Psychological Research (2021) 85:1801–1813
https://doi.org/10.1007/s00426-020-01345-w

ORIGINAL ARTICLE

Perceptual errors are related to shifts in generalization of conditioned responding

Jonas Zaman[1,2] · Dieter Struyf[2,3] · Eva Ceulemans[4] · Bram Vervliet[5] · Tom Beckers[2,6]

Received: 29 October 2019 / Accepted: 13 April 2020 / Published online: 24 April 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Studies of perceptual generalization have recently demonstrated a close relationship between stimulus perception and conditioned responding, suggesting that incorrect stimulus perception might account for certain characteristics of generalization gradients. In this study, we investigated whether common phenomena, such as the area and peak shift in conditioned responding, relate to perceptual errors. After a differential conditioning procedure, in which one circle was paired with the presentation of an aversive picture whereas a different-sized circle was not, we combined a generalization test with a three-alternative forced-choice perceptual categorization task where participants had to indicate on every trial whether the presented circle was one of the two circles from the conditioning phase or a different one, after which US-expectancy ratings were collected. The typical peak and area shift were observed when conditioned responses were plotted on a physical dimension. However, when stimulus perception was incorporated generalization gradients diverged from the typical gradient. Both the area and peak shift largely disappeared when accounting for perceptual errors. These findings demonstrate the need to incorporate perceptual mechanisms in associative models.

Introduction

The concept of generalization increasingly gained scientific attention since the observation that a trained response in humans and animals occurs in variety of new situations as a function of how similar the situations are (Ghirlanda & Enquist, 2003; Mednick & Freedman, 1960). Since the discovery of this effect by Pavlov (1927) much research has demonstrated its robustness, yet at the same time, failed to unravel its underlying mechanisms (Struyf, Zaman, Vervliet, & Van Diest, 2015). As we and others have previously argued, one reason for this is the difficulty to separate generalization from other factors such as perceptual indiscriminability that may yield the same behavior (Jäkel, Schölkopf, & Wichmann, 2008; Lashley & Wade, 1946; Struyf et al., 2015). To cite Jäkel and colleagues (2008): 'As a theoretical construct, generalization refers to a covert process that leads a subject to respond to a new stimulus in the same way as to a previously learned stimulus, despite the ability of the subject to tell the stimuli apart (p.258).' Yet, the bulk of generalization research investigates the spreading of a trained response along a physical spectrum, with small differences between test stimuli, and without ensuring the crucial prerequisite of discriminability (Struyf et al., 2015). As a consequence, it remains unclear to which extent obtained

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00426-020-01345-w) contains supplementary material, which is available to authorized users.

✉ Jonas Zaman
jonas.zaman@kuleuven.be
https://ppw.kuleuven.be/ogp

1 Health Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Box 3726, 3000 Leuven, Belgium

2 Center for the Psychology of Learning and Experimental Psychopathology, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Box 3712, 3000 Leuven, Belgium

3 Creative and Innovative Business, Thomas More, Zandpoortvest 60, Mechelen, Belgium

4 Quantitative Psychology and Individual Differences Research Unit, KU Leuven, Tiensestraat 102, Box 3731, Leuven, Belgium

5 Laboratory for Biological Psychology, KU Leuven, Tiensestraat 102, Box 3714, Leuven, Belgium

6 Leuven Brain Institute, KU Leuven, Leuven, Belgium

generalization patterns are affected by the issue of discriminability (Lashley & Wade, 1946). We recently found a close relationship between stimulus perception and the strength of generalized responding across a physical dimension (Struyf, Zaman, Hermans, & Vervliet, 2017; Zaman, Struyf, Ceulemans, Beckers, & Vervliet, 2019b; Zaman, Ceulemans, Hermans, & Beckers, 2019a). Irrespective of which test stimulus was presented, the perception of a stimulus as the trained stimulus was associated with increased generalized responding, whereas perception of the exact same stimulus as different from the trained stimulus was associated with attenuated responding (Struyf et al., 2017; Zaman et al., 2019a, b). These findings demonstrate that patterns of generalized responding could be largely a byproduct of perceptual errors, with potential far reaching theoretical consequences.

The observation that a trained response not necessarily peaks at the original stimulus has inspired a wide range of learning and cognitive models to account for these effects (Blough, 1969; Spence, 1937; Thomas & Thomas, 1974). The bulk of research on the (ir)regularities and mechanisms underlying behavioral generalization relies on conditioning protocols. First, a response (conditioned response or CR) is trained through the pairing of a stimulus (conditioned stimulus or CS) with a motivational outcome (called the unconditioned stimulus or US). Next, the extent to which test stimuli (generalization stimuli or GSs) other than the original trained stimulus trigger a conditioned response is assessed. Across the test dimension (i.e., range of stimuli that differ from the CS on one physical dimension such as frequency, size…), responding tends to peak around the CS after training with one CS (i.e., simple conditioning). However, when two stimuli (located on this dimension) are differentially paired with the US and its absence (i.e., differential conditioning)—the response peak has been found to shift away from the reinforced stimulus (CS+) in the direction opposite to the location of the non-reinforced stimulus (CS−) (Hanson, 1957, 1959; Purtle, 1973), a phenomenon known as peak shift (Fig. 1). Furthermore, gradients of conditioned responding across the test dimension tend to be symmetrical after simple conditioning (Honig & Urcuioli, 1981), whereas differential conditioning skews the gradient in the direction away from the CS−, which is termed area shift (Fig. 1).

According to associative models, the interplay between excitatory and inhibitory response distributions across the stimulus spectrum—which capture the potential of stimuli to trigger or inhibit a conditioned response—drives the area and the peak shift (Blough, 1975; McLaren & Mackintosh, 2002; Spence, 1937). Due to the difference between those distributions, with the excitatory distribution peaking at the CS+ and the inhibitory distribution at the CS− representation, their summation results in lower overall associative strength for the CS+ compared to some

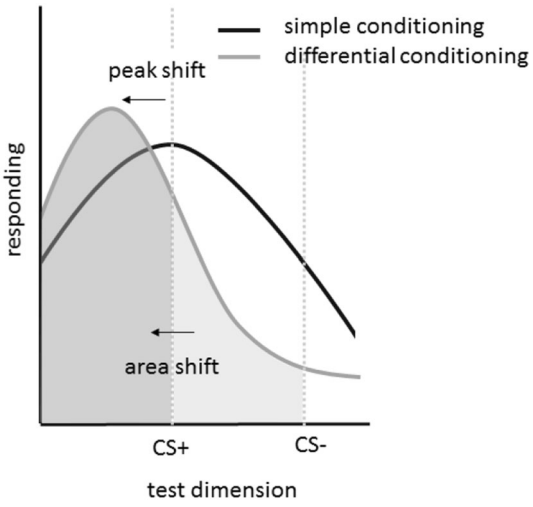

Fig. 1 Illustration of typical gradients in conditioned responding following simple conditioning (i.e., single cue paired to an outcome) and differential conditioning (i.e., one cue paired with an outcome and another cue paired with the absence of that outcome). *CS+* conditioned stimulus paired with the outcome, *CS−* conditioned stimulus paired with the absence of the outcome

other stimuli located on the perceptual spectrum close to but in the direction away from the CS− (see Fig. 1). These models make predictions regarding conditioned responding only, not regarding stimulus perception. Alternatively, adaptation theory assumes that during differential learning humans form an average stimulus representation (located between the CS+ and CS−) and that responding is based on the absolute distance between a stimulus and this average representation (Thomas, Mood, Morrison, & Wiertelak, 1991; Thomas & Thomas, 1974). Changes in the average representation—due to the presentation of different test stimuli—are used to explain the peak and area shift. Most studies in support of this theory are based on tasks where subjects have to emit a certain response to stimulus A and another response to stimulus B. During the test phase, a range of different stimuli is presented, while participants are instructed to continue pressing either of the response options. High rates of responding to the test stimuli are interpreted as generalization. Alternatively, they might merely reflect the inability of participants to perceive test stimuli as different from the trained stimuli (Struyf et al., 2015). In a large sample, after differential conditioning, participants were presented with either the CS+, the CS− or one of a number of GSs in a one-trial generalization phase (Struyf et al., 2017). We found a perceptual area shift as GSs on the side of the CS+ away from the CS− had a higher probability of being misidentified as the CS+ than GSs on the CS− side which was paralleled by higher conditioned responding. Albeit preliminary,

these findings suggest that perceptual mechanisms might drive canonical phenomena such as the area and peak shift.

Therefore, the aim of the current study was to investigate to which extent the area and peak shift relate to perceptual errors. We tested the extent of generalization across a spectrum of circles with increasing physical size, after differential conditioning with circles of two different sizes as CS+ [reinforced in 80% of the trials by the presentation of an aversive picture (US)] and CS−. During the generalization phase, participants had to indicate on every trial whether the presented circle was one of the circles from the conditioning phase or a different one [3-alternative forced-choice (AFC) task]. Next, US-expectancy ratings were collected, which have been demonstrated to be a valid measure of conditioned responding (Boddez et al., 2013), and the US was either presented or not. We expected that many GSs would be misidentified as CSs, especially for GSs located near the CSs at the most extreme values of the stimulus spectrum. Furthermore, such perceptual errors were hypothesized to enhance conditioned responding, yielding a peak shift and an area shift. Most importantly, we predicted that when those perceptual errors were accounted for, the peak and area shift within the gradient would disappear. Second, as large interindividual differences in the probability distributions of committed perceptual errors were observed that were associated with distinct generalization gradients (Zaman et al., 2019a, b). A second aim was to further explore the extent to which interindividual differences in committed perceptual errors relate to differences in gradients of conditioned responding. To this end, similar to Zaman et al. (2019a, b), participants were grouped (using statistical clustering methods) based on their differences in perceptual accuracy. We then assessed how this clustering related to differences in generalization gradients. Cluster analysis is an explorative, data-driven analytical tool that allows to group individuals based on similar data patterns (Hofmans, Ceulemans, Steinley, & Van Mechelen, 2015).

Materials and methods

Participants

A total of 200 participants took part in the study in exchange for course credits, 199 (169 females, mean age: 18.28, SD = 0.9) of which were included in the final analyses (one subject became unwell prior to the start of the experiment). This number is sufficient given benchmarking studies on cluster procedures (Steinley & Brusco, 2011). Participants were all first year bachelor students in psychology at KU Leuven and received partial course credit as compensation. The study was approved by the local university ethics committee (G-201610641). The study has been preregistered on the Open Science Framework (OSF) (https://osf.io/u38b5).

Stimuli and apparatus

Ten differently sized white outlines of circles (varying in diameter from 5.08 to 11.94 cm in steps of 0.762 cm; S1–S10) were depicted against a black background (similar to Zaman et al., 2019a, b). S3 (diameter = 6.60 cm) and S8 (diameter = 10.41 cm) randomly served as CS+ and CS− (in 111 participants S3 served as CS+, whereas in 88 participants, S8 was used as CS+). The remainder of the circles served as GSs. An aversive picture from the International Affective Picture System [IAPS; (Lang, Bradley, & Cuthbert, 2008)] was used as US. Prior to the start of the experiment, participants received written content descriptions of three aversive photos varying in their level of aversiveness (Struyf et al., 2017). They were instructed to select a picture that they anticipated to be aversive yet tolerable. Upon making a selection, the US was presented once and participants had the opportunity to change the US if they thought it was not sufficiently or exceedingly aversive. The most strongly aversive US [selected by 61.8% of the participants ($n = 123$)] was an image of a bloodied corpse, the moderately aversive US [selected by 31.7% of the participants ($n = 63$)] was a picture of a disabled child, and the mildly aversive US [selected by 6.5% of the participants ($n = 13$)] was a picture of a holstered gun. The moderate US received an average unpleasantness score of 4.50 (SD = 2.26) and the strongly aversive US was rated 6.38 (SD = 2.19) on a visual analogue scale (VAS) ranging from 0 ('not unpleasant') to 10 ('very unpleasant'); for the mild US only one rating was obtained (3).[1] The experiment was programmed in Affect4 (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010).

Procedure

The experiment was conducted in a classroom equipped with Dell desktop computers. Participants were invited in groups of maximum 20 and were separated by an empty desk. At the beginning of the experiment, participants received written instructions to work individually and in silence at their

[1] These averages are based on only 12% of the participants, as the majority failed to provide an unpleasantness rating due to a relatively short response window. In a study by Grühn and Scheibe (2008) the same USs received the following ratings in a sample of young adults: the most aversive US was rated as a 7.77 on arousal (1 = calm, 9 = excited) and as a 1.77 on valence (1 = unpleasant, 9 pleasant), the mild aversive US received an arousal rating of 5.86 and a valence rating of 3.14 and the mildly aversive US was rated as a 4.92 on arousal and as a 3.65 for valence (Grühn & Scheibe, 2008).

own pace. An experimenter was present to provide assistance and to make sure that no communication occurred among participants. During a practice phase (five trials), participants were familiarized with the experimental task. First, two differently sized stimuli (note that in this phase squares instead of circles were used, called square S and L) were presented together with the instruction that in the next trials participants had to decide whether the presented stimulus was one of the previously depicted squares or a novel square in a 3AFC perceptual categorization task (response options: square S, square L or new square). During a practice trial, one of the two squares was presented in the middle of the screen for 8000 ms. Categorization response buttons appeared simultaneously with stimulus onset at the bottom of the screen for 3000 ms after which a US-expectancy scale appeared for 5000 ms (1 = certainly no picture, 10 = certainly picture). All trials were separated by a 2000-ms intertrial interval (ITI) during which the screen remained black. The actual experiment comprised a habituation, acquisition and generalization phase. The habituation (3 CS+ and 3 CS− trials) and acquisition (8 CS+ and 8 CS− trials) phase were presented successively without break. During each habituation trial, either the CS+ or CS− was presented on screen for 8000 ms. The US-expectancy scale appeared at the bottom of the screen 3000 ms after CS onset and remained present until the CS disappeared. Acquisition trials had an identical trial structure apart from the fact that 87.5% of the CS+ trials ended in a 1500-ms presentation of the US, onset of which coincided with CS+ offset. Prior to the generalization phase, participants received on-screen instructions to identify the circle presented on the upcoming trials as either the 'small' or the 'large' circle presented during the preceding phase or as a 'different' circle. For the sake of simplicity, we will describe categorization responses in function of being perceived as a CS+, CS− or GS in the remainder of the paper, although response labels were irrespective of contingencies. Identical to practice trials, categorization response buttons appeared at the start of each generalization trial simultaneously with circle onset for 3000 ms after which US-expectancy ratings were assessed. The location of the cursor, required for responding, was reset to the middle of the screen at the onset of each trial and again at the onset of the US-expectancy scale. The generalization phase (48 trials) contained 8 CS+, 8 CS− and 32 GS trials (4 trials per GS) equally divided across over four consecutive blocks. Presentation order within blocks was randomized. When participants failed to respond within the provided time window (3 s for the categorization response, 5 s for the US-expectancy rating), the value was registered as missing (8.4% of the categorization data and 0.8% of the US expectancy data).

Interindividual differences in trait anxiety were measured using the STAI-T questionnaire (range: 20—80, mean: 45.70, SD: 8.26), which assesses an individual's tendency to appraise situations as threatening and to respond with anxiety across 20 self-report questions. The validated Dutch version was used (Van der Ploeg, Defares, & Spielberger, 2000). STAI-T scores were missing for 25 participants due to a programming glitch. Analyses on the effects of STAI-T on the categorization and US expectancy data can be found in the supplemental information (SI).

Data analysis

All conducted analyses and (computed) measures are reported. The datasets generated during and analyzed during the current study are available in the OSF repository (https://osf.io/yrxsz/).

The combined data from the habituation and acquisition phase were analyzed with a linear mixed model (also called a multilevel model) which is a type of regression model that contains fixed effects (i.e., variation that is explained by the independent variables of interest) and random effects (i.e., variation that is not explained by the independent variables of interest). Mixed (and marginal) models provide powerful and flexible approaches to analyze repeated measures data (Blackwell, de Leon, & Miller, 2006) and use, compared to ANOVA, the Satterthwaite approximation (Satterthwaite, 1946) to estimate the degrees of freedom which were rounded off to the closest integer. As for the manipulation check, we tested if US expectancies evolved differentially across trials between the CSs [as such the model included as fixed effects: Trial (continuous variable), CS (CS+ vs. CS−) and Trial × CS; random effect: subject dependent intercept].

The data of the generalization phase comprised the perceptual categorizations in addition to US expectancy. Categorization data were transformed, through the calculation of the different response probabilities (one for each response option) per stimulus per participant resulting in three probability distributions across the stimulus spectrum (see Fig. 3a). Differences in the shape of the probability distributions across the stimulus range between the different response options were tested. A second-order polynomial of stimulus was used to model a quadratic curve. The following fixed effects were included: Stimulus (continuous variable: 1–10), Stimulus^2 (to model a quadratic curve), Categorization (perceived as the CS+, the CS− or as a GS), Categorization × Stimulus and Categorization × Stimulus^2. To account for the repeated measures, a first-order autoregressive moving average repeated-measure covariance structure (ARMA11) was selected.

Next, we wanted to identify the number of relatively homogeneous subgroups within the total sample based on the grouping (or clustering) of participants with similar patterns in categorization probability distributions across the stimulus spectrum for all three response options. This

cluster analysis was done using the k-means algorithm, as implemented in MATLAB (Hofmans et al., 2015). This algorithm divides the whole group of participants based on their categorization probability distributions into k clusters. Each subject is allocated to the cluster for which its squared Euclidean distance to the cluster centroid is minimal, where the centroids are computed as the mean score profile per cluster. The maximum number of clusters was set to 10. We ran the analysis 10,000 times, each time using a different random initialization of the centroid matrix (Hofmans et al., 2015), to prevent ending in a local optimum. The run with the lowest sum of squared Euclidean distances was retained. Based on the inspection of the different cluster solutions and the automated scree test (Ceulemans & Kiers, 2006; Wilderjans, Ceulemans, & Meers, 2013), the solution with three clusters and thus three centroids was preferred.

US-expectancy data of the generalization phase were analyzed with mixed models. First, we tested whether an S-shaped US expectancy gradient emerged across the stimulus range. The model comprised Trial, Stimulus and Stimulus^2 as fixed effects (Model 1). Next, we added the following fixed effects: Categorization and its interaction with Stimulus and Stimulus^2 to explore to relationship between stimulus perception and the shape of the gradient across the stimulus spectrum (model 2). Explorative, we tested for differences in overall US expectancy gradients—irrespective of categorization—between the different clusters in a third model (fixed effects: Trial, Stimulus and Stimulus^2, Cluster (Cluster 1 vs. Cluster 2 vs. Cluster 3), Cluster × Stimulus and Cluster × Stimulus^2). In a final model (Model 4), we tested whether the effect of perception differed between the three clusters as the model was further extended with Categorization as well as its interactions with Stimulus, Stimulus^2, and Cluster. In all models, the random effect consisted of a subject-dependent intercept. All performed analyses and (computed) measures are reported. Adjusted Bonferroni correction was applied for post hoc tests. All mixed model analyses were performed in SPSS 24.

Results

Acquisition data (habituation and acquisition combined)

We found a main effect of CS [$F(1, 4028) = 74.04$, $p < 0.001$], Trial [$F(1, 4041) = 95.82$, $p < 0.001$] and a CS × Trial interaction [$F(1, 4027) = 730.93$, $p < 0.001$]. Participants learned the association between CSs and the probability of US occurrence as US expectancy ratings increased across CS+ trials [$\beta = 0.41$, 95% CI (0.38 0.44)] and decreased across CS trials [$\beta = -0.19$, 95% CI (CS− 0.22 − 0.17)] (Fig. 2). Exploratory analyses investigating the effect of CS counterbalancing can be found in the SI.

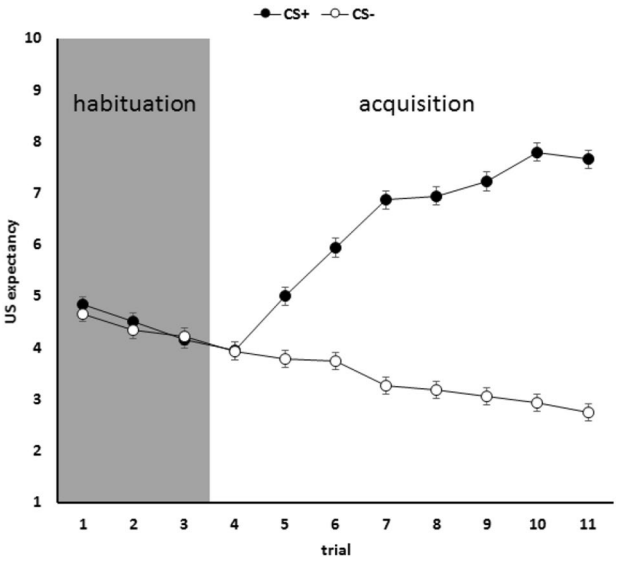

Fig. 2 US expectancy ratings across habituation and acquisition trials for the CS+ and the CS−. *CS+* conditioned stimulus paired with the outcome, *CS−* conditioned stimulus paired with the absence of the outcome. Error bars denote standard errors of the mean

Generalization data

Stimulus categorizations

On the majority of trials, generalization stimuli were incorrectly identified as either the CS+ (30.5%) or the CS− (30.5%), with only on 39.0% of the trials stimuli being correctly perceived as different from the CSs. In, on average, 74.1% of the CS+ trials and 62.8% of the CS− trials the presented stimulus was correctly identified. We found a main effect of Categorization response [$F(2, 3236) = 295.36$, $p < 0.001$] and no effects of Stimulus [$F(1, 3328) = 0.16$, $p = 0.69$] and Stimulus^2 [$F(1, 3283) = 0.56$, $p = 0.46$]. The shape of the probability distributions across the stimulus spectrum differed depending on the Categorization response [Categorization × Stimulus: $F(2, 3639) = 80.40$, $p < 0.001$; Categorization × Stimulus^2: $F(2, 3891) = 5.39$, $p = 0.005$]. Illustrated in Fig. 3a, probability distributions that stimuli were identified as CS+ or CS−, respectively, peaked around their location on the stimulus spectrum and strongly declined as stimuli became more dissimilar to the CSs. The probability that stimuli were identified as novel increased as the physical difference with the CSs increased, resulting in a wave-formed curve peaking at the most extreme and middle GSs.

Post hoc testing revealed a perceptual peak shift for the probability that a stimulus was categorized as CS+ with

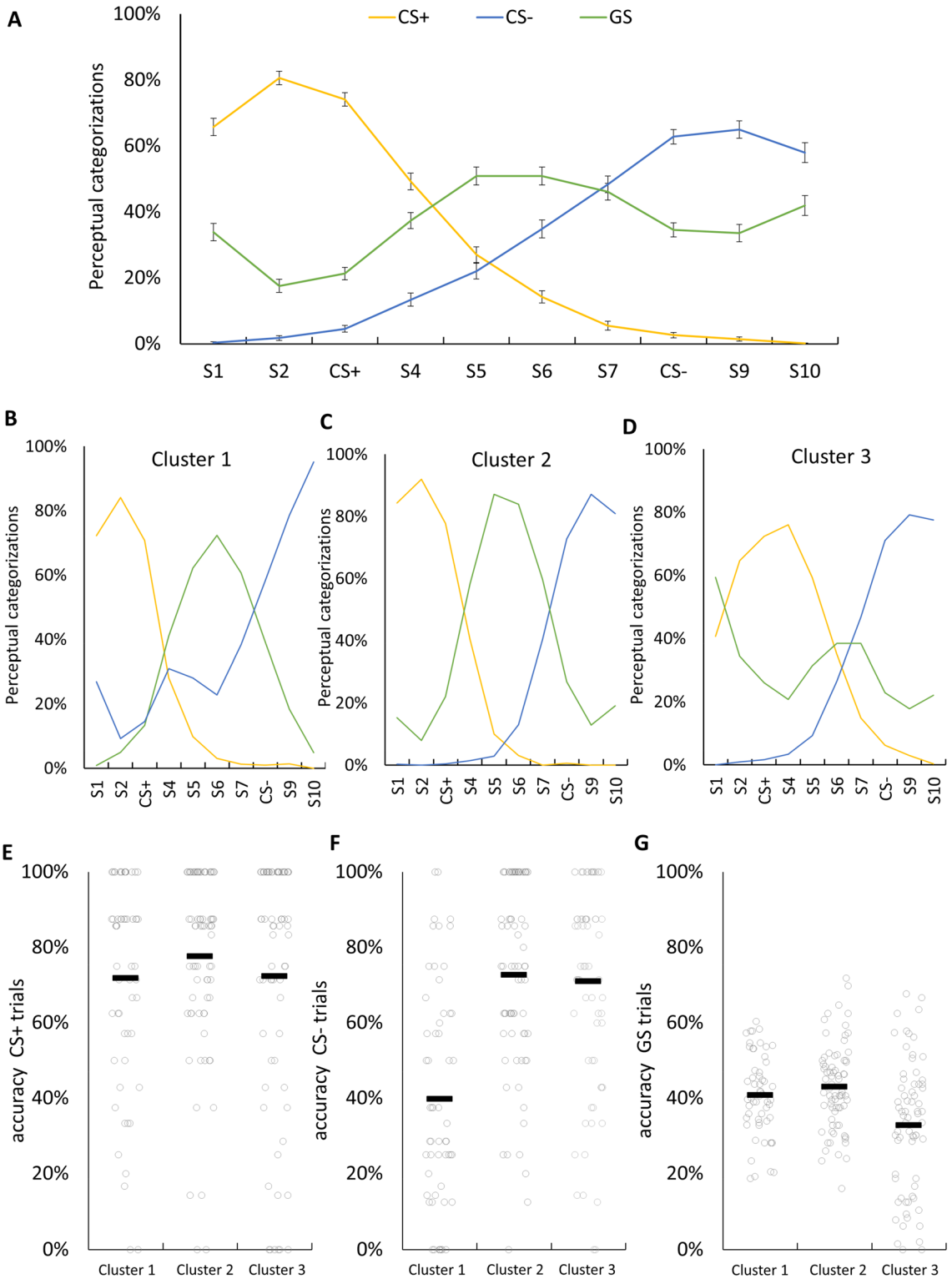

Fig. 3 **a** Probability distribution across the stimulus spectrum for the different perceptual response categories. **b**–**d** The three centroids of the identified clusters. **e**–**g** Mean perceptual accuracy for the different cluster on CS+ trials, CS− trials and GS trials. Error bars denote standard errors of the mean

the highest probability for S2 compared to S1 (S1 vs. S2: $\beta = 0.147$, $SE = 0.02$, *95% CI* [0.11 0.18], $p < 0.001$) and the CS+ (S2 vs CS+: $\beta = 0.07$, $SE = 0.02$, *95% CI* [0.03 0.10], $p < 0.001$). We did not find a negative perceptual peak shift (around the CS−) in the probability distribution of CS− responses as CS− and S9 did not differ ($\beta = -0.22$, $SE = 0.02$, *95% CI* [-0.06 0.02], $p = 0.25$). Finally, we also found evidence for a perceptual area shift around both the CS+ and the CS− as for the most extreme GSs higher probabilities of being perceived as CS+ or CS−, respectively, were found compared to their equidistant counterparts [S1 and S2 vs. S4 and S5: $t(1223) = 16.46$, $p < 0.001$; S6 and S7 vs. S9 and S10: $t(1295) = -8.09$, $p < 0.001$]. Exploratory analyses on the effect of CS allocation can be found in the SI.

Three groups of participants with similar individual probability distributions were identified based on visual inspection of the different k-means solutions and the scree plot (Fig. 3b–d). Exploratory testing in Cluster 1 ($N = 56$) revealed that categorization accuracy was significantly higher on CS+ trials compared to CS− trials [$t(55) = 7.453$, $p < 0.001$[2]] and GS trials [$t(55)$ 7.49, $p < 0.001$[2]], with no difference between GS and CS− trials [$t(55) = 0.224$, $p > 0.82$[2]]. In Cluster 2 ($N = 73$), response accuracies were overall high for the CSs and the GSs located between the CS+ and CS−. Cluster 3 ($N = 69$) was characterized by low response tendencies to perceive stimuli as novel. In both clusters 2 and 3, categorization accuracies for CS+ and CS− trials were significantly higher than for GS trials [Cluster 2: $t(72) = 10.51$, $p < 0.001$[2]; $t(72) = 8.39$, $p < 0.001$[2]; Cluster 3: $t(68) = 7.99$, $p < 0.001$[2]; $t(68) = 8.80$, $p < 0.001$[2]] with no difference between CS+ and CS− trials [Cluster 2: $t(72) = 1.17$, $p = 0.75$[2]; Cluster 3: $t(68) = 0.31$, $p < 0.001$[2]]. We found no difference in accuracy on CS+ trials between clusters (all $ps > 0.78$[2]), whereas accuracy on CS− trials was significantly lower in Cluster 1 compared to the other two clusters [Cluster 1 vs Cluster 2: $t(127) = 6.83$, $p < 0.001$[2]; cluster 1 vs cluster 3: $t(123) = 6.04$, $p < 0.001$[2]], with no difference between Cluster 2 and 3 [$t(140) = 0.38$, $p > 0.99$[2]]. Finally, accuracy on GS trials was significantly lower in Cluster 3 compared to Cluster 2 [$t(115.96) = 4.09$, $p < 0.001$[2]] and 1 [$t(115.49) = 3.14$, $p = 0.006$[2]], with again no difference between the latter two clusters [$t(127) = 1.09$, $p = 0.78$[2]][3] (Fig. 3e–g).

[2] Corrected with a factor 3 for multiple testing.

[3] Exclusion of subjects (N = 11) that seem to have switched the meaning of the response buttons [accuracies both for CS+ and CS- trials < 40%, mean accuracy: CS+ = 15.62% (SE = 4%), CS- = 13.42% (SE = 4%)] did not meaningfully change the results.

US expectancy

First, we investigated the shape of the generalization gradient across the stimulus spectrum (model 1). As expected, we found a S-shaped gradient as indicated by the main effect of Stimulus [$F(1, 10{,}066) = 367.76$, $p < 0.001$] and Stimulus^2 [$F(1, 10{,}066) = 27.52$, $p < 0.001$]. Across the stimulus spectrum, US expectancy increased when stimuli approached the CS+ and decreased as they approached the CS−. We found a main effect of Trial [$F(1, 10{,}065) = 18.18$, $p < 0.001$], as US expectancy increased across generalization trials [$\beta = 0.02$, 95% CI (0.01 0.03)]. Analyses on the effect of CS allocation can be found in the SI.

Next, the influence of perception upon generalized responding was investigated (model 2). The categorization of the stimulus strongly affected US expectancy (main effect of Categorization: $F(2, 9351) = 42.92$, $p < 0.001$]. Overall, CS+ percepts led to significantly higher US expectancy [$\beta = 4.06$, 95% CI (3.95 4.18)] and CS− percepts to significantly lower US expectancy ratings [$\beta = -0.65$, 95% CI (− 0.77 − 0.54)] compared to when stimuli were perceived as novel. Furthermore, depending on how a stimulus was categorized, different gradients emerged across the stimulus spectrum [Categorization × Stimulus: $F(2, 9335) = 44.03$, $p < 0.001$; Categorization × Stimulus^2: $F(2, 9342) = 54.05$, $p < 0.001$]. In Fig. 4a, a much broader gradient is found for the CS+ percepts, whereas for the CS− percepts a steep gradient is observed across the stimulus spectrum. For the full model output see SI.

Exploratory, differences in overall gradients—irrespective of categorization—between the different clusters was tested (model 3). The main effect of Cluster [$F(2, 1445) = 55.14$, $p = 0.001$] was significance. Further, different gradients between the Clusters were observed [Cluster × Stimulus: $F(2, 10{,}011) = 99.85$, $p = 0.001$; Cluster × Stimulus^2: $F(2, 10{,}011) = 89.32$, $p = 0.001$] (see Fig. 4b–d). For the full model output see SI.

Last, we tested whether the effect of a perceptual error was dependent on cluster membership (model 4). We found a significant interaction between Cluster and Categorization [$F(4, 9331) = 7.67$, $p < 0.001$]. Post hoc testing revealed: (1) higher US expectancy when a stimulus was categorized as GS in Cluster 3 compared to Cluster 1 [mean difference = 0.76, SE = 0.20, 95% CI (0.37 1.16), $p < 0.001$[4]] and Cluster 2 [mean difference = 0.83, SE = 0.18, 95% CI (0.46 1.21), $p < 0.001$[5]], with no difference between the latter [mean difference = -0.07, SE = 0.20, 95% CI (− 0.32 0.45), $p > 0.99$[5]]. (2) When a stimulus was perceived as the CS+ in cluster 2, US expectancy ratings were higher compared to the remaining clusters [Cluster 1, mean difference = 0.79,

[4] Corrected with a factor of 3 for multiple testing.

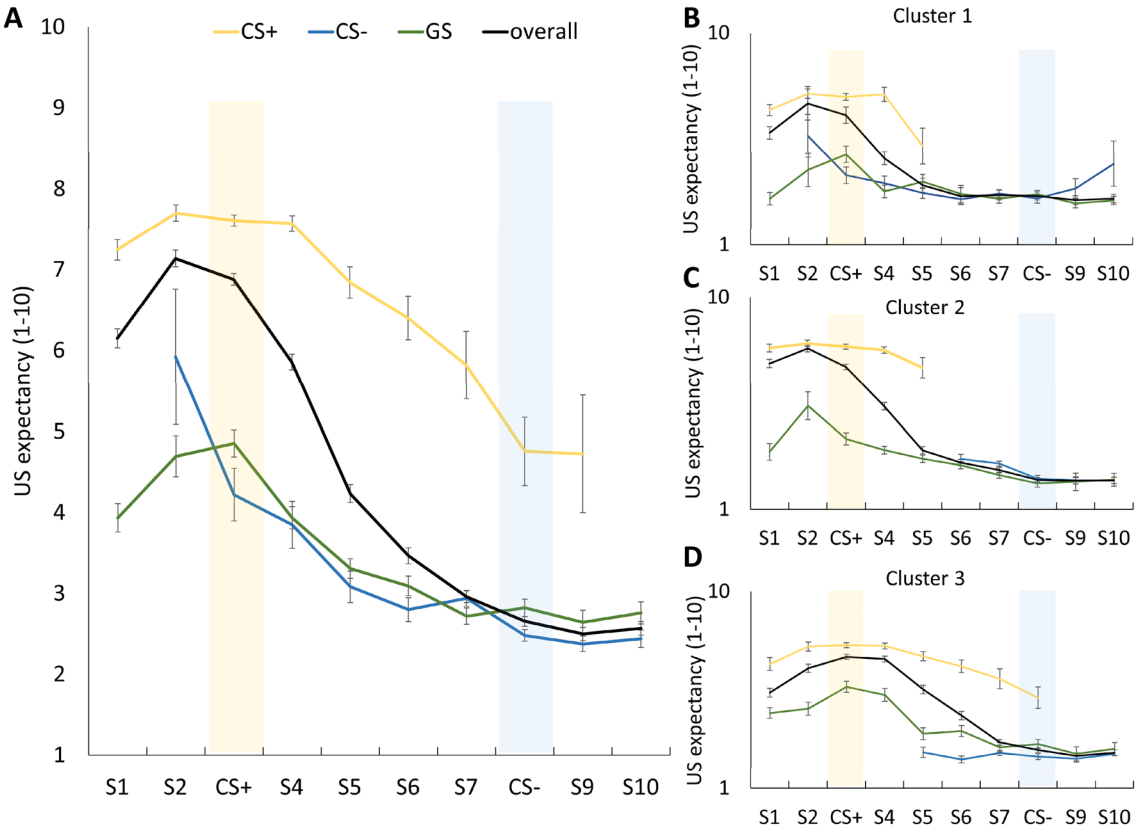

Fig. 4 **a** Mean US expectancy ratings across the stimulus spectrum regardless of stimulus perception (averaged) or depending on whether the stimulus was perceived as the CS+, CS− or GS. **b–d** US expectancy rating for the different Clusters with and without accounting for perception. Error bars denote standard errors of the mean

SE = 0.20, 95% CI (0.39 1.19), $p < 0.001$[5]; Cluster 2: mean difference = 0.59, SE = 0.19, 95% CI (0.22 0.96), $p = 0.005$[5]], with no differences between the latter two [mean difference = − 0.20, SE = 0.20, 95% CI (− 0.60 0.20), $p = 0.96$[5]]. (3) The categorization of a stimulus as CS− led to higher US expectancy in Cluster 1 compared to Cluster 2 and 3 [Cluster 2, mean difference = 0.95, SE = 0.20, 95% CI (0.56 1.35), $p < 0.001$[5]; Cluster 3: mean difference = 0.68, SE = 0.20, 95% CI (0.29 1.08), $p = 0.003$[5]], with no differences between the latter two [mean difference = − 0.27, SE = 0.19, 95% CI (− 0.64 0.11), $p = 0.57$[5]]. Finally, the effect of cluster or its interaction with stimulus was no longer significant, suggesting that differences in generalization gradients between clusters decreased when perceptual errors were taken into account [Cluster: $F(2, 2079) = 2.39$, $p = 0.09$; Cluster × Stimulus: $F(2, 9229) = 2.82$, $p = 0.059$; Cluster × Stimulus[2]: $F(2, 9227) = 3.34$, $p = 0.036$]. For the full model output, see SI.

Averaged vs corrected gradient (peak and area shift) Additional analyses were conducted that focused on the peak shift and area shift and whether they remained present when removing responses preceded by perceptual errors. The gradient obtained by averaging all responses per stimulus (as typically done) is labeled the averaged gradient. The gradient that is based only on trials where the stimulus was categorized correctly is called the corrected gradient. Here, only responses are averaged for trials where the presented and perceived stimulus mapped (e.g., trials where a GS was perceived as a GS).

We found in the averaged gradient a positive area shift as US expectancy ratings were higher for S1 and S2 compared to S4 and S5 [(S1 + S2)/ 2 vs (S4 + S5)/2: $t(10{,}065) = 19.47$, $p < 0.001$[6]] as well as a negative area shift as US expectancy ratings were lower for S9 and S10 compared to S6 and S7 [(S7 + S8)/ 2 vs (S9 + S10)/2: $t(10{,}064) = 8.66$, $p < 0.001$[6]]. In the corrected gradient, a positive area shift was observed albeit to a much lesser extent [$t(4437) = 4.14$, $p < 0.001$,[5] whereas the negative area shift no longer was significant [$t(4447) = 2.31$, $p = 0.08$[6]] (Fig. 5a). Furthermore, a positive peak shift occurred in the averaged gradient as US expectancy was higher to S2 compared to the CS+ [$t(10{,}059) = 2.72$, $p = 0.019$[6]] as well as to S1 [$t(10{,}059) = 7.25$, $p < 0.001$[6]]. No negative peak shift was found as there was no

[5] Corrected with a factor of 4 for multiple testing.

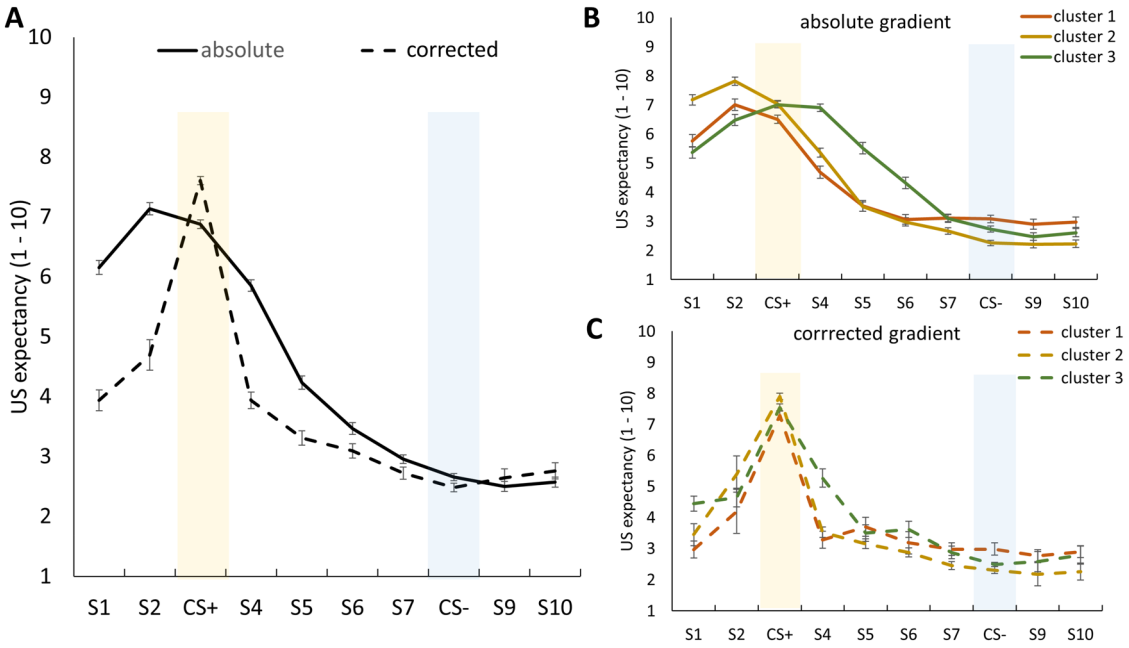

Fig. 5 **a** Mean US expectancy gradient based on all responses (averaged gradient) and on trials with accurate categorization performance only (corrected gradient). **b** Averaged generalization gradients for the different clusters. **c** Corrected generalization gradients for the different clusters. Error bars denote standard errors of the mean

Table 1 Mean US expectancy per cluster for the absolute and corrected gradient

	Averaged gradient			Corrected gradient		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
S1	5.79^{a} $(0.21)^{a}$	7.21 $(0.17)^{b}$	5.46 $(0.19)^{a}$	2.96 $(0.27)^{+}$	3.44 $(0.35)^{+}$	4.44 $(0.25)^{-}$
S2	7.04 $(0.19)^{a}$	7.83 $(0.14)^{b}$	6.41 $(0.18)^{a}$	4.19 $(0.71)^{+}$	5.39 $(0.59)^{+}$	4.63 $(0.30)^{+}$
CS+	6.39 $(0.14)^{a}$	6.98 $(0.12)^{a}$	6.99 $(0.11)^{a}$	7.29 $(0.15)^{+}$	7.90 $(0.10)^{+}$	7.54 $(0.11)^{+}$
S4	4.66 $(0.19)^{a}$	5.37 $(0.15)^{b}$	6.82 $(0.13)^{c}$	3.27 $(0.27)^{+}$	3.51 $(0.17)^{+}$	5.27 $(0.30)^{-}$
S5	3.60 $(0.18)^{a}$	3.62 $(0.15)^{a}$	5.48 $(0.19)^{b}$	3.69 $(0.30)^{+}$	3.15 $(0.15)^{+}$	3.49 $(0.27)^{+}$
S6	3.13 $(0.16)^{a}$	3.05 $(0.14)^{a}$	4.37 $(0.18)^{b}$	3.17 $(0.36)^{+}$	2.87 $(0.15)^{+}$	3.61 $(0.26)^{+}$
S7	3.11 $(0.12)^{a}$	2.72 $(0.11)^{a}$	3.21 $(0.13)^{a}$	2.97 $(0.21)^{+}$	2.45 $(0.13)^{+}$	2.87 $(0.21)^{+}$
CS−	3.07 $(0.12)^{a}$	2.31 $(0.09)^{b}$	2.80 $(0.10)^{a,b}$	2.98 $(0.20)^{+}$	2.30 $(0.11)^{-}$	2.45 $(0.11)^{+-}$
S9	2.92 $(0.16)^{a}$	2.28 $(0.13)^{a}$	2.51 $(0.13)^{a}$	2.76 $(0.19)^{+}$	2.17 $(0.38)^{+}$	2.58 $(0.32)^{+}$
S10	2.94 $(0.17)^{a}$	2.26 $(0.13)^{a}$	2.74 $(0.14)^{a}$	2.89 $(0.18)^{+}$	2.26 $(0.28)^{+}$	2.79 $(0.30)^{+}$

Standard errors in brackets. Different superscripts denote differences between clusters at $p < 0.005$

difference in US expectancy between the CS− and S9 [$t(10{,}059) = 1.47$, $p = 0.57$[6]]. In the corrected gradient, no peak shift occurred as the highest US expectancy was observed for the CS+ compared to S2 [$t(4430) = -15.56$, $p < 0.001$[6]]. No difference between CS− and S9 was observed [$t(4435) = 1.15$, $p > 0.9$] (Fig. 5a). In Table 1, differences between clusters in US expectancy per stimulus are reported once for the averaged gradient and once for the corrected gradient (Fig. 5b, c). After filtering out the effect of perceptual mistakes the gradients between clusters became markedly more similar.

Discussion

The current study investigated the relationship between perceptual errors and generalization phenomena like the area and peak shift. To this end, a generalization protocol was combined with a perceptual categorization task. Our findings demonstrated that many generalization stimuli were misidentified as the stimuli used during training, with a strong correspondence between stimulus perception and conditioned responding. The area and peak shift

were found both for the perceptual and US expectancy gradient. Correcting for perceptual errors resulted in the disappearance of the peak shift and a strong reduction of the area shift in the US expectancy gradient.

The standard approach to analyzing generalization gradients is by averaging the responses to the same stimulus and plotting them across the spectrum of presented stimuli, yielding a gradient in conditioned responding. Underlying this approach is the premise that humans are ideal observers, aware of physical differences, with perception of a given stimulus constant across trials. Our findings together with other recent studies (Holt et al., 2014; Laufer & Paz, 2012; Lovibond, Lee, & Hayes, 2019; Schechtman, Laufer, & Paz, 2010; Struyf et al., 2017; Zaman et al., 2019a, b) question this assumption as we demonstrate that many perceptual errors may occur within a context of perceptual generalization, with average accuracies for GSs in the present study around chance level. Furthermore, conditioned responding to the same stimulus in our study was higher when it was perceived as the CS+ and lower when it was identified as different from the CS+, regardless of whether the stimulus was actually a CS or a GS (Struyf et al., 2017; Zaman et al., 2019a, b). This illustrates the strong relationship between perceptual errors and obtained gradients and challenges the convention of simply averaging responses over a physical stimulus identity, independent of subjective perception.

Large interindividual variations in perceptual errors on CS+, CS− and GS trials were observed. Using a clustering approach, we identified three groups of individuals with different perceptual probability distributions across the stimulus spectrum that related to distinct gradients in conditioned responding. In addition, a cluster-dependent effect of perception was observed as US expectancy ratings after CS− and GS percepts were higher in those clusters where the perceived CS–US and GS–US contingencies were relatively higher (see SI for supporting analyses). When controlling for differences in committed perceptual errors, generalization gradients between clusters became markedly more similar. These findings demonstrate that, for example, wider generalization gradients do not necessarily imply cognitive or learning problems (Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015) but could be attributed to differences in perceptual performance (Laufer, Israeli, & Paz, 2016) and subsequent experienced stimulus-outcomes probabilities. Hence, inferences regarding latent mechanisms between clinical populations and healthy individuals should be made cautiously when made irrespective of perception as differences in perceptual abilities might result in distinct generalization gradients (Laufer et al., 2016).

Unlike traditional behaviorist and associative accounts (Blough, 1975; McLaren & Mackintosh, 2002; Spence, 1937) that typically do not explicitly distinguish between perceptual and cognitive mechanisms, more recent work has attempted to investigate the distinct contribution of cognitive and perceptual processes to generalization. This recent work approaches generalization from a different perspective, in that it attempts to determine what remains of generalization gradients if perceptual mechanisms are accounted (Holt et al., 2014; Lovibond et al., 2019; Struyf et al., 2017; Zaman et al., 2019a, b). However, it currently remained unclear to what extent the peak and area shift remain preserved when responses were plotted across stimuli that were identified as different from the initial learned stimuli. When omitting perceptual errors in our study, a gradient emerged that was very different from the typical gradient. Responses peaked at the CS+ and strongly decreased to the neighboring stimuli, after which responses more gradually decreased as the physical difference increased. Interestingly, when GSs that were mistaken as CS+ were omitted from the data the peak shift disappeared and the extent of the area shift dropped drastically. These findings suggest an alternative account where generalization phenomena such as the peak and area shift in fact strongly relate to perceptual errors. However, they also require further research before a firm conclusion can be drawn. In the current design, both outcomes were measured congruently. Hence, the observed relationship may be a consequence of the adopted design. Put differently, participants might adjust their expectancy judgments so they are in accordance with the response on the 3AFC task. Others have assessed US expectancy and perception in separate blocks (Lovibond et al., 2019) or used discriminable test stimuli (based on individuals psychometric functions) (Holt et al., 2014), circumventing this issue but having other limitations (i.e., accounting for trial-by-trial perceptual variability; Struyf et al., 2015). In previous work, however, a similar relationship between perception and conditioned responding was found despite US expectancy ratings being assessed before perception (without potential carry over-effects of previous trials due to its one-test-trial design) (Struyf et al., 2017). Furthermore, relatively similar gradients have been found with or without the perceptual categorization task, suggesting a limited impact of including a categorization task (Zaman et al., 2019b). In contrast, Lovibond and colleagues (2019) assessed expectancy and perceptual categorization in separate blocks. They found different US expectancy gradients depending on when categorization was tested. Yet, as the authors did not control for the effect of mere extra stimulus exposure, the extent of a potential carry-over effect remains unclear. Interestingly, in their study, a peak shift in US expectancy ratings was only observed when additional test stimuli were added that had high probabilities of being incorrectly categorized as the CS+, because of their close resemblance to the CS+. Combined these findings are suggestive of the importance of stimulus indiscriminability to observe a peak shift.

The current data cannot exclude the possibility that both perception and conditioned responding are affected by the same latent processes. For example, the shape of our perceptual gradients remarkably matches the gradient predicted by associative learning accounts (Blough, 1975; McLaren & Mackintosh, 2002; Spence, 1937). Yet, the generalizability of these associative models to perceptual categorizations remains to be tested as they were developed to account for gradients in conditioned responding. Similarly, it remains unclear to which extent exemplar models of perceptual categorization like the general context model (Nosofsky, 1986) or the ALCOVE model (Kruschke, 1992) can account for a perceptual peak shift or generalization gradients of other response outcomes, like expectancy ratings. A shift in perception has been found previously in perceptual categorization research (Davis & Love, 2010). According to prototype categorization models, shifts in the memory representation of learned categories can give rise to a perceptual peak shift. During differential learning, two categories are learned (safe stimulus vs. threatening stimulus) based on perceptual features that are located on the same physical dimension (Davis & Love, 2010). Due to contrastive learning mechanisms, category averages (i.e., prototypes) are not determined solely by the statistical properties of the perceived stimuli but can shift depending on situational demands in the direction opposite to the contrasted category (Davis & Love, 2010). Future research should investigate to which extent perceptual gradients and gradients in conditioned responding are driven by identical or different mechanisms.

At the same time, our data clearly demonstrate that perceptual generalization cannot be reduced to problems with stimulus perception. A gradient in conditioned responding across the stimulus dimension remained while accounting for perception, suggesting the involvement of cognitive, non-perceptual processes. In line, recent work has demonstrated the influence of cognitive rules on the shape of the generalization gradient across a perceptual dimension (Lee, Hayes, & Lovibond, 2018; Lee, Lovibond, Hayes, & Navarro, 2018a, b). The current paper's focus was on perceptual generalization as it is the most studied form, yet many other forms of generalization exist, where perception is of subordinate importance (Dunsmoor & Murphy, 2015; Feather, 1965). For instance, a trained fear response has been found to generalize across a range of distinct objects belonging to the same conceptual category (Dunsmoor, White, & LaBar, 2011).

Strong similarities exist between the perceptual categorization task in our study and those used in the context of adaptation theory (Thomas et al., 1991; Thomas & Thomas, 1974), apart from the included third response option in the current experiment. Nonetheless, it would be interesting to compare the extent to which the obtained perceptual gradient matches with predictions derived from adaptation theory. The symmetrical distribution of presented GSs around the CSs should not affect the location of the average representation and hence no peak or area shift should occur according to adaptation theory. Yet, the obtained perceptual area and peak shift seem to contradict this. A recent study further challenged this theory as a peak shift was found after a one-trial generalization paradigm, which again would not be expected departing from the principles of adaptation theory (Struyf, Iberico, & Vervliet, 2014).

Interestingly, compared to a previous study that used identical generalization stimuli but relied on a simple conditioning procedure (Zaman et al., 2019a, b), perceptual accuracies on CS trials were much higher in the current study. Combined with the finding that CS percepts elevate conditioned responding, this provides an alternative explanation for the typical observation that conditioned responding to the CS+ is higher in a generalization task when preceded by differential compared to simple conditioning (Hanson, 1959, 1961). Our findings suggest an alternative explanation where increases in perceptual accuracy on CS trials drive the increase in averaged conditioned responding in differential relative to non-differential procedures. Which methodological difference determines perceptual performance (or changes therein) remains unclear. Preliminary studies suggest that variations in the physical difference between the CS+ an CS− moderate perceptual performance (Åhs, Miller, Gordon, & Lundström, 2013; Aizenberg & Geffen, 2013; Li, Howard, Parrish, & Gottfried, 2008). However, apart from these studies, systematic research is lacking.

In conclusion, the custom approach to average conditioned responding regardless of perception led to the inclusion of many cases where stimuli were mistaken for the trained stimuli. Inferences on these generalization gradients may lead to an overestimation of individual's generalization tendencies. We demonstrated that the congruent assessment of stimulus perception and conditioned responding is a promising approach to gain better insight into the mechanism driving the generalization of behavior.

Author contributions JZ developed the paradigm. JZ and DS collected the data. JZ analyzed the data. JZ wrote the manuscript. JZ, EC, DS, BV and TB revised the manuscript.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. JZ is a postdoctoral Research Fellow of the Research Foundation Flanders (FWO, 12P8619N), and was supported by the "Asthenes'' long-term structural funding (METH/15/011)—Methusalem grant by the Flemish Government, and a Krediet aan Navorsers (FWO, 1500620 N).

Data availability Data are available on the open science framework.

Compliance with ethical standards

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

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.

Aizenberg, M., & Geffen, M. N. (2013). Bidirectional effects of aversive learning on perceptual acuity are mediated by the sensory cortex. *Nature Neuroscience, 16*(8), 994–996. https://doi.org/10.1038/nn.3443.

Blackwell, E., de Leon, C. F. M., & Miller, G. E. (2006). Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. *Psychosomatic Medicine, 68*(6), 870–878. https://doi.org/10.1097/01.psy.0000239144.91689.ca.

Blough, D. (1969). Generalization gradient shape and summation in steady-state tests. *Journal of the Experimental Analysis of Behavior, 12*(1), 91–104. https://doi.org/10.1901/jeab.1969.12-91.

Blough, D. S. (1975). Steady sate data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 1*(1), 3–21.

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 201–206. https://doi.org/10.1016/j.jbtep.2012.08.003.

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology, 59*(1), 133–150. https://doi.org/10.1348/000711005X64817.

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.

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.

Dunsmoor, J. E., White, A. J., & LaBar, K. S. (2011). Conceptual similarity promotes generalization of higher order fear learning. *Learning & Memory (Cold Spring Harbor, NY), 18*(3), 156–160. https://doi.org/10.1101/lm.2016411.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: systematic review and implications for anxiety disorder research. *Behavior Therapy, 46*(5), 561–582. https://doi.org/10.1016/j.beth.2014.10.001.

Feather, B. W. (1965). Semantic generalization of classically conditioned responses: A review. *Psychological Bulletin, 63*(6), 425–441. https://doi.org/10.1037/h0022003.

Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour, 66*(1), 15–36. https://doi.org/10.1006/anbe.2003.2174.

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods, 40*(2), 512–521. https://doi.org/10.3758/BRM.40.2.512.

Hanson, H. M. (1957). Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science, 125*(3253), 888–889. https://doi.org/10.1126/science.125.3253.888.

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.

Hanson, H. M. (1961). Stimulus generalization following three-stimulus discrimination training. *Journal of Comparative and Physiological Psychology, 54*(2), 181–185. https://doi.org/10.1037/h0041866.

Hofmans, J., Ceulemans, E., Steinley, D., & Van Mechelen, I. (2015). On the added value of bootstrap analysis for K-means clustering. *Journal of Classification, 32*(2), 268–284. https://doi.org/10.1007/s00357-015-9178-y.

Holt, D. J., Boeke, E. A., Wolthusen, R. P. F., Nasr, S., Milad, M. R., & Tootell, R. B. H. H. (2014). A parametric study of fear generalization to faces and non-face objects: Relationship to discrimination thresholds. *Frontiers in Human Neuroscience, 8*(SEP), 1–12. https://doi.org/10.3389/fnhum.2014.00624.

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.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin and Review, 15*(2), 256–271. https://doi.org/10.3758/PBR.15.2.256.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8.*

Lashley, K. S., & Wade, M. (1946). The Pavlovian theory of generalization. *Psychological Review, 53*(2), 72–87. https://doi.org/10.1037/h0059999.

Laufer, O., Israeli, D., & Paz, R. (2016). Behavioral and neural mechanisms of overgeneralization in anxiety. *Current Biology, 26*(6), 713–722. https://doi.org/10.1016/j.cub.2016.01.023.

Laufer, O., & Paz, R. (2012). Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 32*(18), 6304–6311. https://doi.org/10.1523/JNEUROSCI.6281-11.2012.

Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2018a). Peak Shift and Rules in Human Generalization. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000558.

Lee, J. C., Lovibond, P. F., Hayes, B. K., & Navarro, D. J. (2018b). Negative evidence and inductive reasoning in generalization of associative learning the role of inductive reasoning in generalization of human associative learning view project. *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.

Lovibond, P. F., Lee, J. C., & Hayes, B. K. (2019). Stimulus discriminability and induction as independent components of generalization. *Journal of Experimental Psychology Learning, Memory, and Cognition (September).*. https://doi.org/10.1037/xlm0000779.

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.

Mednick, S. A., & Freedman, J. L. (1960). Stimulus generalization. *Psychological Bulletin, 57*(3), 169–200. https://doi.org/10.1037/h0041650.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford University Press.

Purtle, R. B. (1973). Peak shift: A review. *Psychological Bulletin, 80*(5), 408–421. https://doi.org/10.1037/h0035233.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics, 2*(6), 110–114. https://doi.org/10.2307/3002019.

Schechtman, E., Laufer, O., & Paz, R. (2010). Negative valence widens generalization of learning. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 30*(31), 10460–10464. https://doi.org/10.1523/JNEUROSCI.2377-10.2010.

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimenssion. *Psychological Review, 44*(5), 430–444. https://doi.org/10.1037/h0062885.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: a free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*(1), 36–45. https://doi.org/10.1027/1618-3169/a000005.

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods, 16*(1), 63–79. https://doi.org/10.1037/a0022673.

Struyf, D., Iberico, C., & Vervliet, B. (2014). Increasing predictive estimations without further learning: the peak-shift effect. *Experimental Psychology, 61*(2), 134–141. https://doi.org/10.1027/1618-3169/a000233.

Struyf, D., Zaman, J., Hermans, D., & Vervliet, B. (2017). Gradients of fear: How perception influences fear generalization. *Behaviour Research and Therapy, 93*, 116–122. https://doi.org/10.1016/j.brat.2017.04.001.

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

Thomas, D. R., Mood, K., Morrison, S., & Wiertelak, E. (1991). Peak shift revisited: A test of alternative interpretations. *Journal of Experimental Psychology: Animal Behavior Processes, 17*(2), 130–140. https://doi.org/10.1037/0097-7403.17.2.130.

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.

Van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (2000). Handleiding bij de zelf-beoordelings vragenlijst. Een Nederlandstalige bewerking van de Spielberger State-Trait Anxiety Inventory [Manual for the State-Trait Anxiety Inventory. A Dutch translation]. Lisse: The Netherlands

Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull-based model selection method. *Behavior Research Methods, 45*(1), 1–15. https://doi.org/10.3758/s13428-012-0238-5.

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

Zaman, J., Struyf, D., Ceulemans, E., Beckers, T., & Vervliet, B. (2019b). Probing the role of perception in fear generalization. *Scientific Reports, 9*(1), 10026. https://doi.org/10.1038/s41598-019-46176-x.

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