

**Representational shifts as a distinct mechanism in associative learning and  
generalization**

Kenny Yu<sup>1</sup>, Steven Verheyen<sup>2</sup>, Tom Beckers<sup>3</sup>, Wolf Vanpaemel<sup>1</sup>, Francis Tuerlinckx<sup>1</sup>, and  
Jonas Zaman<sup>3,4,5</sup>

<sup>1</sup>Quantitative Psychology and Individual Differences, KU Leuven


<sup>2</sup>Department of Psychology, Education and Child Studies, Erasmus University Rotterdam


<sup>3</sup>Centre for Learning and Experimental Psychopathology, KU Leuven


<sup>4</sup>REVAL Rehabilitation Research, Faculty of Rehabilitation Sciences, UHasselt


<sup>5</sup>Center for Translational Neuro- and Behavioral Sciences, University of Duisburg-Essen


**Author Note**


Kenny Yu  <https://orcid.org/0000-0002-0665-9354>

Steven Verheyen  <https://orcid.org/0000-0002-6778-6744>

Tom Beckers  <https://orcid.org/0000-0002-9581-1505>

Wolf Vanpaemel  <https://orcid.org/0000-0002-5855-3885>

Francis Tuerlinckx  <https://orcid.org/0000-0002-1775-7654>

Jonas Zaman  <https://orcid.org/0000-0002-2218-3018>

Correspondence concerning this article should be addressed to Kenny Yu, Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, box 3726, 3000 Leuven, Belgium. E-mail: [kenny.yu@kuleuven.be](mailto:kenny.yu@kuleuven.be). The raw data for the experiment, codes for the computational model and analysis, and Supplementary Materials which contains more modelling details, are available at the OSF repository: <https://osf.io/a8c4g/>.

*Word count: 7613*

### Abstract

The peak shift effect in generalization, a robust phenomenon where the maximal response occurs to stimuli shifted away from the conditioned stimulus in the direction opposite the inhibitory stimulus, has been predominantly attributed to associative learning. However, recent findings point to biases in stimulus representation as an alternative mechanism. Our research, comprising two experiments (Experiment 1:  $N = 76$ ; Experiment 2:  $N = 234$ ), investigated whether the peak shift effect could emerge from a stimulus contrast effect, independent of associative learning. By comparing stimulus identification patterns between simple and differential conditioning groups both after stimulus exposure and following associative learning, we found that shifts in the identification of conditioned stimuli were primarily driven by a stimulus contrast effect, rather than associative learning per se. This representational shift was closely associated with the peak shift phenomenon observed in post-learning generalization patterns, with identification errors strongly impacting generalized responding. These results underscore the importance of considering stimulus representations as an independent process in understanding learning-based behavior, offering new insights into the mechanisms underlying generalization phenomena.

*Keywords:* Associative Learning; Generalization; Identification; Contrast effect

## **Representational shifts as a distinct mechanism in associative learning and generalization**

### **Introduction**

To function adaptively in different environments, humans rely not only on the cognitive transfer of prior learning to new situations but also on imperfect perceptual or memory representations of physical reality. These representations can facilitate the detection of threatening contexts through altered processing of physical inputs. For instance, recent research has shown that aversive odors or auditory stimuli can alter discrimination thresholds, serving as a distinct mechanism that allows individuals to respond more effectively to newly encountered potential threats (Resnik & Paz, 2015; Resnik et al., 2011).

The transfer of past learning, known as generalization, is a cornerstone in psychology and ethology research (Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981; Mednick & Freedman, 1960). This process allows humans and other animals to apply knowledge about previous experiences to new, similar situations, thereby reducing the need for re-learning and conserving cognitive resources. Recent research has increasingly focused on exploring the mechanisms behind generalization, ranging from sensory or perceptual errors (Zaman, Struyf, Ceulemans, et al., 2019; Zaman, Ceulemans, et al., 2019; Zaman, Yu, & Verheyen, 2023; Zaman et al., 2021, 2022; Zenses et al., 2021) to complex higher-order cognitive processes (Boddez et al., 2017; Dunsmoor & Murphy, 2015; Lee & Livesey, 2018). Understanding these mechanisms is essential for gaining deeper insights into how individuals utilize prior knowledge to effectively navigate and respond to diverse environments.

In the study of generalization, one intriguing phenomenon is the peak shift effect (Hanson, 1959; Purtle, 1973). This effect typically emerges after differential conditioning, where one neutral stimulus (CS+, such as a red circle) is paired with an outcome (US, like an electric shock), while another neutral stimulus (CS-, such as a blue circle) is associated

with the absence of that outcome. The effect is most commonly observed with stimuli that share a physical dimension (Purtle, 1973). In subsequent generalization tests using stimuli that vary along a continuum including the conditioned stimuli (e.g., circles in various shades between red and blue), the peak response is often displaced from the CS+ in the direction opposite the CS-. This peak shift effect is consistently found in differential conditioning paradigms. Furthermore, the averaged responses to stimuli closer to the CS+ are often observed to be stronger in the direction opposite to the CS- compared to those closer to the CS- (the area shift), resulting in generalization gradient asymmetry. Conversely, in simple conditioning paradigms, where only the CS+ is associated with the US, generalization gradients are typically symmetrical around the CS+, without the asymmetry and peak displacement observed in differential conditioning.

Traditionally, the peak shift effect has been explained as resulting from an interaction between excitatory and inhibitory associations that are formed during differential learning (Blough, 1975; Spence, 1937; Thomas & Thomas, 1974). Excitatory strength, developed through CS-US pairings, supports the production of a conditioned response, while inhibitory strength, established through CS-noUS pairings, suppresses it. According to this theory, stimuli adjacent to the CS+ on the side opposite from the CS- acquire similar excitatory strength as the CS+ but, being further away from the CS-, accumulate less inhibitory strength during associative learning. This imbalance results in a stronger response to these specific adjacent stimuli compared to the CS+ itself. Elemental theories (Ghirlanda & Enquist, 1999; McLaren & Mackintosh, 2002) have expanded on this framework, proposing that stimuli are represented by multiple elements along a continuum. When a stimulus is presented, it activates these elements to varying extents. Notably, elements strongly activated by the CS+ do not acquire the most associative strength because they are also activated by the CS-, leading to a combination of excitatory and inhibitory strengths and a subsequently lower net response. Instead, the stimulus that elicits the greatest difference between excitatory and inhibitory response strengths

generates the strongest response during generalization.

An alternative explanation for the peak shift phenomenon is provided by adaptation theory (Thomas & Switalski, 1966; Thomas & Thomas, 1974). This theory suggests that when an individual is repeatedly exposed to two stimuli along the same physical dimension, the brain forms an average or adapted representation that lies between these stimuli. The response to a new stimulus is then influenced by the relative difference between this stimulus and the adapted average. As new stimuli are presented, this average shifts incrementally. According to this theory, when an individual is repeatedly exposed to two stimuli (CS+ and CS-) along the same physical dimension, the brain forms an adapted average representation between these stimuli, rather than simply learning to distinguish between them. The response to a new stimulus is then influenced by its relative difference from this adapted average, not just its absolute properties. The subsequent presentation of other test stimuli influences the average representation leading to a shift in peak responding since responding is relative to this average. Both the associative learning explanation and adaptation theory for the peak shift effect have limitations. The associative learning approach fails to account for the malleability of stimulus representations. While adaptation theory does consider stimulus representations, it still falls short in fully capturing the complexity of human perception. Crucially, both explanations neglect the critical role of contextual factors in shaping perceptual experiences and behavioral responses. This oversight is particularly problematic given the substantial body of evidence demonstrating that human perception and memory frequently deviate from physical reality and are profoundly influenced by various perceptual contexts (Bays et al., 2024; Fougner et al., 2012; Guo et al., 2004; Petzschner et al., 2015; Purves et al., 2014). Increasing empirical evidence demonstrates that the perception of encountered physical objects is significantly shaped by prior perceptual experiences (Adams et al., 2004; Aru et al., 2016; Chopin & Mamassian, 2012; Samaha et al., 2018; Snyder et al., 2015; Yon & Frith, 2021). Additionally, perception is profoundly influenced by contextual factors (Albright & Stoner,

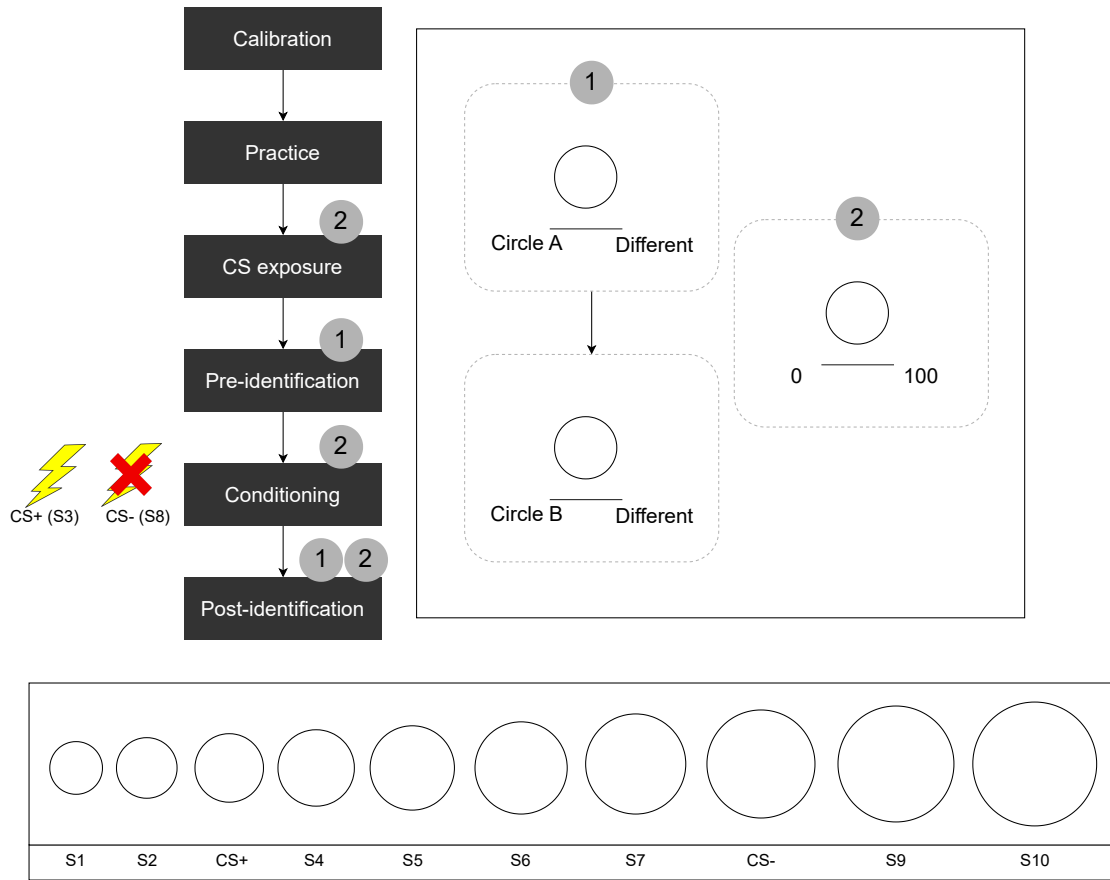
2002). A well-known example is the stimulus contrast effect, where individuals perceive physical reality differently depending on the contrasted baseline that is established by contextual factors or prior experiences (Helson, 1964). For instance, individuals tend to perceive circles as smaller when larger circles are presented simultaneously (McCarthy et al., 2013; Mruczek et al., 2017; Urale & Schwarzkopf, 2023; Weintraub & Schneck, 1986).

Moreover, recent research in perceptual working memory emphasizes that perceptual inputs stored in working memory do not exist in isolation during retrieval; rather, they interact, especially when they share physical characteristics (Bae & Luck, 2017; Yang et al., 2024). For instance, evidence suggests that perceptual working memory incorporates higher-level statistical structures instead of merely item-specific information (Brady & Alvarez, 2011, 2015). Furthermore, studies have shown that perceptual memory representations can be readily modulated by other simultaneously encoded information (Bae & Luck, 2017; Rademaker et al., 2015). This modulation illustrates the group effect in perceptual encoding, where the same perceptual item is not represented identically in memory when stored with different groups of items (Clevenger & Hummel, 2014; Xu, 2006; Xu & Chun, 2007). Additionally, variations in memory responses observed during recall can originate from perceptual processes, not solely from interactions during retrieval (Bae et al., 2015; Olkkonen et al., 2010, 2014). In a series of delay estimation experiments (Bae et al., 2015), participants were asked to recall colors that were initially presented near the boundary between two categories, such as a specific hue of blue or green. The results revealed a systematic bias in how participants recalled these colors, particularly for those close to category boundaries. This indicates that variations in recall were significantly shaped by the processes involved in perceptual encoding, rather than solely by memory retrieval.

Research into human perception and perceptual memory raises an intriguing explanation for the peak shift effect observed in conditioned responses: it may stem from shifts in the stimulus representation of the CS+ due to a stimulus contrast effect, rather

than from associative learning per se (Zaman et al., 2021). If confirmed, this would challenge traditional theories of peak shift, potentially rendering a central explanatory mechanism obsolete. Indirect evidence comes from a recent study where the peak shift in conditioned responses vanished after correcting for perceptual errors, which appeared to result from a stimulus contrast effect. Specifically, the stimulus most often identified as the CS+ was not the CS+ itself, but an adjacent stimulus (Zaman et al., 2021). However, previous studies exploring the relationship between perceptual errors and generalization (Zaman, Struyf, Ceulemans, et al., 2019; Zaman, Ceulemans, et al., 2019; Zaman, Yu, & Verheyen, 2023; Zaman et al., 2021, 2022; Zenses et al., 2021) only measured stimulus identifications after associative learning, making it difficult to disentangle their contribution from the associative learning process. It is possible that the area shift in generalization, along with the associated perceptual errors observed in these studies, were driven by the same learning process. This underscores the need for an investigation that tests identification prior and after associative learning combined with post-learning generalization responses to examine how they interact.



**Figure 1**

*The experimental paradigm and the visual stimuli.*

In this study, we aimed to investigate response patterns of stimulus identification after stimulus exposure and after conditioning in both simple and differential learning groups, and to explore how these patterns interact with post-learning generalization behavior. To this end, we conducted two experiments (Experiment 1 (laboratory):  $N = 76$ ; Experiment 2 (online):  $N = 234$ ). Participants were divided into two groups: in one group a single stimulus (CS+) was used during visual exposure and conditioning, while in the other group two stimuli (CS+ and CS-) were used. Stimulus identification of ten physically similar stimuli was measured twice, after stimulus exposure and conditioning by presenting 10 different sized circles repeatedly (Figure 1). Additionally, in the second experiment,

generalized responding (US expectancy) was measured concurrently with stimulus identification during the post-conditioning identification task. This experimental design allowed us to examine whether previously biased stimulus representations emerged due to a stimulus contrast effect or due to associative learning and their impact on shaping post-learning generalization behavior.

## Methods

The study was pre-registered on the Open Science Framework (Experiment 1: <https://osf.io/bqkgy>; Experiment 2: <https://osf.io/k36ug>). We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All relevant materials, including experiment scripts, data, and analysis scripts, are accessible at <https://osf.io/a8c4g/>. Ethical approval was obtained from KU Leuven’s Social and Societal Ethics Committee (G-2022-5873-R4).

Initially, we registered the analysis plan for Experiment 1 using a frequentist statistical approach, which informed our sample size determination. However, we ultimately adopted Bayesian statistics for parameter estimation in both experiments. Despite this methodological shift, we employed weakly informative priors in the inference process, yielding results anticipated to be comparable to those obtained through frequentist methods.

## Experiments

The current study comprised two experiments: one conducted in a laboratory setting ( $N = 76$ ) and the other conducted online ( $N = 234$ ).

### *Participants*

The sample size for Experiment 1 was determined based on data from two previous studies (Zaman, Ceulemans, et al., 2019; Zaman et al., 2021), which examined post-learning stimulus identification gradients following simple or differential fear conditioning. An effect size of  $\eta_p^2 = 0.05$  (95% CI = [0.03, 0.08]) was calculated for the interaction between experiment type and stimulus. To ensure a conservative estimate, we

used half of this effect size, resulting in the following G\*Power (3.1.9.7) settings:

$\eta_p^2 = 0.025$ ,  $f = 0.16$ ,  $\alpha = 0.05$ ,  $\beta = 0.95$ , 2 groups, 10 measures, correlation within measures = 0.3, non-sphericity correction  $\epsilon = 0.8$ . This yielded a target sample size of 78 participants. For Experiment 2, the sample size was tripled to 234 participants to account for potential additional noise from the online setting.

In Experiment 1, 92 participants were recruited (simple conditioning: 46; differential conditioning: 46). 76 participants (simple conditioning: 43; differential conditioning: 33) were left for the analysis after applying the exclusion criteria, which included reporting non-serious responses or having more than 20% missing data in any of the four phases. Participants were divided equally into two groups: one group underwent simple conditioning (one conditioned stimulus), and the other underwent differential conditioning (two conditioned stimuli). Recruitment was conducted through the participant pool of the KU Leuven Faculty of Psychology and Educational Sciences, and participants received research participation credit or €8 as compensation. The experiment lasted approximately one hour, and all instructions were provided in Dutch.

In Experiment 2, 275 participants were recruited through Prolific (simple conditioning: 143; differential conditioning: 132). After applying exclusion criteria, 234 participants (simple conditioning: 117; differential conditioning: 117) were included in the analysis (simple conditioning: 53% female; mean age = 28.51 years, SD = 7.30; differential conditioning: 51% female; mean age = 28.22 years, SD = 6.96). Participants received €8 for their participation, and the experiment took approximately one hour. All instructions were provided in English.

### *Stimuli*

**Visual stimuli.** For both experiments, the visual stimuli consisted of 10 circles with diameters ranging from 50.80 mm to 119.42 mm, each differing by 7.624 mm. The circles, labeled S1 to S10, were presented as white outlines against a black background. Circle size has been widely used in perceptual generalization studies with both healthy and

clinical populations (Lange et al., 2017; Lissek, 2012; Lissek et al., 2008, 2014; Yu et al., 2023; Zaman, Ceulemans, et al., 2019). For both the simple and differential conditioning groups, the third circle (S3) served as the conditioned stimulus (CS+). In the differential conditioning group, the eighth circle (S8) served as the second conditioned stimulus (CS-). The remaining circles were used as test stimuli. During the practice trials, squares of different side lengths (50, 80, 100 mm) were presented, with two trials for each size, to allow participants to practice the identification task.

**Unconditioned stimuli.** In Experiment 1 (laboratory), participants experienced a 2-ms aversive electrocutaneous stimulus as the unconditioned stimulus (US). This stimulus was administered using a Constant Current Stimulator (DS7) through a pair of Ag/AgCl electrodes (each 8 mm in diameter) placed on the non-dominant wrist and lubricated with K-Y gel. The US intensity was individually adjusted using the Ascending Method of Limits approach (Yarnitsky et al., 1995) to achieve a pain rating of 8 on a Visual Analog Scale (VAS) ranging from 0 (no pain) to 10 (worst imaginable pain). Starting at 2 mA, the intensity increased by 0.2 mA per step to ensure that pain sensations remained tolerable. On average, the selected intensity was 21.43 mA (SD = 18.65).

In Experiment 2 (online), a visual image of a yellow lightning bolt served as the US. Participants were instructed to imagine a hypothetical situation where they encountered a strange machine that occasionally delivers electrical shocks. This machine displays different symbols, which may or may not predict whether a shock will occur. Their task was to determine which symbols predicted the hypothetical electrical shock. A similar predictive learning paradigm has been used in recent studies (Lee et al., 2019; Ng et al., 2022).

### ***Procedure***

In Experiment 1 (laboratory), participants were required to maintain a fixed head position using a headrest throughout the experiment, except during breaks between trials, to ensure consistent viewing distance from the computer screen (53 cm). Informed consent was obtained after participants received all necessary information. The experiment

comprised six phases: a calibration phase, a practice phase, a CS exposure phase, a pre-conditioning identification phase, a conditioning phase, and a post-conditioning identification phase. The electrical stimulation (US) was administered only during the conditioning phase. On each trial, a fixation cross appeared, followed by the visual stimulus and a response option. At the beginning of each trial, the mouse position was reset to the bottom center of the screen to equalize the distance between response options.

After calibrating the intensity of the electrical stimulation, participants engaged in a practice block to familiarize themselves with the experimental identification task. The practice block required participants to categorize squares instead of circles, with one square used for simple conditioning and two squares for differential conditioning. In the simple conditioning task, participants determined whether the presented square matched or differed from a previously shown square (5 mm). In contrast, the differential conditioning task involved comparing the square to two previously shown squares (5 mm and 8 mm). The practice block consisted of six trials, with each of the three differently sized squares (5 mm, 8 mm, 10 mm) presented twice. No electrical stimulation (US) was administered during these practice trials.

During the CS-exposure phase, the CS+ stimulus (labeled as CIRCLE A) was presented eight times in both simple and differential conditioning. In the differential conditioning task, the CS- stimulus (labeled as CIRCLE B) was also presented eight times, with the CS+ and CS- presented in a random order. No electrical stimulation (US) was given during this phase. Each trial featured a simultaneous display of the CS and a visual analogue scale (VAS) for 5 seconds. Participants rated their expectation of receiving a shock on the US expectancy VAS, which ranged from 0 (no shock) to 100 (shock).

The pre-conditioning identification phase consisted of four blocks. Each of the ten stimuli was presented a total of 12 times (120 trials in total), evenly distributed across the four blocks (3 times each). All stimuli were displayed in a random order. Participants were required to identify each presented circle as being either identical to or different from the

CS+ (labeled as CIRCLE A) from the previous phase, relying on their memory as the target stimuli were not presented for direct comparison. In each trial, one of the 10 stimuli appeared alongside two response options at the bottom of the screen. For the simple conditioning task, the response options were “CIRCLE A” or “DIFFERENT.” The trial concluded when a response option was selected. For the differential conditioning task, the response options were also “CIRCLE A” or “DIFFERENT.” If participants identified the stimulus as CIRCLE A (S3), the trial ended immediately. If they chose the “DIFFERENT” option, they were then prompted to decide whether the stimulus was identical to CIRCLE B (S8) or a different circle altogether. The circle and response options remained visible for a maximum of 5 seconds.

The conditioning phase closely mirrored the CS-exposure phase, with one key difference: the unconditioned stimulus (US) was administered at the end of the CS+ presentation in six out of the eight CS+ trials (75% reinforcement rate). In the differential conditioning task, none of the CS- trials were reinforced with the US. After the conditioning phase, participants completed another identification phase with a structure identical to the pre-conditioning phase. After the post-conditioning identification phase, participants were debriefed.

Each block of trials was separated by a 20-second break. Throughout the experiment when participants failed to respond within the allotted time, their response was recorded as a missing value. Participants did not receive feedback on their answers. For all four phases, a fixation cross appeared during the 1.5-second intertrial interval (ITI) separating the trials.

Experiment 2 (online) was essentially identical to Experiment 1, with the following differences. First, at the beginning of the experiment, participants were instructed to place a physical credit card on the screen where a virtual credit card was displayed. They then adjusted the size of the virtual credit card to match the physical one. This calibration captured the screen dimensions of each participant, ensuring that participants with

different screen sizes viewed the circles at the correct sizes. Second, during the post-conditioning identification phase, participants provided their US expectancy ratings for each presented circle after giving their identification responses. As in Experiment 1, no US was administered during this phase. To prevent extinction, participants were informed that during this phase no feedback would be provided on the delivery of virtual electric shocks and were instructed to use their experiences from the conditioning phase to indicate their US expectancy.

### **Analysis**

All analyses were conducted using Bayesian mixed-effects models to evaluate how various experimental variables influenced the response variables across different phases and conditions of the experiments. Hypothesis testing was carried out by calculating Bayes factors (BF) using the Savage-Dickey ratio (Dickey, 1971; Wagenmakers et al., 2010), which provides a quantitative measure of evidence for or against the null hypothesis. Specifically, the Bayes factor was computed as the ratio of the probability of the data under the alternative hypothesis (that the effect in the mixed model differs from zero) to that under the null hypothesis (that the effect in the mixed model is zero). In this context, a Bayes factor greater than 1 indicates stronger evidence against the null hypothesis, suggesting that the effect differs from zero. Conversely, a Bayes factor less than 1 provides stronger evidence in favor of the null hypothesis, indicating that the effect does not differ from zero. According to the scale proposed by Kass and Raftery (1995), BF values between  $10^{-2}$  and  $10^{-1}$  indicate strong evidence, and those between  $10^{-1}$  and  $10^{-0.5}$  indicate substantial evidence, in favor of the null hypothesis. Conversely, BF values between  $10^{0.5}$  and  $10^1$  indicate substantial evidence, and those between  $10^1$  and  $10^2$  indicate strong evidence, in favor of the alternative hypothesis.

### ***Associative learning***

To investigate changes in US expectancy during the CS-exposure and conditioning phases, US expectancy was used as the response variable in the analysis. For the simple

conditioning group, the predictors included Repetition (the number of times a stimulus was presented) and Phase (CS-exposure and conditioning). In the differential conditioning group, Stimulus (CS+ and CS-) was also included as a predictor. Additionally, a separate model was constructed using the combined data from both groups to determine whether US expectancy patterns during CS+ trials differed between the simple and differential conditioning groups, with Group (simple vs. differential conditioning) as an additional predictor. Interaction effects between Repetition, Phase, Stimulus, and Group were incorporated to assess how these factors might jointly influence US expectancy.

To further investigate the learning process, we applied a multilevel error-driven learning model to determine whether participants adjusted their responses based on the presence or absence of the US (Rescorla & Wagner, 1972). For mathematical details see Supplementary Information. Participants were classified as Non-Learners if their learning rate parameters were estimated as not differing from zero. In Experiment 1, no participants were identified as Non-Learners. However, in Experiment 2, 12 participants in the simple conditioning group and 30 in the differential conditioning group were identified as Non-Learners. For all post-conditioning analyses, we ran models using both the full dataset and the data excluding Non-Learners. The results patterns were consistent across both analyses. We report the results from the full dataset.

### ***Identification***

To investigate identification patterns before and after conditioning, we analyzed the sum of CIRCLE A (CS+; S3) identifications per stimulus as the response variable for the simple conditioning group. For the differential conditioning group, we separately analyzed the sum of CIRCLE A (CS+; S3) identifications and the sum of CIRCLE B (CS-; S8) identifications per stimulus as response variables. The first model explored how identification patterns were influenced by the interaction between Stimulus (S1 to S10) and Phase (pre-conditioning vs. post-conditioning identification) within each group. In the second model, Group (simple vs. differential conditioning) was introduced as an additional



predictor to compare identification patterns across the two groups. For this combined analysis, only the sum of CIRCLE A identifications was used as the response variable for both groups. Interaction effects among Stimulus, Phase, and Group were included to assess how these factors jointly influenced identification patterns.

### ***Associative learning and identification***

In Experiment 2, where US expectancy was measured alongside identification, we explored the relationship between US expectancy and identification within each conditioning group. US expectancy was used as the response variable, with trial-by-trial identifications as predictors (simple conditioning: CS+ or others; differential conditioning: CS+, CS-, or others). Additionally, we ran models where Stimulus alone served as the predictor and compared this single-predictor model to a two-predictor model that included both Stimulus and identification. This comparison aimed to assess how much the inclusion of stimulus identification improved the predictive accuracy of US expectancy.

Model comparison was conducted using Leave-One-Out Cross-Validation (LOO), a statistical technique that estimates the predictive accuracy of models by systematically leaving out one observation at a time from the dataset. The model is then trained on the remaining data and tested on the omitted observation. This process is repeated for each observation in the dataset. The average performance across all iterations provides an estimate of how well each model generalizes to new data. This approach was used to determine the added value of including stimulus identification as a predictor for US expectancy.

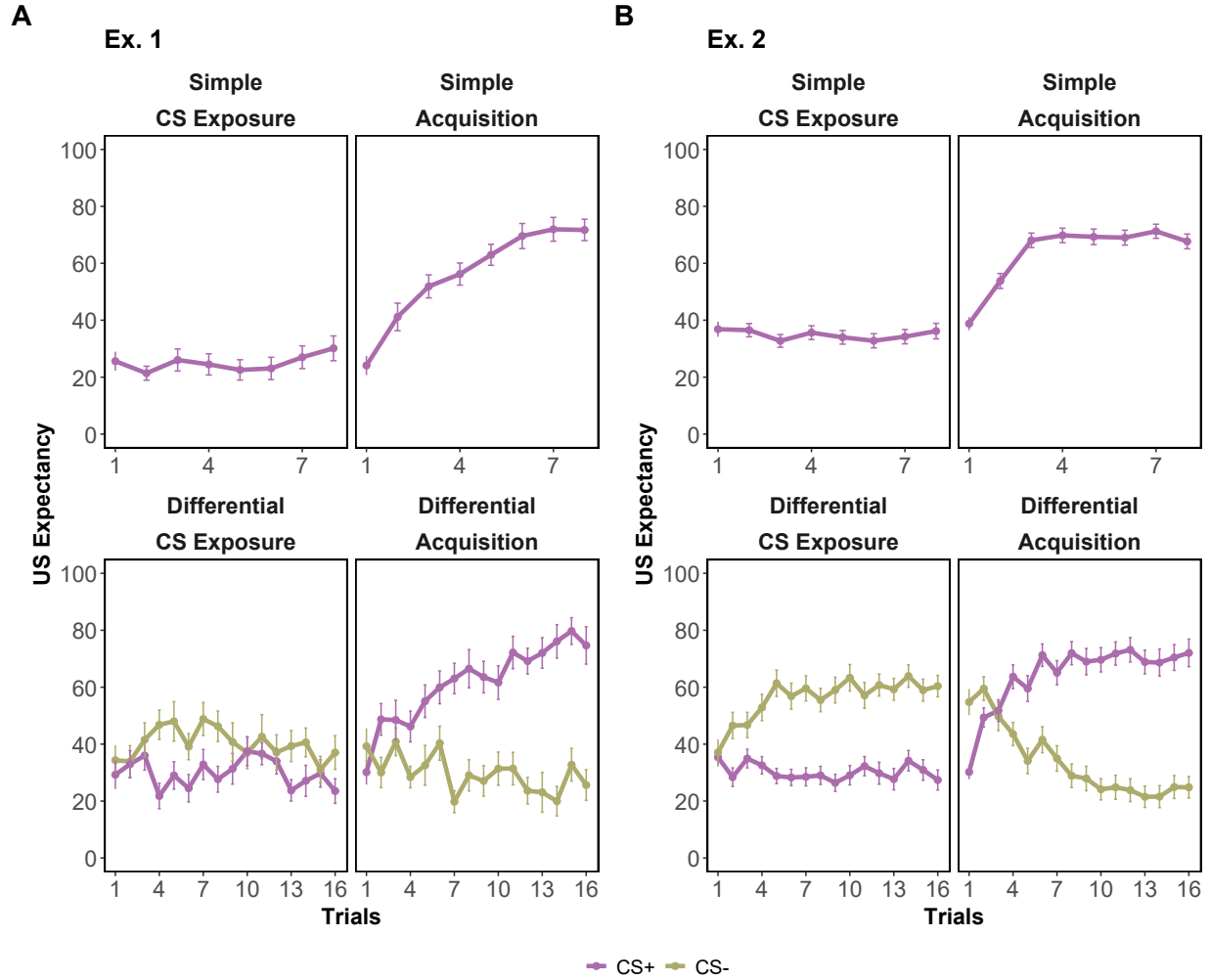
### ***Statistical inference***

For all mixed-effects models, statistical inferences were performed using the **brms** package (Bürkner, 2017) in R (R Core Team, 2021), which utilizes Hamiltonian Monte Carlo (HMC) through Stan (Carpenter et al., 2017), specifically employing the No-U-Turn Sampler (NUTS). For the error-driven learning model, JAGS (Plummer, 2003) was used for sampling, interfaced with R via the **jagsUI** package (Kellner, 2021). We employed four

Markov Chain Monte Carlo (MCMC) chains, each with 20,000 iterations. To ensure robust parameter estimates, the first 10,000 iterations of each chain were discarded as burn-in, resulting in a final sample of 40,000 samples per parameter. Convergence of the parameter estimates was assessed by visually inspecting the chains for irregular patterns and checking the  $\hat{R}$  value based on the Gelman-Rubin diagnostic (Brooks & Gelman, 1998; Gelman & Rubin, 1992), with a threshold of  $\hat{R} < 1.1$  for acceptable convergence. All post-sampling analyses were performed in R.

## Results

## Associative learning

**Figure 2**

US expectancy patterns for Experiment 1 (laboratory) and Experiment 2 (online). Each panel illustrates the group mean of US expectancy over trials for CS+ (S3) and CS- (S8) stimuli during the CS exposure and acquisition phases. Error bars represent the standard error.

***CS exposure***

In the CS exposure phase (Figure 2), without receiving any unconditioned stimulus (US), participants responded randomly without a specific pattern following CS repetitions for both simple (Experiment 1:  $\beta = 0.75$ , 95% CI [-0.39, 1.88], BF = 0.03; Experiment 2:  $\beta = -0.27$ , 95% CI [-0.98, 0.44], BF = 0.02) and differential (Experiment 1:  $\beta = -0.42$ , 95%CI[-1.50, 0.66], BF = 0.02; Experiment 2:  $\beta = -0.24$ , 95%CI[-1.05, 0.57], BF = 0.02) conditioning groups in both experiments.

In the differential group, an interesting trend emerged: without forming any learned associations, participants tended to give higher US expectancy ratings to the larger stimulus (S8). This trend was observed consistently across both Experiment 1 ( $\beta = 10.49$ , 95% CI [2.57, 18.29], BF = 2.27) and Experiment 2 ( $\beta = 17.97$ , 95% CI [12.22, 23.70], BF > 100). This pattern may suggest a natural tendency to perceive larger stimuli as more threatening compared to smaller ones.

***Conditioning***

Overall, participants exhibited an increase in US expectancy with repeated CS exposures in both the simple (Experiment 1:  $\beta = 5.66$ , 95% CI [4.09, 7.21], BF > 100; Experiment 2:  $\beta = 3.60$ , 95% CI [2.59, 4.60], BF > 100) and differential (Experiment 1:  $\beta = 5.84$ , 95% CI [4.33, 7.35], BF > 100; Experiment 2:  $\beta = 5.16$ , 95% CI [4.05, 6.28], BF > 100) conditioning groups. Additionally, differential learning was evident in the decreasing US expectancy with more CS- (S8) repetitions during the conditioning phase (Experiment 1:  $\beta = -7.37$ , 95% CI [-9.51, -5.26], BF > 100; Experiment 2:  $\beta = -10.57$ , 95% CI [-12.14, -9.01], BF > 100).

When further exploring group differences, the analysis indicated that all fixed and interaction effects involving the conditioning group had Bayes factors smaller than 1, suggesting no significant difference between the conditioning groups in terms of changes in US expectancy.

## Identification

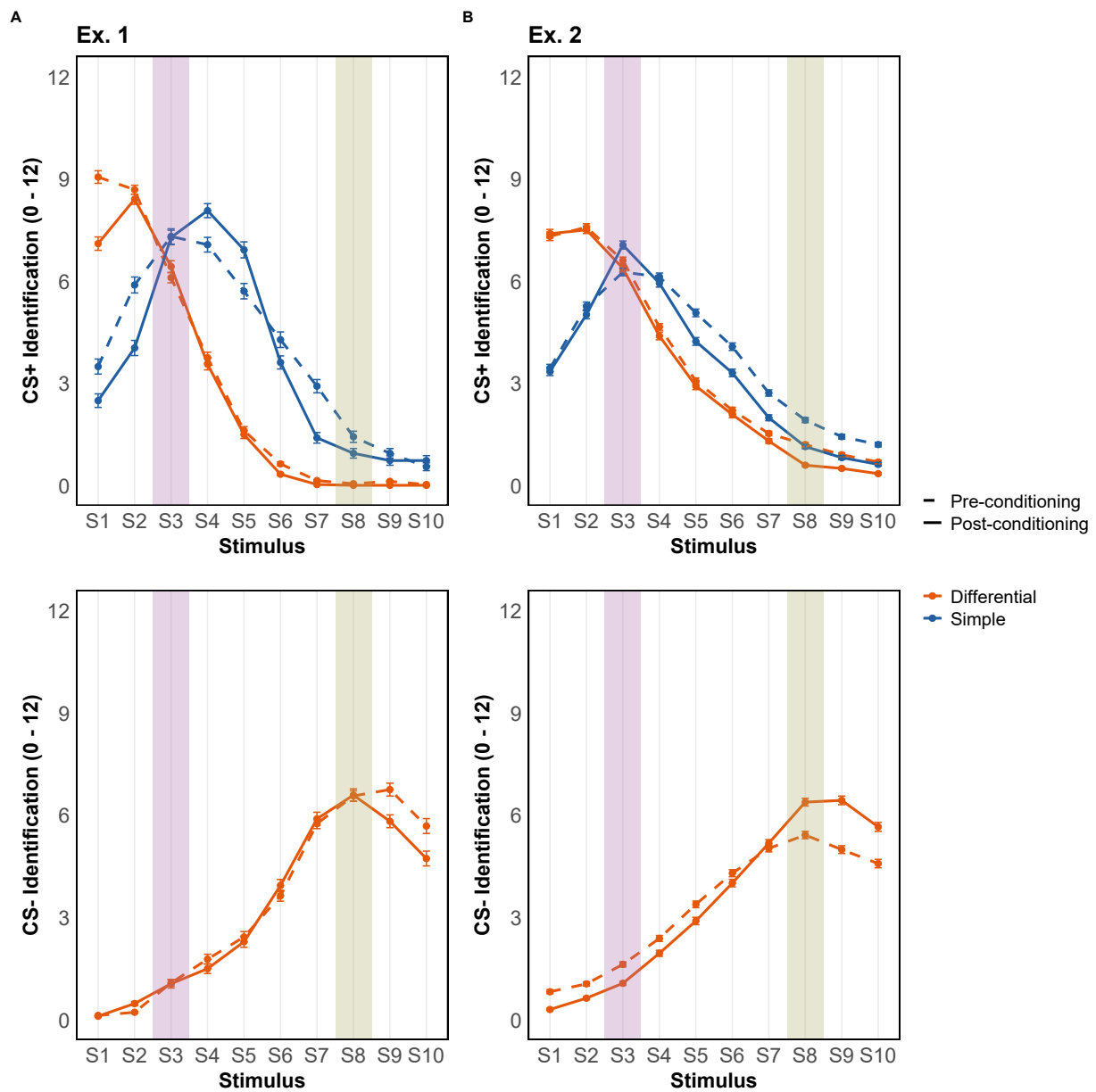


Figure 3

Identification patterns for Experiment 1 (laboratory) and Experiment 2 (online). Each panel illustrates the group mean of CS identification (0 - 12) for each stimulus during the pre-conditioning and post-conditioning identification phases. Error bars represent the standard errors

*Pre-conditioning*

In the pre-conditioning identification phase for the simple conditioning group (Figure 3), the marginal effects showed that the stimulus most frequently identified as the CS+ (S3) was the CS+ itself in both experiments (Experiment 1: 7.30, 95% CI [6.62, 8.00],  $BF > 100$ ; Experiment 2: 6.25, 95% CI [5.93, 6.57],  $BF > 100$ ). However, the difference in identification between S3 and S4 was negligible (Experiment 1: 0.24, 95% CI [-0.24, 0.73],  $BF < 0.01$ ; Experiment 2: 0.14, 95% CI [-0.12, 0.39],  $BF < 0.01$ ). The likelihood of identifying a stimulus as the CS+ decreased as it became more physically different from the CS+.

In the differential conditioning group, a stimulus contrast effect was observed, where stimuli smaller than the CS+ (S3) were more frequently identified as the CS+ than the CS+ itself. In both experiments, S1 (Experiment 1: 9.05, 95% CI [8.73, 9.36],  $BF > 100$ ; Experiment 2: 7.31, 95% CI [6.98, 7.62],  $BF > 100$ ) and S2 (Experiment 1: 8.67, 95% CI [8.36, 8.99],  $BF > 100$ ; Experiment 2: 7.57, 95% CI [7.25, 7.88],  $BF > 100$ ) were the stimuli most often misidentified as the CS+. The contrasts between S1 and S3 (Experiment 1: 2.95, 95% CI [2.66, 3.25],  $BF > 100$ ; Experiment 2: 0.71, 95% CI [0.49, 0.93],  $BF > 100$ ), and between S2 and S3 (Experiment 1: 2.58, 95% CI [2.28, 2.88],  $BF > 100$ ; Experiment 2: 0.97, 95% CI [0.75, 1.19],  $BF > 100$ ), demonstrated that S1 and S2 were significantly more likely to be identified as the CS+ than the CS+ itself. Moreover, as shown in Figure 3, an asymmetric identification gradient appeared in the differential conditioning group, with the misidentifications of S1 and S2 as CS+ being more frequent than those of S4 and S5 in both Experiment 1 ( $\frac{S1+S2}{2}$  vs.  $\frac{S4+S5}{2}$ : 6.19, 95% CI [5.98, 6.40],  $BF > 100$ ) and Experiment 2 ( $\frac{S1+S2}{2}$  vs.  $\frac{S4+S5}{2}$ : 3.59, 95% CI [3.43, 3.74],  $BF > 100$ ).

Regarding CS- (S8) identification in the differential group, the stimuli most frequently identified as the CS- were S8 itself (Experiment 1: 6.55, 95% CI [6.14, 6.97],  $BF > 100$ ; Experiment 2: 5.42, 95% CI [5.12, 5.71],  $BF > 100$ ) and S9 (Experiment 1: 6.74, 95% CI [6.33, 7.15],  $BF > 100$ ; Experiment 2: 4.99, 95% CI [4.70, 5.28],  $BF > 100$ ), with

no significant difference between the two in Experiment 1 (S8 vs. S9: -0.19, 95% CI [-0.59, 0.22],  $BF < 0.01$ ) and a small difference in Experiment 2 (S8 vs. S9: 0.43, 95% CI [0.18, 0.67],  $BF = 1.49$ ).

### ***Post-conditioning***

After conditioning, the simple conditioning group exhibited significant negative effects on S1 (-1.00, 95% CI [-1.49, -0.52],  $BF > 100$ ), S2 (-1.85, 95% CI [-2.34, -1.36],  $BF > 100$ ), and S7 (-1.51, 95% CI [-2.00, -1.03],  $BF > 100$ ) in Experiment 1, as well as on S5 through S10 in Experiment 2, with median estimates ranging from -0.58 to -0.84. In contrast, significant positive effects were observed on S4 (1.00, 95% CI [0.51, 1.48],  $BF > 100$ ) and S5 (1.21, 95% CI [0.72, 1.70],  $BF > 100$ ) in Experiment 1, and on S3 (0.79, 95% CI [0.54, 1.05],  $BF > 100$ ) in Experiment 2. Despite these effects, the magnitudes of change were relatively minor, and the overall trend of identification gradients remained similar to the pre-conditioning phase. In Experiment 1, S3 and S4 continued to be the stimuli most frequently identified as the CS+. However, in Experiment 2, S3 was more often identified as the CS+ compared to S4 (S3 post vs. S4 post: 1.12, 95% CI [0.87, 1.37],  $BF > 100$ ).

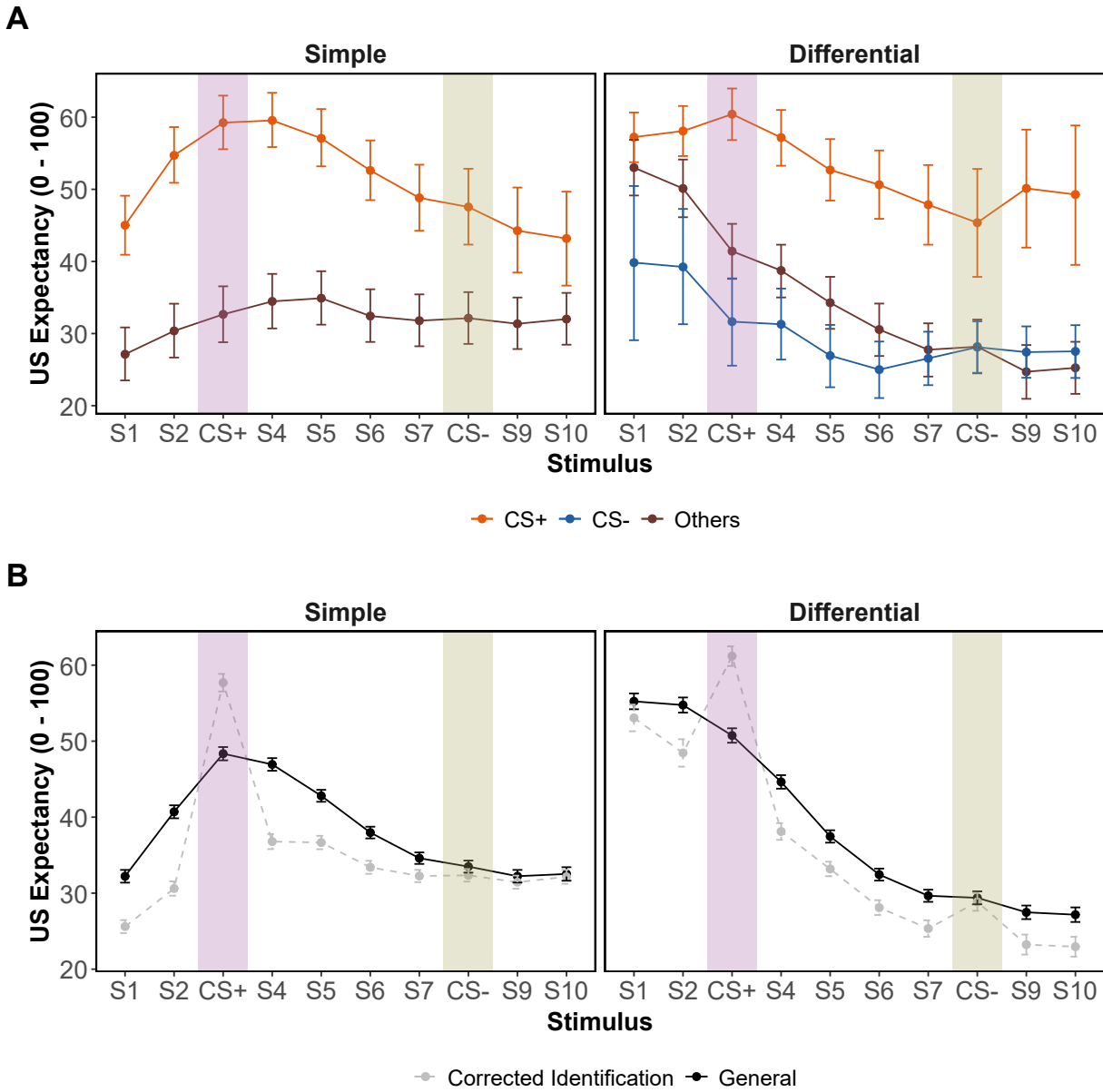
In the differential conditioning group, the conditioning phase had a significant effect only on S1 in Experiment 1 (-1.95, 95% CI [-2.26, -1.66],  $BF > 100$ ) and S9 in Experiment 2 (-0.40, 95% CI [-0.62, -0.18],  $BF > 100$ ). As a result, the identification gradients remained highly similar to those observed during the pre-conditioning phase. Despite a decrease in the identification of S1 as the CS+ in Experiment 1, S1 continued to be identified as the CS+ more frequently than the CS+ itself (S1 post vs. S3 post, 1.02, 95% CI [0.80, 1.23],  $BF > 100$ ). For CS- identification, the overall pattern remained consistent, with S8 and S9 continuing to be the most frequently identified as the CS- after conditioning. While minor negative effects were observed for S9 and S10 (median estimates: -0.93 and -0.95) in Experiment 1, and for S1 to S6 (median estimates: -0.29 to -0.56) in Experiment 2, positive effects were noted for S8 to S10 (median estimates: 0.97 to 1.44) in Experiment 2. Despite these subtle changes, the general identification trends were

largely unaffected by the conditioning process.

**Generalization and identification**

In Experiment 2, US expectancy was additionally measured during the post-conditioning identification task.



**Figure 4**

Panel A: Marginal effects of stimulus identification on US expectancy, conditioned on whether the stimulus is identified as the CS or as a different stimulus. Panel B: Generalization gradients comparing responses with and without identification errors. Error bars represent the standard errors.

For both the simple and differential conditioning groups, marginal effects indicated that CS+ identification consistently led to higher responses across all stimuli (panel A,

Figure 4). This suggests that participants tended to respond more strongly when they identified the encountered stimulus as the CS+ compared to when they identified it as a different test stimulus. Conversely, in the differential conditioning group, lower marginal effects were observed for CS- identification, indicating that responses were reduced when the presented stimulus was identified as the CS-. To assess whether including stimulus identification improved the predictive power of generalization behavior, we compared models using stimulus alone as a predictor with models incorporating both stimulus and stimulus identification. This comparison was performed using Leave-One-Out Cross-Validation (LOO) to compute the Expected Log Predictive Density (ELPD). The results showed that, for both the simple conditioning group (stimulus-only model: ELPD = -850.6, SE = 46.5) and the differential conditioning group (stimulus-only model: ELPD = -322.7, SE = 28.4), the inclusion of stimulus identification significantly improved the predictive accuracy of the model for generalization behavior. Additionally, the Bayesian  $R^2$  indicated that including stimulus identification in the model accounted for an additional 8% of the variance in the simple conditioning group (stimulus-only model:  $R^2 = 0.36$ , 95% CI [0.35, 0.37]; full model:  $R^2 = 0.44$ , 95% CI [0.43, 0.45]) and an additional 3% of the variance in the differential conditioning group (stimulus-only model:  $R^2 = 0.32$ , 95% CI [0.31, 0.33]; full model:  $R^2 = 0.35$ , 95% CI [0.34, 0.36]).

Next, similar to the previous study (Zaman et al., 2021), we compared the generalization gradients both with and without accounting for perceptual errors (panel B, Figure 4). When perceptual errors were excluded, the CS+ elicited a stronger response (Simple: 58.92, 95% CI [54.76, 63.07]; Differential: 60.02, 95% CI [55.99, 63.99]) compared to when perceptual errors were included (Simple: 48.46, 95% CI [45.01, 51.99]; Differential: 51.02, 95% CI [47.65, 54.44]). In the differential conditioning group, a peak shift effect was evident, with the highest response shifting from the CS+ to S1 (S1 vs. S3: 4.45, 95% CI [2.23, 6.65], BF = 33) and S2 (S2 vs. S3: 3.77, 95% CI [1.55, 5.98], BF = 3.03). However, this peak shift effect vanished when only the responses without perceptual errors were

analyzed, with a notable reversal in the trend (S1 vs. S3: -5.61, 95% CI [-8.57, -2.67],  $BF = 33$ ; S2 vs. S3: -8.67, 95% CI [-11.75, -5.61],  $BF > 100$ ). An area shift was observed in the differential conditioning group, both when generalization gradients were analyzed with ( $\frac{S1+S2}{2} - \frac{S4+S5}{2}$ : 14.29, 95% CI [12.74, 15.87],  $BF > 100$ ) and without ( $\frac{S1+S2}{2} - \frac{S4+S5}{2}$ : 15.27, 95% CI [13.04, 17.52],  $BF > 100$ ) accounting for perceptual errors. In both cases, responses were stronger in the region away from the CS- compared to the region toward the CS-.

## Discussion

In this research, we explored whether exposure to two contrasting stimuli could generate a biased mental representation, offering an alternative explanation for the observed shift in peak conditioned responses following differential conditioning. Across two experiments, our findings demonstrated that when participants were exposed to two contrasting circle sizes in a differential conditioning paradigm, the peak identification of the CS+ shifted away from the CS+ itself and toward the direction opposite to the CS-. Notably, this perceptual contrast effect emerged even before CS-US associative learning took place and remained largely intact after subsequent associative learning. Furthermore, consistent with the findings of Zaman et al. (2021), we observed that the peak shift effect disappeared when we considered only generalized responding for correct stimulus identifications. These results highlight the intricate relationship between mental representation and generalization behavior. They suggest that the patterns of post-learning identification and generalization observed in previous studies (Zaman et al., 2021) and in the current study were significantly shaped by a stimulus contrast effect resulting from CS exposure, underscoring the critical influence of perceptual mechanisms in these behaviors.

Our findings not only challenge the traditional theoretical account that the peak shift effect is primarily driven by the interaction between excitatory and inhibitory strength distributions, but they also highlight a more fundamental insight: generalization behavior is shaped by the interplay of multiple inferential systems. Generalization has long been conceptualized as a cognitive inferential process, operating within a mental space

where behavior is inferred based on the perceived distances between stimuli (Austerweil et al., 2019; Douven et al., 2023; Osherson et al., 1990; Shepard, 1957, 1987; Tenenbaum & Griffiths, 2001). Typically, shorter distances result in greater behavioral transfer. However, this model assumes that the mental distance between physical stimuli remains fixed regardless of changing contexts, thus lacking any dynamic or inferential adjustments. The traditional associative learning explanation of the peak shift effect aligns with this fixed-distance model, suggesting that the distribution of excitatory and inhibitory strengths is primarily determined by static stimulus features (Blough, 1975; Ghirlanda & Enquist, 1998; McLaren & Mackintosh, 2002; Spence, 1937; Thomas & Thomas, 1974). Alternatively, adaptation theory explains the peak shift effect by suggesting that averaged representations shift with each newly encountered stimulus (Thomas & Switalski, 1966; Thomas & Thomas, 1974). However, this theory oversimplifies the process by applying a fixed rule for representational shifts, overlooking the context-dependent nature of representation. Such assumptions are increasingly seen as inadequate, particularly in light of recent advances that highlight the inferential and context-sensitive aspects of human mental representation. Current research suggests that the encoding, storage, and retrieval of physical stimuli follow a more complex and context-sensitive structure (Bays et al., 2024; Fougner et al., 2012; Petzschnner et al., 2015; Purves et al., 2014). In fact, research in categorization has emphasized the impact of context on the formation of mental coordinates for similarity judgments (Goldstone et al., 1997; Medin et al., 1993; Tversky, 1977). However, these contextual effects are often examined within the scope of higher-order cognitive processes, rather than the fundamental representation of physical inputs. Our results demonstrate that perceptual contrast effects play a crucial role in shaping generalization patterns. This suggests that to fully understand generalization, it is essential to consider the interaction of multiple cognitive systems. Furthermore, the perceptual contrast effect observed in our study bears resemblance to the caricature effects found in categorization (Ameel & Storms, 2006; Davis & Love, 2010; Goldstone, 1996; Goldstone et al., 2003; Palmeri & Nosofsky,

2001), where categories are often represented not by their prototypes, but by exaggerated features that are distinct from contrasting categories. This connection opens up possibilities for applying our findings to more complex, multi-dimensional stimuli.

Moreover, as emphasized in previous studies (Zaman, Struyf, Ceulemans, et al., 2019; Zaman, Ceulemans, et al., 2019; Zaman, Yu, Andreatta, et al., 2023; Zaman et al., 2022; Zenses et al., 2021), identification errors are common in both simple and differential conditioning groups. Individuals often misidentify different stimuli as the conditioned stimulus (CS+), and these perceptual errors are strongly associated with generalized responses. Specifically, greater generalization tends to occur when a stimulus is mistakenly identified as the CS+, whereas less generalization is observed when it is not. In the current study, similar findings emerged, where the inclusion of stimulus identification as a predictor significantly improved the predictive accuracy of US expectancy. Yet, previous experimental designs concurrently measured stimulus identifications and generalization post-learning, leaving open the possibility that their strong relationship stemmed from a shared learning mechanism. The present findings indicate that perceptual errors emerged immediately after CS exposure and persisted largely unchanged after conditioning, suggesting that these errors were shaped by an independent perceptual process rather than by associative learning per se. This underscores the importance of recognizing the distinct contributions of perception to generalization, independent of the learning mechanisms that may also influence these behaviors. The evidence from our study highlights the critical role of perceptual errors in shaping generalized responses and suggests that these errors represent a key factor in the interaction between perception and generalization, rather than being merely a byproduct of associative learning.

Recent research exploring the relationship between associative aversive learning and identification ability has yielded inconsistent findings (Åhs et al., 2013; Li et al., 2008; Resnik et al., 2011; Shalev et al., 2018; Zaman, Yu, Andreatta, et al., 2023). In our study, we did not observe a clear trend in changes in stimulus identification following fear learning

in either Experiment 1, which involved aversive electrical stimulation, or Experiment 2, which used a visual stimulus. Specifically, the misidentification patterns across stimuli remained largely unchanged for both the simple and differential conditioning groups. The absence of changes in identification patterns cannot be attributed to a lack of learning, as successful learning was evident through both the observed patterns in the mixed-effects model and the latent learning processes revealed by the error-driven learning model, at both the group and individual levels. However, as previously discussed, identification behavior involves complex representational processes that rely on both recent perceptual representations and the memory representation of the targeted stimulus. The alternative forced-choice task used in this study may be limited in its ability to capture subtle changes in either perception or memory. This limitation leaves open the possibility that perceptual or memory representations in the current study were indeed modulated by the learning process, causing the mental representations of stimuli to move closer together or further apart in the mental space following learning. Future research is needed to explore the interaction between mental representation and aversive learning more deeply, aiming to disentangle these processes and clarify how they jointly contribute to the observed outcomes after learning.

Overall, the current findings provide evidence that the mental representation process, as a system distinct from associative learning, can produce the peak shift effect. However, this evidence does not suggest that the peak shift effect or other generalization phenomena are solely generated by the perceptual system. Instead, it advocates for an integrative approach to understanding generalization behavior. This perspective suggests that, rather than examining perception (Zaman, Struyf, Ceulemans, et al., 2019; Zaman, Ceulemans, et al., 2019; Zenses et al., 2021) and cognition (Boddez et al., 2017; Dunsmoor & Murphy, 2015; Lee & Livesey, 2018) as isolated ends of a spectrum, we should consider them as intertwined processes. By conceptualizing the spectrum as a continuum that can be folded, we can better explore learning and post-learning behaviors as joint, interactive

phenomena. Our recent work (Yu et al., 2023) proposed a computational model that integrates dynamic perceptual estimation data into the traditionally pure cognitive similarity generalization function. Nevertheless, the assumptions in the model are limited by our current understanding of the different processes underlying generalization behavior. Future research should empirically explore how various processes and systems jointly shape behavior. These investigations would serve as the foundation for developing more comprehensive formal models of human generalization.

### **Limitation**

Experiment 2 was conducted in an online environment, where we calibrated participants’ screen sizes to ensure consistent circle sizes across all participants. However, unlike the tightly controlled conditions of Experiment 1, we had limited control over certain factors in the online setting that could potentially influence perception, such as head position and distance from the screen. These uncontrolled variables could introduce additional measurement errors in Experiment 2. Despite these challenges, the replication of findings from the more controlled laboratory experiment 1 to a less controlled online environment strengthens the robustness and generalizability of our results.

Additionally, in Experiment 1, participants received electrical stimuli to elicit emotional fear responses. However, our measurements were limited to self-reported behavioral data for assessing learning and generalization behavior, lacking additional biological indicators such as physiological or neurological measurements. This limitation constrains our ability to fully understand the impact of fear learning on changes in identification patterns. Previous research has shown that different response channels can yield inconsistent results concerning fear learning (LeDoux & Brown, 2017; LeDoux & Pine, 2016; Lipp & Purkis, 2005) and perception (Rossi & Berglund, 2011). A unified approach that includes both behavioral and biological measurements would likely provide a more comprehensive understanding of the interactions between fear learning and perception, offering a more holistic view of the phenomena investigated in the current study.

### Constraints on generality

The results are anticipated to generalize to other types of stimuli with one-dimensional sensory features, such as orientation and color. Additionally, we anticipate that these findings will extend to more complex, multi-dimensional stimuli. This expectation is supported by evidence from categorization studies, which have shown that categorization behavior can be influenced by exposure to previously encoded contrasting elements (Ameel & Storms, 2006; Davis & Love, 2010; Goldstone, 1996; Goldstone et al., 2003; Palmeri & Nosofsky, 2001). For example, face recognition patterns have been found to be modulated by the sequence of previously encountered faces (Cabeza et al., 1999). We also have no reasons to believe that the findings will be constrained by different sample populations.

### Acknowledgements

KY is supported by an FWO research project (co-PI: JZ, G079520N). JZ is a Postdoctoral Research Fellow of the Research Foundation Flanders (FWO, 12P8623N) and received funding from the Alexander von Humboldt Stiftung. The research leading to the results reported in this paper was also supported in part by the Research Fund of KU Leuven (C14/19/054 and C16/19/002). Material infrastructure for Experiment 1 was supported by an infrastructure grant of the Research Foundation Flanders (FWO) and the Research Fund of KU Leuven, Belgium (I011320N; AKUL/19/06). Further resources and services used in this work were 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 statement

**Kenny Yu:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review and editing. **Steven Verheyen:** Conceptualization, Writing—review and editing. **Tom Beckers:** Conceptualization, Writing—review and editing. **Wolf Vanpaemel:**



Conceptualization, Supervision, Writing—review and editing. **Francis Tuerlinckx:**

Conceptualization, Supervision, Writing—review and editing. **Jonas Zaman:**

Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review and editing.

### **Data availability**

The raw and processed data for the experiments in this study can be accessed at the following Open Science Framework (OSF) repository: <https://osf.io/a8c4g/>.

### **Code availability**

The code for the computational model and analysis, as well as supplementary information with additional information about the model and results, can be found at the same repository as the data: <https://osf.io/a8c4g/>. The Bayesian sampling was conducted with JAGS (version 4.3.1) and stan (version 2.32.2), and the post-sampling analysis and visualization were conducted with R (version 4.3.0).

### **Competing interests**

The authors declare no competing interests. The evaluation, opportunities for promotion, and ability to obtain research funding of KY, SV, TB, WV, FT and JZ are partly dependent on the number of articles they publish.

## References

- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, 7(10), 1057–1058.  
<https://doi.org/10.1038/nn1312>
- Åhs, F., Miller, S. S., Gordon, A. R., & Lundström, J. N. (2013). Aversive learning increases sensory detection sensitivity. *Biological Psychology*, 92(2), 135–141.  
<https://doi.org/10.1016/j.biopsycho.2012.11.004>
- Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience*, 25(1), 339–379.  
<https://doi.org/10.1146/annurev.neuro.25.112701.142900>
- Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language*, 55(3), 402–421.  
<https://doi.org/10.1016/j.jml.2006.05.005>
- Aru, J., Rutiku, R., Wibral, M., Singer, W., & Melloni, L. (2016). Early effects of previous experience on conscious perception. *Neuroscience of Consciousness*, niw004.  
<https://doi.org/10.1093/nc/niw004>
- Austerweil, J. L., Sanborn, S., & Griffiths, T. L. (2019). Learning How to Generalize. *Cognitive Science*, 43(8). <https://doi.org/10.1111/cogs.12777>
- Bae, G.-Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, & Psychophysics*, 79(8), 2376–2395.  
<https://doi.org/10.3758/s13414-017-1404-8>
- Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144(4), 744–763. <https://doi.org/10.1037/xge0000076>

- Bays, P. M., Schneegans, S., Ma, W. J., & Brady, T. F. (2024). Representation and computation in visual working memory. *Nature Human Behaviour*, 8(6), 1016–1034. <https://doi.org/10.1038/s41562-024-01871-2>
- Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 1(1), 3–21. <https://doi.org/10.1037/0097-7403.1.1.3>
- Boddez, Y., Bennett, M. P., Van Esch, S., & Beckers, T. (2017). Bending rules: The shape of the perceptual generalisation gradient is sensitive to inference rules. *Cognition and Emotion*, 31(7), 1444–1452. <https://doi.org/10.1080/02699931.2016.1230541>
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384–392. <https://doi.org/10.1177/0956797610397956>
- Brady, T. F., & Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, 15(15), 6. <https://doi.org/10.1167/15.15.6>
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455. <https://doi.org/10.1080/10618600.1998.10474787>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. *Memory & Cognition*, 27(1), 139–151. <https://doi.org/10.3758/BF03201220>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).

- Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, 22(7), 622–626. <https://doi.org/10.1016/j.cub.2012.02.021>
- Clevenger, P. E., & Hummel, J. E. (2014). Working memory for relations among objects. *Attention, Perception, & Psychophysics*, 76(7), 1933–1953. <https://doi.org/10.3758/s13414-013-0601-3>
- Davis, T., & Love, B. C. (2010). Memory for Category Information Is Idealized Through Contrast With Competing Options. *Psychological Science*, 21(2), 234–242. <https://doi.org/10.1177/0956797609357712>
- Dickey, J. M. (1971). The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics*, 42(1), 204–223. <https://doi.org/10.1214/aoms/1177693507>
- Douven, I., Verheyen, S., Elqayam, S., Gärdenfors, P., & Osta-Vélez, M. (2023). Similarity-based reasoning in conceptual spaces. *Frontiers in Psychology*, 14, 1234483. <https://doi.org/10.3389/fpsyg.2023.1234483>
- Dunsmoor, J. E., & Murphy, G. L. (2015). Categories, concepts, and conditioning: How humans generalize fear. *Trends in Cognitive Sciences*, 19(2), 73–77. <https://doi.org/10.1016/j.tics.2014.12.003>
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3(1), 1229. <https://doi.org/10.1038/ncomms2237>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4). <https://doi.org/10.1214/ss/1177011136>
- Ghirlanda, S., & Enquist, M. (1998). Artificial neural networks as models of stimulus control. *Animal Behaviour*, 56(6), 1383–1389. <https://doi.org/10.1006/anbe.1998.0903>
- Ghirlanda, S., & Enquist, M. (1999). The geometry of stimulus control. *Animal Behaviour*, 58(4), 695–706. <https://doi.org/10.1006/anbe.1999.1187>

- Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour*, *66*(1), 15–36. <https://doi.org/10.1006/anbe.2003.2174>
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*(5), 608–628. <https://doi.org/10.3758/BF03201087>
- Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition*, *25*(2), 237–255. <https://doi.org/10.3758/BF03201115>
- Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31*(2), 169–180. <https://doi.org/10.3758/BF03194377>
- Guo, K., Nevado, A., Robertson, R. G., Pulgarin, M., Thiele, A., & Young, M. P. (2004). Effects on orientation perception of manipulating the spatio-temporal prior probability of stimuli. *Vision Research*, *44*(20), 2349–2358. <https://doi.org/10.1016/j.visres.2004.04.014>
- Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *58*(5), 321–334. <https://doi.org/10.1037/h0042606>
- Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior*. New York.
- Honig, W. K., & Urcuioli, P. J. (1981). The legacy of guttman and kalish (1956): Twenty-five years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior*, *36*(3), 405–445. <https://doi.org/10.1901/jeab.1981.36-405>
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kellner, K. (2021). *Jagsui: A wrapper around 'rjags' to streamline 'jags' analyses* [R package version 1.5.2]. <https://CRAN.R-project.org/package=jagsUI>
- Lange, I., Goossens, L., Michielse, S., Bakker, J., Lissek, S., Papalini, S., Verhagen, S., Leibold, N., Marcelis, M., Wichers, M., Lieveise, R., van Os, J., van Amelsvoort, T., & Schruers, K. (2017). Behavioral pattern separation and its link to the neural

- mechanisms of fear generalization. *Social Cognitive and Affective Neuroscience*, 12(11), 1720–1729. <https://doi.org/10.1093/scan/nsx104>
- LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, 114(10). <https://doi.org/10.1073/pnas.1619316114>
- LeDoux, J. E., & Pine, D. S. (2016). Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework. *American Journal of Psychiatry*, 173(11), 1083–1093. <https://doi.org/10.1176/appi.ajp.2016.16030353>
- Lee, J. C., & Livesey, E. J. (2018). Rule-based generalization and peak shift in the presence of simple relational rules (F. A. Soto, Ed.). *PLOS ONE*, 13(9), e0203805. <https://doi.org/10.1371/journal.pone.0203805>
- Lee, J. C., Lovibond, P. F., Hayes, B. K., & Navarro, D. J. (2019). Negative evidence and inductive reasoning in generalization of associative learning. *Journal of Experimental Psychology: General*, 148(2), 289–303. <https://doi.org/10.1037/xge0000496>
- Li, W., Howard, J. D., Parrish, T. B., & Gottfried, J. A. (2008). Aversive Learning Enhances Perceptual and Cortical Discrimination of Indiscriminable Odor Cues. *Science*, 319(5871), 1842–1845. <https://doi.org/10.1126/science.1152837>
- Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition and Emotion*, 19(2), 269–282. <https://doi.org/10.1080/02699930441000319>
- Lissek, S. (2012). Toward an account of clinical anxiety predicated on basic, neurally-mapped mechanisms of pavlovian fear-learning: The case for conditioned overgeneralization. *Depression and anxiety*, 29(4), 257–263. <https://doi.org/10.1002/da.21922>
- Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, R. P., Pine, D. S., & Grillon, C. (2008). Generalization of conditioned fear-potentiated startle in humans:

- Experimental validation and clinical relevance. *Behaviour Research and Therapy*, 46(5), 678–687. <https://doi.org/10.1016/j.brat.2008.02.005>
- Lissek, S., Bradford, D. E., Alvarez, R. P., Burton, P., Espensen-Sturges, T., Reynolds, R. C., & Grillon, C. (2014). Neural substrates of classically conditioned fear-generalization in humans: A parametric fMRI study. *Social Cognitive and Affective Neuroscience*, 9(8), 1134–1142. <https://doi.org/10.1093/scan/nst096>
- McCarthy, J. D., Kupitz, C., & Caplovitz, G. P. (2013). The Binding Ring Illusion: Assimilation affects the perceived size of a circular array. *F1000Research*, 2, 58. <https://doi.org/10.12688/f1000research.2-58.v2>
- McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. generalization and discrimination. *Animal Learning & Behavior*, 30(3), 177–200. <https://doi.org/10.3758/BF03192828>
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100(2), 254–278. <https://doi.org/10.1037/0033-295X.100.2.254>
- Mednick, S. A., & Freedman, J. L. (1960). Stimulus generalization. *Psychological Bulletin*, 57(3), 169–200. <https://doi.org/10.1037/h0041650>
- Mruczek, R. E. B., Blair, C. D., Strother, L., & Caplovitz, G. P. (2017, June). Size contrast and assimilation in the delboeuf and ebbinghaus illusions. In A. G. Shapiro & D. Todorovic (Eds.), *The oxford compendium of visual illusions* (1st ed., pp. 262–268). Oxford University Press New York. <https://doi.org/10.1093/acprof:oso/9780199794607.003.0028>
- Ng, D. W., Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2022). Generalization following symmetrical intradimensional discrimination training. *Journal of Experimental Psychology: Animal Learning and Cognition*, 48(3), 179–189. <https://doi.org/10.1037/xan0000327>

- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11), 5–5. <https://doi.org/10.1167/14.11.5>
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, *10*(9), 16–16. <https://doi.org/10.1167/10.9.16>
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*(2), 185–200. <https://doi.org/10.1037/0033-295X.97.2.185>
- Palmeri, T. J., & Nosofsky, R. M. (2001). Central Tendencies, Extreme Points, and Prototype Enhancement Effects in Ill-Defined Perceptual Categorization. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 197–235. <https://doi.org/10.1080/02724980042000084>
- Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293. <https://doi.org/10.1016/j.tics.2015.03.002>
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Working Papers*.
- Purtle, R. B. (1973). Peak shift: A review. *Psychological Bulletin*, *80*(5), 408–421. <https://doi.org/10.1037/h0035233>
- Purves, D., Monson, B. B., Sundararajan, J., & Wojtach, W. T. (2014). How biological vision succeeds in the physical world. *Proceedings of the National Academy of Sciences*, *111*(13), 4750–4755. <https://doi.org/10.1073/pnas.1311309111>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental*



*Psychology: Human Perception and Performance*, 41(6), 1650–1665.

<https://doi.org/10.1037/xhp0000110>

Rescorla, R., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory* (pp. 64–69, Vol. Vol. 2). New York: Appleton-Century-Crofts.

Resnik, J., & Paz, R. (2015). Fear generalization in the primate amygdala. *Nature Neuroscience*, 18(2), 188–190. <https://doi.org/10.1038/nn.3900>

Resnik, J., Sobel, N., & Paz, R. (2011). Auditory aversive learning increases discrimination thresholds. *Nature Neuroscience*, 14(6), 791–796. <https://doi.org/10.1038/nn.2802>

Rossi, G. B., & Berglund, B. (2011). Measurement involving human perception and interpretation. *Measurement*, 44(5), 815–822.

<https://doi.org/10.1016/j.measurement.2011.01.016>

Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports*, 8(1), 6606.

<https://doi.org/10.1038/s41598-018-25093-5>

Shalev, L., Paz, R., & Avidan, G. (2018). Visual Aversive Learning Compromises Sensory Discrimination. *The Journal of Neuroscience*, 38(11), 2766–2779.

<https://doi.org/10.1523/JNEUROSCI.0889-17.2017>

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22(4), 325–345.

<https://doi.org/10.1007/BF02288967>

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317–1323. <https://doi.org/10.1126/science.3629243>

- Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D., & Melloni, L. (2015). How previous experience shapes perception in different sensory modalities. *Frontiers in Human Neuroscience*, *9*. <https://doi.org/10.3389/fnhum.2015.00594>
- Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, *44*(5), 430–444. <https://doi.org/10.1037/h0062885>
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640. <https://doi.org/10.1017/S0140525X01000061>
- Thomas, D. R., & Switalski, R. W. (1966). Comparison of stimulus generalization following variable-ratio and variable-interval training. *Journal of Experimental Psychology*, *71*(2), 236–240. <https://doi.org/10.1037/h0022880>
- Thomas, D. R., & Thomas, D. H. (1974). Stimulus labeling, adaptation level, and the central tendency shift. *Journal of Experimental Psychology*, *103*(5), 896–899. <https://doi.org/10.1037/h0037385>
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. <https://doi.org/10.1037/0033-295X.84.4.327>
- Urale, P. W. B., & Schwarzkopf, D. S. (2023). Effects of cortical distance on the ebbinghaus and delboeuf illusions. *Perception*, *52*(7), 459–483. <https://doi.org/10.1177/03010066231175014>
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Weintraub, D. J., & Schneck, M. K. (1986). Fragments of delboeuf and ebbinghaus illusions: Contour/context explorations of misjudged circle size. *Perception & Psychophysics*, *40*(3), 147–158. <https://doi.org/10.3758/BF03203010>

- Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, *68*(5), 815–828. <https://doi.org/10.3758/BF03193704>
- Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(47), 18766–18771. <https://doi.org/10.1073/pnas.0705618104>
- Yang, J., Zhang, H., & Lim, S. (2024, April). Sensory-memory interactions via modular structure explain errors in visual working memory. <https://doi.org/10.7554/eLife.95160.1>
- Yarnitsky, D., Sprecher, E., Zaslansky, R., & Hemli, J. A. (1995). Heat pain thresholds: Normative data and repeatability. *Pain*, *60*(3), 329–332. [https://doi.org/10.1016/0304-3959\(94\)00132-X](https://doi.org/10.1016/0304-3959(94)00132-X)
- Yon, D., & Frith, C. D. (2021). Precision and the bayesian brain. *Current Biology*, *31*(17), R1026–R1032. <https://doi.org/10.1016/j.cub.2021.07.044>
- Yu, K., Tuerlinckx, F., Vanpaemel, W., & Zaman, J. (2023). Humans display interindividual differences in the latent mechanisms underlying fear generalization behaviour. *Communications Psychology*, *1*(1), 5. <https://doi.org/10.1038/s44271-023-00005-0>
- Zaman, J., Struyf, D., Ceulemans, E., Beckers, T., & Vervliet, B. (2019). Probing the role of perception in fear generalization. *Scientific Reports*, *9*(1), 10026. <https://doi.org/10.1038/s41598-019-46176-x>
- Zaman, J., Ceulemans, E., Hermans, D., & Beckers, T. (2019). Direct and indirect effects of perception on generalization gradients. *Behaviour Research and Therapy*, *114*, 44–50. <https://doi.org/10.1016/j.brat.2019.01.006>
- Zaman, J., Struyf, D., Ceulemans, E., Vervliet, B., & Beckers, T. (2021). Perceptual errors are related to shifts in generalization of conditioned responding. *Psychological Research*, *85*(4), 1801–1813. <https://doi.org/10.1007/s00426-020-01345-w>

- Zaman, J., Yu, K., Andreatta, M., Wieser, M. J., & Stegmann, Y. (2023). Examining the impact of cue similarity and fear learning on perceptual tuning. *Scientific Reports*, 13(1), 13009. <https://doi.org/10.1038/s41598-023-40166-w>
- Zaman, J., Yu, K., & Lee, J. C. (2022). Individual differences in stimulus identification, rule induction, and generalization of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0001153>
- Zaman, J., Yu, K., & Verheyen, S. (2023). The idiosyncratic nature of how individuals perceive, represent, and remember their surroundings and its impact on learning-based generalization. *Journal of Experimental Psychology: General*, 152(8), 2345–2358. <https://doi.org/10.1037/xge0001403>
- Zenses, A.-K., Lee, J. C., Plaisance, V., & Zaman, J. (2021). Differences in perceptual memory determine generalization patterns. *Behaviour Research and Therapy*, 136, 103777. <https://doi.org/10.1016/j.brat.2020.103777>