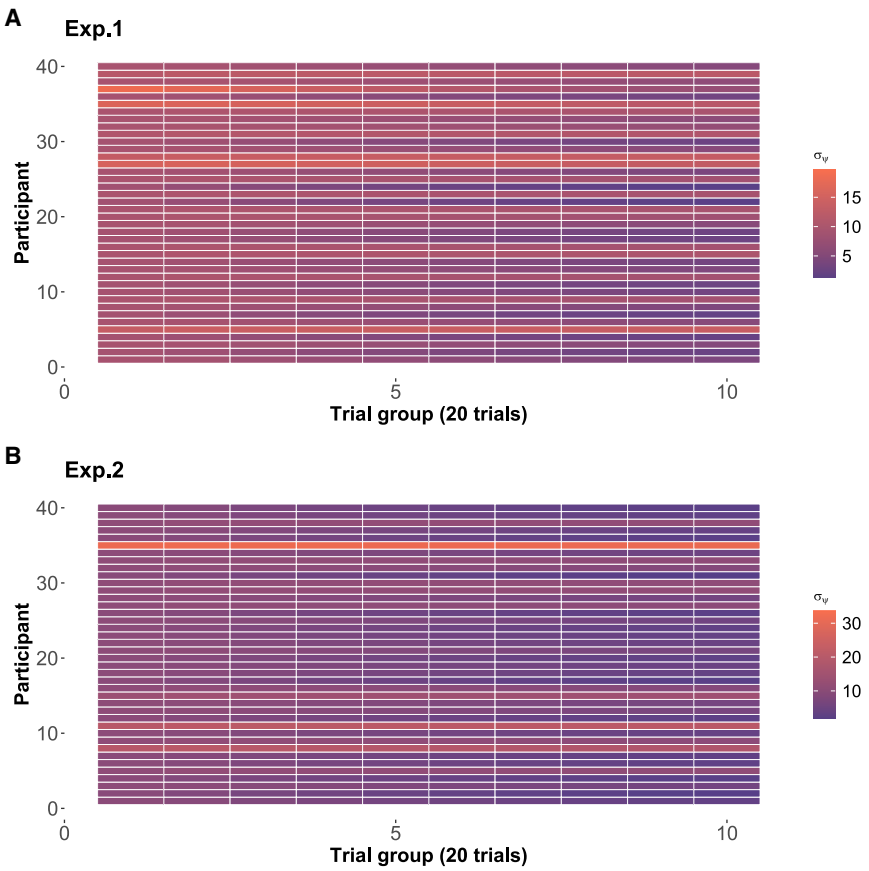


Figure 7. Perceptual uncertainty

The two panels illustrate the mean of perceptual uncertainty (from 50th quantile of MCMC samples) for every 20 trials in Experiment 1 (A) and Experiment 2 (B).

overall model performance was equivalent, suggests that modeling perceptual uncertainty reveals important individual differences in how perception shapes generalization. The different patterns of model preference between experiments might be attributed to several factors: the increased attention demands of differential conditioning could have led to more consistent perceptual responses, reducing the importance of modeling uncertainty, or participants might have exhibited greater variability in how they approached the differential conditioning task. Future research should systematically investigate these potential explanations to better understand how task demands shape the relationship between perceptual processes and generalization behavior, particularly focusing on when and why explicitly modeling perceptual uncertainty becomes crucial for understanding generalization patterns.

In the process of transitioning from learning associations within a specific context to applying that learning in new situations, humans continually engage with physical inputs. Yet, the findings within the realm of perception,[17–20] in conjunction with our current findings, make it evident that humans do not possess flawless perception. Instead, human perception operates through a non-linear sensory mapping, and individuals perceive the physical reality through the lens of inherent perceptual uncertainties. Regrettably, current theories of stimulus representation in learning and generalization, such as elemental representation models[22,23] or neural network models,[24] all presume that stimulus representation maps solely to the physical dimension. This assumption suggests that more similar physical features result in greater overlaps in the distribution of stimulus representations, ultimately falling short in capturing the idiosyncratic and dynamic nature of physical-to-perception mappings. The multidimensional scaling method,[2] while acknowledging the mapping from physical to perceptual space, assumes that perception for a stimulus is a fixed point in mental space, remaining invariant over time or individuals. This perspective, treating perception as a deterministic system, overlooks the inferential and probabilistic nature inherent in perception.

In this work, we advance the understanding of generalization by modeling perception as dynamic probability distributions rather than fixed points in mental space. These distributions capture three fundamental aspects of human perception: how individuals process incoming sensory information, how they integrate this information with past experiences, and how confident they are in their perceptions. When determining whether learning should transfer between stimuli, our framework examines the overlap between their perceptual distributions - a high degree of overlap suggests strong similarity, while minimal overlap indicates distinct percepts. This probabilistic approach reveals subtle but important phenomena that point-based models miss. For instance, two stimuli might be perceived quite similarly on average, yet rarely trigger generalization if people perceive them with high certainty, creating narrow distributions with minimal overlap. Conversely, even when average perceptions differ moderately, uncertain perception of both stimuli can create a focused region of distributional overlap, promoting consistent generalization. By capturing both the content of perception (where distributions are centered) and its quality (how precise they are), our framework provides new insights into why the same physical similarities between stimuli can lead to markedly different patterns of generalization across individuals and contexts.

Generalization holds a significant theoretical foundation not only within other cognitive domains such as learning[40,41] and memory,[42,43] it constitutes a foundational cognitive process that underpins a range of behaviors, including but not limited to categorization,[44] motor learning,[45] language processing,[46] and face recognition.[47] This serves to accentuate its pivotal role within the domain of cognitive science. While the

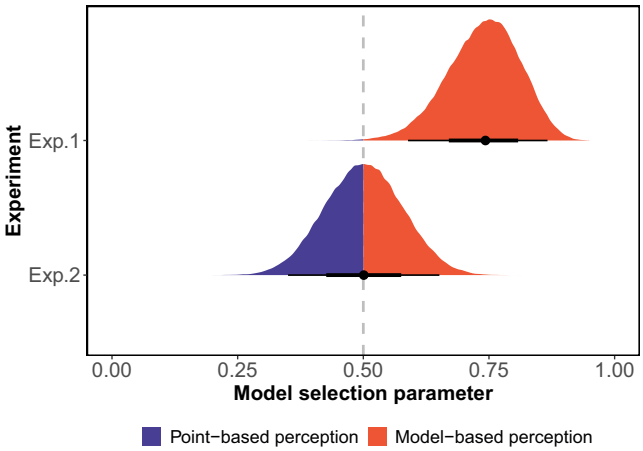


Figure 8. Posterior distributions of the model selection parameter

The two models, each based on different assumptions about perceptual distances, are combined into a single super model. The likelihoods of both models are integrated in a linear manner, with a model selection parameter denoted as β_M. The prior distribution for β_M is specified as $Uniform(1,1)$. The figure displays the posterior distribution of β_M, each comprising 120,000 samples. Values greater than 0.5 are indicated with the red color, suggesting that the observed data show a preference for the model-based perceptual assumption. Conversely, values less than 0.5 are shown in blue color, indicating that the observed data exhibit a preference toward the descriptive perceptual assumption.

significance of variability in the context of learning and generalization has been extensively deliberated,[48] the bulk of this discourse has predominantly centered around the physical domain. Yet, the exploration of variability within the human perceptual system and its consequential impact on relevant behavior remains relatively understudied. The increased proportion of individuals allocated to the Perceptual Generalizer group through the probabilistic perceptual similarity underscores the role of internal variability within the perceptual system in shaping observed generalization behavior.

The importance of perceptual variability, or more broadly, the intricacies of the perceptual process, also extends beyond theoretical implications to encompass clinical perspectives. Aberrant generalization has garnered empirical attention as a potential driving factor in anxiety disorders.[25–27] Moreover, evidence indicates that individuals with anxiety disorders exhibit poorer stimulus identification abilities compared to their healthy counterparts.[29] However, the causal direction between anxiety disorder symptoms and stimulus identification remains an open empirical question. Furthermore, the association between dynamic and probabilistic perceptual processes and clinical disorder symptoms remains largely unexplored, although there are emerging reports of altered perceptual inference processes in certain patient populations or personality traits.[21] In this context, the utilization of the current computational model not only provides insights into the associations between the generalization behavior of an individual and the underlying perceptual process but also into the dynamics through which the perceptual process engenders such behavior. For instance, individuals exhibiting a gradual decrease in Kalman gain over time tend to display an increasing bias toward previous perceptual experiences, a phenomenon with a profound impact observed in several mental disorders, contributing significantly to the overall shaping of perceptual patterns.[49–51]

Additionally, the perceptual sensitivity to different stimulus might also provide valuable insights, an example for this can be found in a recent study in which stimulus-based auditory perceptual plasticity is found to explain the overgeneralization behavior of the generalized anxiety disorder (GAD) patients.[29] Examining how diverse perceptual processes shape generalization behavior can provide clinicians with valuable insights not only into the extent to which problematic generalization behavior is influenced by biased perception but also into the workings of inherent perceptual processes that contribute to the eventual manifestation of such behavior. In sum, the knowledge of inherent perceptual process may offer clinicians a deeper understanding of the underlying mechanisms involved in anxiety disorders and aid in developing more targeted therapeutic interventions.[52]

Limitations of the study

Currently, there is no unified and coherent assumption regarding the optimality of human perception within the Bayesian framework.[15] Moreover, a consensus theory on how perceptual prior should be formulated remains elusive. Some studies have adopted a fixed perceptual prior distribution to represent a macro belief about the physical world,[17,19] while others have embraced a more experience-driven perceptual prior distribution.[18,53] In our study, we adopted a simple assumption for formulating the perceptual prior, considering past perceptual experiences as the sole prior for the current perceptual system. However, future research can delve deeper into the formulation of perceptual priors. For instance, in the context of fear generalization, the conditioned stimulus often carries more salient experiences, such as fear or pain. A pertinent question arises regarding whether these highly salient experiences contribute to the formulation of a macro perceptual prior that governs the dynamic updating process of perception. Further exploration and refinement of perceptual prior formulations could offer valuable insights into the complex interactions between perceptual experiences and the updating of perceptual systems. By addressing these issues, we can advance our understanding of how human perception adapts to various contexts and experiences, paving the way for more comprehensive models of perceptual processes and their implications for generalization behaviors. Moreover, delving deeper into the investigation of the perceptual prior can also aid in comprehending the influence of fear learning on the modification of the perceptual system,[54] particularly considering the lack of consistent empirical behavioral findings in this domain.[55]

Another limitation of the current study pertains to its exclusive reliance on self-report responses to assess fear learning, perceptual, and fear generalization processes. While there is some evidence suggesting a strong correlation between self-report responses and physiological or neuronal measures in fear learning, an ongoing debate remains regarding the extent to which we can achieve a consensus on fear-related behavior[56,57] through diverse response channels.[58,59] This unsolved debate can also apply to the field of perception.[60] Therefore,

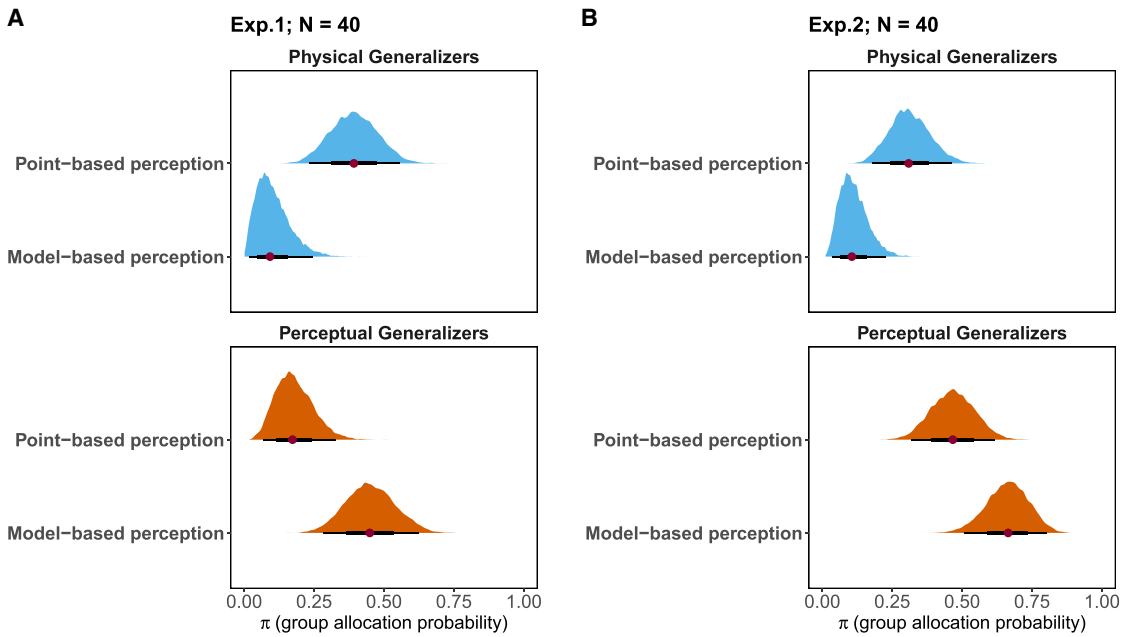


Figure 9. Posterior MCMC samples of group allocation parameter

The proportion of the posterior MCMC samples of the group allocation parameter in Experiment 1 (A) and 2 (B). The point distance distribution depicts the posterior samples when the model adopts descriptive perceptual patterns. In this approach, the current perception is calculated as the absolute difference between the current perceptual response and the cumulative average of the perceptual response to the conditioned stimulus. On the other hand, the modeled distance distribution illustrates the posterior samples when the model incorporates model-based perceptual patterns, integrating probabilistic and dynamical assumptions of perception.

we may expect a more complicated relationship when we attempt to investigate both processes at the same time. Previous empirical studies have demonstrated the impact of perceptual variability on generalization with startle eye-blink responses[31]—a physiological manifestation of the body's automatic reaction to sudden stimuli. The findings of our study hold potential for broader generalizability with such a physiological measure. Furthermore, there is a need for further exploration of the generalizability of the modeling findings to different types of learning paradigms, stimuli with elevated levels of complexity, and multi-dimensional features.

In the present study, the determination of similarity between the currently encountered stimulus and the previously learned stimulus relies solely on perceptual data. However, this approach assumes a precise alignment between the perceptual memory associated with the learned stimulus and the perceptual encoding captured during the initial encounter. Regrettably, this assumption overlooks the potential influence of memory processes, which can exert a significant impact on both the encoding and retrieval of memories. Inter-individual and intra-individual differences during encoding, retrieval, or mechanisms acting on these processes may introduce substantial biases in memory that can deviate considerably from the objective representation of stimuli.[61,62] Consequently, the exclusive reliance on perceptual data to determine perceptual distance may fall short in capturing the complexities and variations introduced by memory processes. This situation calls for further investigation and consideration in the study of generalization behaviors. In future studies, it is encouraged to develop models that incorporate recent theories of human memory and collect memory data throughout generalization testing.[63] This approach would provide a more accurate representation of how individuals perceive and memorize the physical world, thereby advancing our understanding of the intricate interplay between perceptual and memory processes and their combined impact on human generalization behavior.

Conclusion

Human behavior is a multifaceted phenomenon frequently involving multiple psychological processes and mechanisms. Generalization, as one such behavior, manifests observable traits that can emerge from distinct mechanisms. The comprehension of the congruent relationship between diverse cognitive mechanisms and generalization behavior necessitates a deeper understanding. In this research, we formulate a modeling framework to integrate perceptual mechanisms characterized by their probabilistic and dynamic nature into the process of generalization. This framework enables us to perceive generalization behavior as an outcome arising from the intricate interplay among learning, perception, and generalization mechanisms. Consequently, it affords us a more profound insight into how humans generalize, not merely due to an inherent inclination to transfer previous learning but also owing to the profound interaction between cognition and perception of physical reality. Ultimately, this integrated approach may contribute to the refinement of generalization theories and advance our knowledge of the fundamental mechanisms that shape human generalization behavior.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Kenny Yu (kenny.yu@kuleuven.be).

Materials availability

This study did not generate new data. All data used in this study were from previously published fear conditioning experiments, as described in Yu et al. (2023).[3]

Data and code availability

- **Data:** The fear conditioning experiment data (total $N = 80$) have been deposited at the Open Science Framework repository as https://doi.org/10.17605/OSF.IO/SXJAK and are publicly available as of the date of publication (see also Yu et al. (2023)[3]).
- **Code:** All original code, including the computational models (perceptual and generalization models) and the scripts for simulations, statistical inference, and analysis (implemented in R 4.1.1 and JAGS 4.3.1), have been deposited at the Open Science Framework repository as https://doi.org/10.17605/OSF.IO/SXJAK and are publicly accessible as of the date of publication.
- **Additional Information:** Supplementary information for the modeling details are available at the Open Science Framework repository as https://doi.org/10.17605/OSF.IO/SXJAK and are publicly accessible as of the date of publication.

ACKNOWLEDGMENTS

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. , W.F., 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.

AUTHOR CONTRIBUTIONS

Conceptualization, K.Y., W.F., F.T., and J.Z.; methodology, K.Y., W.F., F.T., and J.Z.; software, K.Y.; formal analysis, K.Y.; investigation, K.Y.; resources, K.Y.; data curation, K.Y.; writing – original draft, K.Y.; writing – review and editing, K.Y., W.F., F.T., and J.Z.; visualization, K.Y.; supervision, F.T. and J.Z.; funding acquisition, J.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human participants
- METHOD DETAILS
  - Experiments
  - Computational model of generalization
  - Model architecture
  - New perceptual stimulus distance
  - Perceptual model
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.112228.

Received: May 3, 2024
Revised: January 7, 2025
Accepted: March 12, 2025
Published: March 17, 2025

REFERENCES

1. Ghirlanda, S., and Enquist, M. (2003). A century of generalization. Animal Behaviour *66*, 15–36. https://doi.org/10.1006/anbe.2003.2174.
2. Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. Science *237*, 1317–1323. https://doi.org/10.1126/science.3629243.
3. Yu, K., Tuerlinckx, F., Vanpaemel, W., and Zaman, J. (2023). Humans display interindividual differences in the latent mechanisms underlying fear generalization behaviour. Commun. Psychol. *1*, 5. https://doi.org/10.1038/s44271-023-00005-0.
4. Honig, W.K., and Urcuioli, P.J. (1981). The legacy of guttman and kalish (1956): Twenty-five years of research on stimulus generalization. J. Exp. Anal. Behav. *36*, 405–445. https://doi.org/10.1901/jeab.1981.36-405.
5. Estes, W.K., and Burke, C.J. (1953). A theory of stimulus variability in learning. Psychol. Rev. *60*, 276–286. https://doi.org/10.1037/h0055775.
6. Zaman, J., Chalkia, A., Zenses, A.-K., Bilgin, A.S., Beckers, T., Vervliet, B., and Boddez, Y. (2021). Perceptual variability: Implications for learning and generalization. Psychon. Bull. Rev. *28*, 1–19. https://doi.org/10.3758/s13423-020-01780-1.
7. Purves, D., Monson, B.B., Sundararajan, J., and Wojtach, W.T. (2014). How biological vision succeeds in the physical world. Proc. Natl. Acad. Sci. USA *111*, 4750–4755. https://doi.org/10.1073/pnas.1311309111.
8. Wei, X.-X., and Stocker, A.A. (2017). Lawful relation between perceptual bias and discriminability. Proc. Natl. Acad. Sci. USA *114*, 10244–10249. https://doi.org/10.1073/pnas.1619153114.
9. Shepard, R.N. (1994). Perceptual-cognitive universals as reflections of the world. Psychon. Bull. Rev. *1*, 2–28. https://doi.org/10.3758/BF03200759.
10. Spratling, M.W. (2016). Predictive coding as a model of cognition. Cogn. Process. *17*, 279–305. https://doi.org/10.1007/s10339-016-0765-6.
11. Friston, K.J., and Stephan, K.E. (2007). Free-energy and the brain. Synthese *159*, 417–458. https://doi.org/10.1007/s11229-007-9237-y.
12. Knill, D.C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. *27*, 712–719. https://doi.org/10.1016/j.tins.2004.10.007.
13. Funamizu, A., Kuhn, B., and Doya, K. (2016). Neural substrate of dynamic Bayesian inference in the cerebral cortex. Nat. Neurosci. *19*, 1682–1689. https://doi.org/10.1038/nn.4390.
14. Yon, D., and Frith, C.D. (2021). Precision and the Bayesian brain. Curr. Biol. *31*, R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044.
15. Acerbi, L., Vijayakumar, S., and Wolpert, D.M. (2014). On the Origins of Suboptimality in Human Probabilistic Inference. PLoS Comput. Biol. *10*, e1003661. https://doi.org/10.1371/journal.pcbi.1003661.
16. Stengård, E., and van den Berg, R. (2019). Imperfect Bayesian inference in visual perception. PLoS Comput. Biol. *15*, e1006465. https://doi.org/10.1371/journal.pcbi.1006465.
17. Weiss, Y., Simoncelli, E.P., and Adelson, E.H. (2002). Motion illusions as optimal percepts. Nat. Neurosci. *5*, 598–604. https://doi.org/10.1038/nn0602-858.
18. Adams, W.J., Graf, E.W., and Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. Nat. Neurosci. *7*, 1057–1058. https://doi.org/10.1038/nn1312.

19. Stocker, A.A., and Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. Nat. Neurosci. *9*, 578–585. https://doi.org/10.1038/nn1669.

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

21. Hoskin, R., Berzuini, C., Acosta-Kane, D., El-Deredy, W., Guo, H., and Talmi, D. (2019). Sensitivity to pain expectations: A Bayesian model of individual differences. Cognition *182*, 127–139. https://doi.org/10.1016/j.cognition.2018.08.022.

22. McLaren, I.P.L., and Mackintosh, N.J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. Animal Learning & Behavior *28*, 211–246. https://doi.org/10.3758/BF03200258.

23. Mclaren, I.P.L., and Mackintosh, N.J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. Anim. Learn. Behav. *30*, 177–200. https://doi.org/10.3758/BF03192828.

24. Ghirlanda, S., and Enquist, M. (2007). How training and testing histories affect generalization: a test of simple neural networks. Philos. Trans. R. Soc. Lond. B Biol. Sci. *362*, 449–454. https://doi.org/10.1098/rstb.2006.1972.

25. Dymond, S., Dunsmoor, J.E., Vervliet, B., Roche, B., and Hermans, D. (2015). Fear generalization in humans: Systematic review and implications for anxiety disorder research. Behav. Ther. *46*, 561–582. https://doi.org/10.1016/j.beth.2014.10.001.

26. Lissek, S., Bradford, D.E., Alvarez, R.P., Burton, P., Espensen-Sturges, T., Reynolds, R.C., and Grillon, C. (2014). Neural substrates of classically conditioned fear-generalization in humans: A parametric fMRI study. Soc. Cogn. Affect. Neurosci. *9*, 1134–1142. https://doi.org/10.1093/scan/nst096.

27. Fraunfelter, L., Gerdes, A.B.M., and Alpers, G.W. (2022). Fear one, fear them all: A systematic review and meta-analysis of fear generalization in pathological anxiety. Neurosci. Biobehav. Rev. *139*, 104707. https://doi.org/10.1016/j.neubiorev.2022.104707.

28. Struyf, D., Zaman, J., Vervliet, B., and Van Diest, I. (2015). Perceptual discrimination in fear generalization: Mechanistic and clinical implications. Neurosci. Biobehav. Rev. *59*, 201–207. https://doi.org/10.1016/j.neubiorev.2015.11.004.

29. Laufer, O., Israeli, D., and Paz, R. (2016). Behavioral and neural mechanisms of overgeneralization in anxiety. Curr. Biol. *26*, 713–722. https://doi.org/10.1016/j.cub.2016.01.023.

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

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

32. Zaman, J., Struyf, D., Ceulemans, E., Vervliet, B., and Beckers, T. (2021). Perceptual errors are related to shifts in generalization of conditioned responding. Psychol. Res. *85*, 1801–1813. https://doi.org/10.1007/s00426-020-01345-w.

33. Zaman, J., Yu, K., and Lee, J.C. (2023). Individual differences in stimulus identification, rule induction, and generalization of learning. J. Exp. Psychol. Learn. Mem. Cogn. *49*, 1004–1017. https://doi.org/10.1037/xlm0001153.

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

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

36. Zaman, J., Yu, K., and Verheyen, S. (2023). The idiosyncratic nature of how individuals perceive, represent, and remember their surroundings and its impact on learning-based generalization. J. Exp. Psychol. Gen. *152*, 2345–2358. https://doi.org/10.1037/xge0001403.

37. Robertson, E.M. (2018). Memory instability as a gateway to generalization. PLoS Biol. *16*, e2004633. https://doi.org/10.1371/journal.pbio.2004633.

38. Jasnow, A.M., Cullen, P.K., and Riccio, D.C. (2012). Remembering another aspect of forgetting. Front. Psychol. *3*, 175. https://doi.org/10.3389/fpsyg.2012.00175.

39. Rescorla, R., and Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, *2* (Appleton- Century-Crofts), pp. 64–99.

40. Wheeler, D.S., Amundson, J.C., and Miller, R.R. (2006). Generalization Decrement in Human Contingency Learning. Q. J. Exp. Psychol. *59*, 1212–1223. https://doi.org/10.1080/17470210600576342.

41. Brandon, S.E., Vogel, E.H., and Wagner, A.R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. Behav. Brain Res. *110*, 67–72. https://doi.org/10.1016/S0166-4328(99)00185-0.

42. Brown, G.D.A., Neath, I., and Chater, N. (2007). A temporal ratio model of memory. Psychol. Rev. *114*, 539–576. https://doi.org/10.1037/0033-295X.114.3.539.

43. Kumaran, D., and McClelland, J.L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. Psychol. Rev. *119*, 573–616. https://doi.org/10.1037/a0028681.

44. Jäkel, F., Schölkopf, B., and Wichmann, F.A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. Psychon. Bull. Rev. *15*, 256–271. https://doi.org/10.3758/PBR.15.2.256.

45. Braun, D.A., Aertsen, A., Wolpert, D.M., and Mehring, C. (2009). Motor Task Variation Induces Structural Learning. Curr. Biol. *19*, 352–357. https://doi.org/10.1016/j.cub.2009.01.036.

46. Singh, L. (2008). Influences of high and low variability on infant word recognition. Cognition *106*, 833–870. https://doi.org/10.1016/j.cognition.2007.05.002.

47. Balas, B., and Saville, A. (2015). N170 face specificity and face memory depend on hometown size. Neuropsychologia *69*, 211–217. https://doi.org/10.1016/j.neuropsychologia.2015.02.005.

48. Raviv, L., Lupyan, G., and Green, S.C. (2022). How variability shapes learning and generalization. Trends Cogn. Sci. *26*, 462–483. https://doi.org/10.1016/j.tics.2022.03.007.

49. Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P.R., Goodyer, I.M., and Fletcher, P.C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc. Natl. Acad. Sci. USA *112*, 13401–13406. https://doi.org/10.1073/pnas.1503916112.

50. O'Callaghan, C., Hall, J.M., Tomassini, A., Muller, A.J., Walpola, I.C., Moustafa, A.A., Shine, J.M., and Lewis, S.J.G. (2017). Visual Hallucinations Are Characterized by Impaired Sensory Evidence Accumulation: Insights From Hierarchical Drift Diffusion Modeling in Parkinson's Disease. Biol. Psychiatry. Cogn. Neurosci. Neuroimaging *2*, 680–688. https://doi.org/10.1016/j.bpsc.2017.04.007.

51. Corlett, P.R., Horga, G., Fletcher, P.C., Alderson-Day, B., Schmack, K., and Powers, A.R. (2019). Hallucinations and Strong Priors. Trends Cogn. Sci. *23*, 114–127. https://doi.org/10.1016/j.tics.2018.12.001.

52. Powers, A.R., Mathys, C., and Corlett, P.R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science *357*, 596–600. https://doi.org/10.1126/science.aan3458.

53. Miyazaki, M., Nozaki, D., and Nakajima, Y. (2005). Testing Bayesian Models of Human Coincidence Timing. J. Neurophysiol. *94*, 395–399. https://doi.org/10.1152/jn.01168.2004.

54. McGann, J.P. (2015). Associative learning and sensory neuroplasticity: how does it happen and what is it good for? Learn. Mem. *22*, 567–576. https://doi.org/10.1101/lm.039636.115.

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

56. Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., and Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. J. Behav. Ther. Exp. Psychiatry *44*, 201–206. https://doi.org/10.1016/j.jbtep.2012.08.003.

57. Constantinou, E., Purves, K.L., McGregor, T., Lester, K.J., Barry, T.J., Treanor, M., Craske, M.G., and Eley, T.C. (2021). Measuring fear: Association among different measures of fear learning. J. Behav. Ther. Exp. Psychiatry *70*, 101618. https://doi.org/10.1016/j.jbtep.2020.101618.

58. Lipp, O.V., and Purkis, H.M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. Cogn. Emot. *19*, 269–282. https://doi.org/10.1080/02699930441000319.

59. LeDoux, J.E., and Brown, R. (2017). A higher-order theory of emotional consciousness. Proc. Natl. Acad. Sci. USA *114*, E2016–E2025. https://doi.org/10.1073/pnas.1619316114.

60. Rossi, G.B., and Berglund, B. (2011). Measurement involving human perception and interpretation. Measurement *44*, 815–822. https://doi.org/10.1016/j.measurement.2011.01.016.

61. Brady, T.F., and Alvarez, G.A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. Psychol. Sci. *22*, 384–392. https://doi.org/10.1177/0956797610397956.

62. Luck, S.J., and Vogel, E.K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. Trends Cogn. Sci. *17*, 391–400. https://doi.org/10.1016/j.tics.2013.06.006.

63. Yu, K., Vanpaemel, W., Tuerlinckx, F., and Zaman, J. (2024). The representational instability in the generalization of fear learning. NPJ Sci. Learn. *9*, 78. https://doi.org/10.1038/s41539-024-00287-x.

64. Spence, K.W. (1937). The differential response in animals to stimuli varying within a single dimension. Psychol. Rev. *44*, 430–444. https://doi.org/10.1037/h0062885.

65. Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. J. Basic Eng. *82*, 35–45. https://doi.org/10.1115/1.3662552.

66. Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003) (Austria: Vienna). March 20-22.

67. R Core Team, R (2021). A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

68. Kellner, K. (2021). jagsUI: A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses. r package version *1.5.2*.

69. Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. Stat. Sci. *7*, 1177011136. https://doi.org/10.1214/ss/1177011136.

70. Brooks, S.P., and 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.

71. Mootoovaloo, A., Bassett, B.A., and Kunz, M. (2016). Bayes Factors via Savage-Dickey Supermodels. Preprint at: arXiv. https://doi.org/10.48550/arXiv.1609.02186

72. Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. J. Am. Stat. Assoc. *90*, 773–795. https://doi.org/10.1080/01621459.1995.10476572.

73. Berger, J.O., and Pericchi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. J. Am. Stat. Assoc. *91*, 109–122. https://doi.org/10.1080/01621459.1996.10476668.

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

75. Dickey, J.M. (1971). The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. Ann. Math. Statist. *42*, 204–223. https://doi.org/10.1214/aoms/1177693507.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Fear conditioning experiment data	Open Science Framework (OSF)	https://doi.org/10.17605/OSF.IO/SXJAK
Software and algorithms		
Analysis codes	Open Science Framework (OSF)	https://doi.org/10.17605/OSF.IO/SXJAK
Other		
Supplementary Information	Open Science Framework (OSF)	https://doi.org/10.17605/OSF.IO/SXJAK

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No new experimental data were collected for the current study. Instead, we conducted a re-analysis of two previously published datasets from Yu et al. (2023).[3] Below, we provide a summary of the original experimental procedures and participant details for context.

Human participants

The original study included a total of 80 healthy adult participants, with 40 participants in each of two experiments. Both experiments included 40 participants (65% female, 35% male), with mean ages of 22.0 years (SD = 5.3) in Experiment 1 and 24.0 years (SD = 8.9) in Experiment 2. All participants were recruited via KU Leuven's online experiment management system and received either course credits or monetary compensation (€12 for Experiment 1 and €16 for Experiment 2). The study was approved by KU Leuven's Social and Societal Ethics Committee (Approval Reference: G-201610641) and all participants provided written informed consent prior to participation. While the current study did not examine potential sex or gender effects on fear generalization, a recent investigation using a comparable experimental paradigm found no gender differences in fear generalization behavior and underlying processes.[34]

In the original study, sample size was determined based on previous studies investigating fear generalization. In Experiment 1 (simple conditioning paradigm), all participants underwent the same experimental procedure. In Experiment 2 (differential conditioning paradigm), participants were counterbalanced in their assignment of CS+ and CS-, where half of the participants received the smallest circle (S1; 50.8 mm) as CS+ and the largest circle (S10; 119.42 mm) as CS-, while the other half received the opposite assignment.

The original experimental procedures received ethical approval from the Social and Societal Ethics Committee at KU Leuven (Approval Reference: G-201610641).

METHOD DETAILS

This study presents a computational reanalysis of previously published experimental datasets[3] using a two-step modeling approach that separates perceptual and generalization processes. First, we applied principles from Bayesian perception theory to model participants' trial-by-trial perceptual judgments, capturing how physical stimuli are transformed into probabilistic mental representations. From this perceptual model, we derived a new perceptual distance metric based on distribution overlaps, which was then incorporated into our previously established generalization model[3] to predict fear learning responses. This approach allowed us to evaluate how different assumptions about perception influence generalization behavior while maintaining a tractable modeling framework.

Importantly, we did not model generalization itself as a Bayesian process; instead, we applied Bayesian principles only to model perceptual processes, and then used the outputs from this perceptual model to derive distances that served as inputs to our established generalization model.[3] This two-step approach allowed us to compare how different perceptual assumptions shaped generalization behavior while maintaining a consistent generalization framework. By separating the perceptual modeling from the generalization process, we ensured that the probabilistic nature of our perceptual model did not inadvertently bias our comparison of different perceptual assumptions.

For complete experimental details of the original data collection, see Yu et al. (2023).[3]

Experiments

The datasets used in this reanalysis came from two experiments that each comprised 40 participants. In both original experiments, participants underwent an acquisition phase, where they learned the associations between conditioned stimuli (CSs) and an unconditioned stimulus (US), followed by a generalization phase where test stimuli (TSs) were presented. The CSs and TSs were represented by circles with white outlines against a black background. The unconditioned stimulus used in both experiments was a noxious electrocutaneous stimulus.

The overall stimulus set of CSs and TSs comprised ten circles, labelled as S1 to S10, with diameters ranging from 50.80 to 119.42 mm, increasing by 7.624 mm between each step. In Experiment 1, which employed a simple conditioning paradigm, a subset of seven circles (S4 to S10) was utilized. The middle circle (S7; 96.54 mm) served as the CS+, while the remaining six stimuli were designated as TSs. In Experiment 2, a differential conditioning paradigm was employed, utilizing the entire stimulus set (S1 to S10). The assignment of the CS+ and CS- was counterbalanced among participants. The smallest circle (S1; 50.8 mm) and the largest circle (S10; 119.42 mm) alternated between being the CS+ or CS-. The remaining eight stimuli exclusively served as TSs and were solely presented during the generalization phase.

In each trial of both experiments, participants were required to provide ratings on two scales. Firstly, they rated the diameter of the stimulus using a size Visual Analogue Scale (VAS), which ranged from 0 to 200 millimeters. Responses on this task show large inter-individual differences but high across-days reliability within individuals.[35] Secondly, they rated their expectancy of experiencing an unconditioned stimulus (US) after observing the presented circle using an expectancy VAS, with ratings ranging from 'no shock' (1) to 'definitely a shock' (10).

In Experiment 1, the acquisition phase consisted of 14 CS+ trials, where the CS+ was paired with the unconditioned stimulus (US) in 7 trials, resulting in a reinforcement rate of 50%. In Experiment 2, the acquisition phase included 12 CS+ trials and 12 CS- trials. Notably, 83% of the CS+ trials were paired with the US, while the CS- trials were never paired with the US. The generalization phase in Experiment 1 encompassed four blocks, whereas Experiment 2 comprised three blocks, with each block being separated by a 3-minute break. Importantly, the US was never paired with the CS- or the TS trials; it was exclusively associated with the CS+ trials in both experiments. Each block in Experiment 1 consisted of 22 CS+ trials and 24 TS trials, and to prevent the extinction of the conditioned response, each block was always initiated with ten consecutive CS+ trials, known as re-acquisition. On the other hand, Experiment 2 incorporated 14 CS+ trials, 8 CS- trials, and 32 TS trials in each block, with each block commencing with six consecutive CS+ trials. Past research successfully used CS+ re-acquisition and large number of stimulus repetitions within a context of fear generalization.[35,55]

Computational model of generalization

Our computational framework advances the quantification of perceptual influences on fear generalization in two key ways (see Figure S5 for the Directed Acyclic Graph). First, we modeled how individuals perceive stimuli using a state-space model that captures both the probabilistic and dynamic nature of perception. Second, we introduced a new method for calculating stimulus distances based on the overlap between perceptual distributions, replacing our previous approach that relied on direct differences between mental coordinates (based on perceptual ratings).[3] Throughout our modeling, we employed weakly to non-informative priors given the current uncertainty about underlying processes (Method S4: Prior Specifications for Computational Models and Tables S1 and S2). A prior sensitivity analysis conducted for both the perceptual and generalization models demonstrates that our results are robust to a range of plausible prior choices (Method S5: Sensitivity Analysis of Model Priors and Figures S6 and S7). The following sections detail first the generalization model's core elements, then our new approach to computing perceptual distances.

Model architecture

The computational model incorporates a mixture structure, enabling the generation of four distinct theoretically or clinically relevant behavior paths. The first group, Non-Learners, is characterized by a learning rate of 0, indicating that the error-driven learning process has not occurred, and consequently, there is no expectation to be generalized. The second group, Overgeneralizers, exhibits a very high generalization tendency (indicated by a low value for the generalization rate). This propensity ensures that the similarity remains consistently higher than 70%, even when encountering the most distant stimuli in the experiment, thereby maintaining the US expectation above 70% of the expectation to the learned stimulus regardless of the encountered stimulus.

The remaining two groups, Physical Generalizers and Perceptual Generalizers, presume that individuals have learned the associations between the CS and US to some extent and subsequently generalize these associations to other stimuli (TS) based on similarity. The distinction between these groups lies in the dimension of generalization - physical or perceptual. Physical Generalizers assume no perceptual variability or that perceptual variability does not influence generalization behavior, while Perceptual Generalizers posit a link between perceptual variability and generalization behavior.

The fundamental assumption of the generalization model posits that at each time point, every participant (indexed by $i = 1,\ldots,n$) holds certain expectations regarding US onset denoted as v_{ij} for the conditioned stimulus on each trial ($j = 1,\ldots,k$) in the learning process. In each conditioned stimulus (CS) trial, the associative strengths of the CS(s) and the unconditioned stimulus (US) for individual i and trial j are updated based on the prediction error, which represents the difference between the current outcome and expectancy, and the individual learning rate (α_i).[39] Concurrently, the generalization process formulates the extend of US expectancy transfer (λ_i) to another stimulus s_{ij} by considering mental stimulus distances (d_{ij})[2] with stimulus coordinates corresponding to either physical stimulus features or perceived stimulus features. Mathematically, this relationship can be expressed in Equations 1 and 2:

$$v_{ij+1} = v_{ij} + \alpha_i(r_{ij} - v_{ij})k_{ij}, \quad \text{(Equation 1)}$$

and

$$g_{ij} = v_{ij}e^{-\lambda_i d_{ij}}. \quad \text{(Equation 2)}$$

where v_{ij} represents the associative strength of the CS(s) in trial j for participant i, reflecting the time-dependent learned expectation of the CS(s) - US association. In the differential learning experiment, we will distinguish further between v_{ij}^{+} and v_{ij}^{-}. Specifically, $v_{ij}^{+} \in [0, 1]$ denotes the associative strength for the CS+ (the excitatory reinforced CS) in the context of the differential learning experiment, while $v_{ij}^{-} \in [-1, 0]$ pertains to the CS- (the inhibitory reinforced CS). The dummy variable k_{ij} is used to control the occurrence of updating, where $k_{ij} \in \{0, 1\}$ with 1 indicating updating and 0 otherwise. Its role is to ensure that learning only happens during CS trials. The variable r_{ij} corresponds to the trial outcomes ($r_{ij} \in \{0, 1\}$ for the CS+ and $r_{ij} \in \{-1, 0\}$ for the CS-). The parameter α_i represents the learning rate, regulating the amount of value learning adaptation for individual i ($\alpha_i \in [0, 1]$), with higher values indicating more learning from the prediction error. The parameter λ_i is the generalization rate, signifying the rate of decay that ensues at a given fixed value of stimulus distance d_{ij}. It concurrently functions as a discriminative factor in identifying Overgeneralizers. An individual is allocated within the Overgeneralizers group if their acquired expectations v_{ij} demonstrate a decay exceeding 70%, even in the presence of the largest stimulus distance.

The mental stimulus distance between the currently encountered stimulus (TS) and the CS(s) is determined by stimulus coordinates, which may correspond to either the physical stimulus features (Physical Generalizers) or the perceived stimulus features (Perceptual Generalizers). Defined as follows:

$$d_{ij} = \left| x^{CS} - x_j^{TS} \right|, \quad \text{(Equation 3)}$$

where x^{CS} represents the coordinate of the CS, and x_j^{TS} signifies the coordinate of the TS on trial j, both situated within the physical dimension. In contrast, for Perceptual Generalizers:

$$d_{ij} = \left| \tilde{x}_{i,1,\ldots,j}^{CS} - \tilde{x}_{ij}^{TS} \right|. \quad \text{(Equation 4)}$$

with $\tilde{x}^{CS} i, 1, \ldots, j$ refers to the cumulative mean derived from the repeated presentations of the CS, capturing the perceived features of the CS up to trial j, while $\tilde{x}_{ij}^{TS}$ specifically denotes the perceived features of the TS at trial j.

The integration of learning and generalization processes yields an generalized associative strength, denoted as g_{ij}. In the context of differential conditioning, g_{ij} is constrained within the range of [-1, 1], $g_{ij} = v_{ij}^{+} e^{-\lambda_i d_{ij}^{+}} + g_{ij} v_{ij}^{-} e^{-\lambda_i d_{ij}^{-}}$,[64] while for simple conditioning, it is limited to [0, 1], $g_{ij} = v_{ij} e^{-\lambda_i d_{ij}}$. Notably, the magnitude of g_{ij} reflects the strength of generalized responses - smaller values result in lower generalized responses, while larger values lead to more potent responses. However, it is important to acknowledge that the scale of g_{ij} does not directly align with the scale of observed behavior, which operates on a 1-10 range. This discrepancy necessitates a scale transformation to establish a meaningful relationship between the two. To address this, we adopted a non-linear sigmoid function, which takes into account both the base rate response and scaling parameters. This sigmoid function effectively bridges the gap, mapping the latent generalized associative strength to the observed response in a manner that aligns with the observed behavioral scale. This can be seen in Equation 5:

$$\theta_{ij} = A + \frac{K - A}{1 + e^{-\left(w_{0_i} + w_{1_i} g_{ij}\right)}}, \quad \text{(Equation 5)}$$

The sigmoid function parameters, A and K, define the lower and upper limits, ensuring that θ_{ij} aligns with the measurement scale $\theta_{ij} \in [1, 10]$ $y_{ij} \in [1, 10]$ used in this study. Hence, we set $A = 1$ and $K = 10$. w_{0i} represents the baseline response parameter, dictating the response in the absence of CS associative strengths. On the other hand, w_{1i} serves as the scaling parameter, determining the relationship between latent and observed responses.

The final generalization response is assumed to follow a normal distribution with θ_{ij} being the mean of the distribution:

$$y_{ij} \sim \mathcal{N}\left(\theta_{ij}, \sigma^2\right). \quad \text{(Equation 6)}$$

The parameter σ plays a crucial role in regulating the level of response noise, which, in turn, is dependent on the specific group. A defining characteristic of the Non-Learners group is that their final responses are entirely random and unrelated to any learning or generalization processes. To accommodate this behavior, a distinct prior parameterized by σ has been assigned specifically for Non-Learners ($m_i = 1$), setting it apart from the prior applied to the other three latent groups ($m_i = 2, 3, 4$).

In the comparison between the generalization model incorporating the point-based perceptual assumption[3] and the current work's model-based assumption, it is noteworthy that no changes were anticipated within the Non-Learners and Overgeneralizers groups. These groups inherently exhibit generalization patterns independent of stimulus distance. However, discernable variations were expected within the Physical Generalizers and Perceptual Generalizers groups. The estimates within these groups assess the extent to which generalized responses align more with perceptual stimulus distance than with physical stimulus distance. Should the new perceptual distance variable result in a closer alignment of generalization responses with perceptual distance compared to the perceptual distance in the previous work,[3] an increase in the number of Perceptual Generalizers and a corresponding decrease in the number of Physical Generalizers are expected. Conversely, if the alignment with perceptual distance is less pronounced, an increase in Physical Generalizers and a decrease in Perceptual Generalizers would be observed.

New perceptual stimulus distance

When computing perceptual distance in the generalization process, our previous study[3] calculated mental inter-stimulus distance d_{ij} as the absolute difference between two point values: the current perceptual response and a memory representation of the CS (computed as the running average of past CS perceptual responses). While computationally straightforward, this approach treated perceptual responses as precise, fixed points in mental space, overlooking two fundamental aspects of human perception: (1) the uncertainty inherent in transforming physical stimuli into mental representations, and (2) the dynamic nature of how current sensory input integrates with prior experiences.

Here, we introduced a theoretically-grounded approach that explicitly incorporates these perceptual processes into distance calculations. With the new operationalization in this current work, distance equates to the amount of distribution overlap (rather than an absolute difference), with CS memory now also conceptualized as a probability distribution. The overlap coefficient provides a method to quantify probabilistic distributions into a deterministic metric, enabling direct comparison with other non-probabilistic assumptions about perceptual distance. This approach not only acknowledges the inherent uncertainty in perception but also allows for a fair and consistent comparison across competing theoretical frameworks. Because the current experimental design does not have information on the memory representation, we postulate that the memory distribution is faithfully encoded during the preceding CS trial and accurately decoded during the subsequent trial. The mean and the standard deviation of perception $\mu_{\psi,ij}$ and $\sigma_{\psi,ij}$ and perceptual memory $\mu^{CS}_{\psi,ij-1}$ and $\sigma^{CS}_{\psi,ij-1}$ are derived from a perceptual model with the perceptual data.

Crucially, while the concept of response distributions arising from feature overlap has been explored in prior generalization theories,[22,23,64] these frameworks were unable to distinguish whether these distributions stemmed from perceptual uncertainty, memory imprecision, or the generalization process itself. Our current framework addresses this limitation by specifically isolating perceptual processes: we compute distances between stimuli by analyzing the overlap between two distinct distributions - the perceptual distribution of the currently encountered stimulus and the memory distribution of the conditioned stimulus (CS).

The perceptual distance, d_{ij}, is calculated using the overlap between two probability densities: $f_1(x)$, representing how the current stimulus is being perceived, and $f_2(x)$, representing how the CS is remembered:

$$d_{ij} = 1 - \int \min(f_1(x), f_2(x))\, dx, \quad \text{(Equation 7)}$$

where

$$f_1(x) = \frac{1}{\sigma_{\psi,ij}\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_{\psi,ij})^2}{2\sigma^2_{\psi,ij}}\right), \quad \text{(Equation 8)}$$

and

$$f_2(x) = \frac{1}{\sigma^{CS}_{\psi,ij-1}\sqrt{2\pi}} \exp\left(-\frac{\left(x-\mu^{CS}_{\psi,ij-1}\right)^2}{2\left(\sigma^{CS}_{\psi,ij-1}\right)^2}\right). \quad \text{(Equation 9)}$$

This overlap-based approach offers several advantages. First, it accounts for the probabilistic nature of human perception, acknowledging that perceptual responses are not fixed but arise from distributions with inherent variability. Second, it provides a more comprehensive measure of similarity by incorporating both the central tendencies and uncertainties of the perceptual and memory distributions. When the two distributions perfectly align (i.e., maximal overlap), the perceptual distance is $d_{ij} = 0$, indicating high similarity. Conversely, minimal overlap results in d_{ij} approaching 1, signifying low similarity. By adopting this method, we ensure that perceptual uncertainty is explicitly considered, enabling a more nuanced understanding of generalization processes. Moreover, this metric bridges probabilistic and deterministic frameworks, facilitating direct comparison with traditional, non-probabilistic assumptions.

The determination of overlapping regions between the perceptual ($\mathcal{N}(\mu_{\psi,ij}, \sigma^2_{\psi,ij})$) and memory ($\mathcal{N}(\mu^{CS}_{\psi,ij-1}, \sigma^{CS^2}_{\psi,ij-1})$) distributions involves calculating the integral that captures the minimum of their respective probability density functions (PDFs) across the entire range, as shown in Equations 7, 8, and 9. To numerically approximate this integral, we employed the Monte Carlo method. By generating a set of random points within the defined range, we estimated the proportion of points that fell within the overlapping region, thereby approximating the corresponding area.

The Monte Carlo implementation involved sampling points uniformly within the range of interest. Specifically, for each randomly generated point, its y-coordinate was drawn from a uniform distribution between zero and the maximum value of the corresponding PDF. By comparing the y-coordinate of each point with the PDF values, we identified the points residing within the overlapping region and estimated the overlap area.

To avoid increasing the complexity of the generalization model, we adopted a two-step approach. First, we used the perceptual model to fit the size estimation data, yielding perception estimates grounded in the model. Next, we calculated the trial-by-trial perceptual distances using these estimates, resulting in a distance matrix $\widehat{d}_{ij}$. These values were then integrated into the generalization model. This two-step methodology provides several advantages. It enables a direct comparison of generalization models with

different perceptual assumptions while maintaining a consistent model structure. Additionally, it minimizes the risk of one process (i.e., perception or generalization) being mis-specified and subsequently influencing the inference of the other process, ensuring a more robust and interpretable modeling framework.

Perceptual model

Our perceptual model captures how humans transform physical stimuli into mental representations through a dynamic process that combines current sensory input with past experiences. We implemented this using a one-dimensional state-space model based on Bayesian principles, where perception emerges from the continuous interaction between incoming sensory evidence and prior expectations. The model's updating mechanism follows Kalman filter dynamics,[65] which provides a mathematically principled way to integrate new information with existing beliefs while accounting for their respective uncertainties. This approach has been implemented before to investigate the dynamic patterns of human perception.[20] The model architecture comprises three fundamental components:

(1) Sensory input processing:

First, physical stimuli are transformed into initial perceptual estimates through a flexible sigmoid mapping function:

$$\mu_{S_{ij}} = \frac{\Omega}{1+e^{-\left(\beta_{0,i}+\beta_{1,i}x_{s[ij]}\right)}}, \quad \text{(Equation 10)}$$

$$\psi_{S,ij} \sim \mathcal{N}\left(\mu_{S,ij}, \sigma^2_{S,i}\right). \quad \text{(Equation 11)}$$

where $\Omega = 200$ represents our rating scale's upper limit, and $\beta_{0,i}$ and $\beta_{1,i}$ are individual-specific parameters capturing personal differences in the perceptual response. Crucially, this initial mapping produces not a single value but a probability distribution, reflecting inherent sensory uncertainty. The sigmoid function offers distinct advantages over traditional psychophysical functions like Weber's law or Stevens' power law for our modeling purposes. While traditional psychophysical functions are derived from averaged data and presuppose uniform perception across individuals, the sigmoid function's mathematical properties afford greater flexibility in capturing individual differences. The dual scaling parameters enable the accommodation of diverse perceptual mapping patterns (Figure 3A), making it particularly suitable for modeling inter-individual variability in stimulus perception.

(2) Dynamic integration:

The heart of our model is a Kalman filter mechanism that governs how current sensory input is integrated with prior perceptual expectations. This integration process unfolds continuously over time, with each moment requiring the brain to combine new sensory evidence with accumulated prior expectations. At any given time point ($j + 1$), the brain receives fresh sensory evidence ($\mu_{S,ij+1}$) about the current stimulus while maintaining its prior expectation ($\mu_{\psi,ij}$) from previous experiences. The Kalman filter determines the optimal way to combine these information sources based on their relative uncertainties. The updated perception is computed through a weighted averaging process:

$$\mu_{\psi,ij+1} = \mu_{\psi,ij} + \alpha_{\psi,ij+1}\left(\mu_{S,ij+1} - \mu_{\psi,ij}\right), \quad \text{(Equation 12)}$$

This equation illustrates how perception is updated in response to new perceptual experiences. Specifically, it reflects the difference between the expected perception ($\mu_{\psi,ij}$) and the actual sensory input ($\mu_{S,ij+1}$). The extent of this adjustment is governed by the Kalman gain ($\alpha_{\psi,ij+1}$), which is defined as:

$$\alpha_{\psi,ij+1} = \frac{\sigma_{\psi,ij}}{\sigma_{\psi,ij}+\sigma_{S,i}}, \quad \text{(Equation 13)}$$

The Kalman gain serves as a dynamic arbiter between new sensory information and prior expectations. When uncertainty in prior expectations ($\sigma_{\psi,ij}$) is high relative to sensory uncertainty ($\sigma_{S,i}$), the gain approaches 1, causing perception to rely more heavily on new sensory input. Conversely, when prior uncertainty is low relative to sensory uncertainty, the gain approaches 0, leading to greater reliance on prior expectations. This adaptive weighting ensures optimal integration based on the reliability of each information source. Consider how this plays out in practice: When someone has developed very stable prior expectations through repeated exposure to similar stimuli, their prior uncertainty ($\sigma_{\psi,ij}$) becomes relatively low. If they then encounter these stimuli in a noisy environment (high $\sigma_{S,i}$), the Kalman gain will be small, causing them to rely more on their well-established prior expectations than on the uncertain sensory input. However, if they encounter something unexpected that challenges their prior expectations, or if the sensory evidence is particularly clear and reliable, the weighting will shift to favor the new information.

(3) Uncertainty evolution:

The evolution of perceptual uncertainty represents another crucial dynamic aspect of our model. As individuals accumulate perceptual experiences, the precision of their perceptual estimates changes systematically. This evolution follows a principled updating rule:

$$\sigma_{\psi,ij+1} = (1 - \alpha_{\psi,ij+1})\sigma_{\psi,ij} + \eta e^{-\omega_i \cdot j}. \quad \text{(Equation 14)}$$

The first term, $(1 - \alpha_{\psi,ij+1})\sigma_{\psi,ij}$, reflects how uncertainty naturally decreases as we gain more experience with a stimulus. However, perception typically maintains some degree of flexibility rather than becoming entirely rigid or deterministic. The second term, $\eta e^{-\omega_i \cdot j}$, introduces process noise that decays over time at an individual-specific rate ω_i. This decay term allows for individual differences in how people balance stability with flexibility in their perceptual systems. The parameter η sets the initial magnitude of this uncertainty, while ω_i determines how quickly an individual's perceptual system stabilizes (Figure 3B).

In summary, our perceptual model captures how sensory information is processed and integrated over time. At each temporal interval, the sensory system transforms physical stimuli into an initial distribution $\mathcal{N}(\mu_{S,ij+1}, \sigma^2_{S,i})$. This sensory information is then integrated with prior expectations based on previous experiences, represented by $\mathcal{N}(\mu_{\psi,ij}, \sigma^2_{\psi,ij})$, to generate the current perceptual distribution $\mathcal{N}(\mu_{\psi,ij+1}, \sigma^2_{\psi,ij+1})$. The integration process reflects both immediate sensory transformations and accumulated perceptual history, with the relative influence of each determined by the Kalman gain $\alpha_{\psi,ij}$.

This formulation naturally accommodates individual differences in perceptual processing. For some individuals, the weight given to prior perceptual experiences strengthens over time, manifested through a decreasing trajectory of $\alpha_{\psi,ij}$ as experiences accumulate. In contrast, others maintain a more flexible perceptual system where current sensory information continues to play a prominent role, evidenced by sustained high levels of $\alpha_{\psi,ij}$ across time. These individual-specific dynamics emerge from the interaction between the Kalman gain mechanism and person-specific parameters governing uncertainty evolution.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical inference was carried out using Markov Chain Monte Carlo (MCMC) with the Gibbs sampling method implemented through JAGS.[66] The analysis was conducted in the statistical computing language R,[67] utilizing the R package jagsUI.[68] To ensure robust results, four MCMC chains were executed for the two models - the perceptual and the generalization model, each comprising 100,000 iterations. A burn-in period of 75,000 iterations was implemented to discard initial samples, and a thinning factor of 10 was applied, resulting in a total of 10,000 ($\frac{(100{,}000 - 75{,}000)}{10} \times 4$) retained samples per parameter. Convergence of the MCMC chains was assessed using the $\widehat{R}$ statistic based on Gelman and Rubin diagnostics.[69,70] The chains were considered to have reached a stabilized state and attained the target distribution when the $\widehat{R}$ value value approached, or was close to, 1.

For the comparison of two models with different assumptions about perceptual distances, we further ran the super model which encompasses both model assumptions and estimated the model selection parameter β_M with the following MCMC rules: Four MCMC chains with 250,000 iterations each. A burn-in period of 100,000 iterations and a thinning factor of 5, resulting in a total of 120,000 ($\frac{(250{,}000 - 100{,}000)}{10} \times 4$) retained samples for the model selection parameter β_M. It is crucial to emphasize that the two models being compared share identical model structures, encompassing the same parameters, variables, and likelihood functions. The sole distinction lies in the application of different inter-stimulus distance metrics—point differences versus overlap in distributions.

The construction of a super model and the estimation of the model selection parameter β_M aim to evaluate whether the model's performance is enhanced when employing a probabilistic perceptual model compared to the previous descriptive approach. To investigate this, we created a nested structure, wherein two models with different perceptual distance assumptions were incorporated into a single model (i.e., a super model), and subsequently, we combined the likelihoods of these two model assumptions in a linear manner.[71] Specifically, considering the parameter vectors θ_1 and θ_2 corresponding to the model utilizing the model-based perceptual distance assumption, denoted as M_1, and the model employing the descriptive perceptual distance assumption, denoted as M_2, we thoroughly examined the likelihood function of the observed data D:

$$P(D|\beta_M, M_1, M_2) = \beta_M P(D|\theta_1, M_1) + (1 - \beta_M)P(D|\theta_2, M_2). \quad \text{(Equation 15)}$$

where $\beta_M \sim Beta(1,1)$, and $\beta_M > 0.5$ infers that M_1 outperforms M_2 and vice versa.

With the properly nested model, we can then compute the Bayes factor[72,73] with the Savage-Dickey method[74,75] based on the parameter β_M to determine to what extent the performance of the two models are different. For this, we computed Bayes factor given H_0 having β_M fixed to 0.5 and H_1 having β_M deviated from 0.5:

$$\frac{P(D|H_0)}{P(D|H_1)} = \frac{P(\beta_M = 0.5|D, H_1)}{P(\beta_M = 0.5|H_1)}. \quad \text{(Equation 16)}$$

This implies that to compute the Bayes factor, we only need to consider the ratio between the posterior of the parameter β_M under the more intricate assumption H_1 , given the observed data D, and the prior of β_M under H_1 (see Method S6: Savage-Dickey Density Ratio for the proof). When the Bayes factor is greater than 1, it signifies that there is stronger evidence suggesting that β_M deviates

from 0.5, indicating that models M_1 and M_2 exhibit distinct performance. Conversely, if the Bayes factor is less than 1, it indicates that the evidence favours β_M being close to 0.5, suggesting that models M_1 and M_2 have similar performance.

Once a determination has been made regarding the presence of enough evidence indicating the superior performance of one model over another, our attention turns to examining the distribution of posterior samples of the model selection parameter β_M to determine whether the data favors model M_1 or M_2.