Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Direct and indirect effects of perception on generalization gradients

Jonas Zaman[a,b,*], Eva Ceulemans[c], Dirk Hermans[b], Tom Beckers[b]

[a] Health Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Box 3726, 3000, Leuven, Belgium
[b] Center for the Psychology of Learning and Experimental Psychopathology, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Box 3712, 3000, Leuven, Belgium
[c] Quantitative Psychology and Individual Differences Research Unit, KU Leuven, Tiensestraat 102, Box 3731, 3000, Leuven, Belgium

ARTICLE INFO

ABSTRACT

For more than a century, researchers have attempted to understand why organisms behave similarly across situations. Despite the robust character of generalization, considerable variation in conditioned responding both between and within humans remains a challenge for contemporary generalization models. The current study aims to investigate the extent to which variation in behavior in a context of generalization can be attributed to differences in perception. We combined a fear conditioning and generalization procedure with a perceptual decision task in humans. We found that the failure to perceive a novel stimulus as different from the trained fear-evoking stimulus led to increased conditioned responding. Furthermore, perceptual errors yielded perceived stimulus-outcome contingencies that differed substantially from the objective contingencies. Final, the impact of a perceptual error was dependent upon these perceived contingencies. These findings suggest that generalization across a perceptual dimension is to a large extent driven by perceptual errors that directly affect behavior but also indirectly as they yield different learning experiences between individuals.

1. Introduction

Humans (and animals) often demonstrate remarkably similar behavior across a variety of situations (Ghirlanda & Enquist, 2003; Mednick & Freedman, 1960), a phenomenon called generalization. Generalization is often investigated using a conditioning protocol, where an initially *neutral* stimulus (conditioned stimulus, CS) starts to elicit a response (conditioned response, CR), such as physiological arousal, after it has been linked to a motivationally significant stimulus (unconditioned stimulus, US, e.g., food, pain). After the establishment of a CS-CR relationship, the extent to which novel test stimuli (generalization stimuli, GS) elicit the CR is tested. Typically, the strength or probability of a CR will decrease as a function of physical dissimilarity between the GS and the CS (Ghirlanda & Enquist, 2003; Mednick & Freedman, 1960), resulting in a bell-shaped gradient peaking around the CS. This generalization behavior seems to rely on the degree of applicability of prior knowledge to the current situation, which in this context seems to be a function of *physical* similarity. Whether this behavior is reflective of an active decision process (i.e., generalization occurs in spite of the fact that the subject can tell the difference between two stimuli) (Dunsmoor & Paz, 2015; Shepard, 1987), is a function of stimulus similarity (due to the overlap in distribution of activation across *units)* (McLaren & Mackintosh, 2002) or reflects merely a failure to

perceptually discriminate between stimuli (Lashley & Wade, 1946; Mostofsky, 1965; Prokasy, Hall, & Ball, 1963; Struyf, Zaman, Vervliet, & Van Diest, 2015) is currently still debated (Onat & Büchel, 2015).

According to the *decisional account*, wider generalization gradients can be attributed to a better-safe-than-sorry strategy, where participants respond with a strong CR despite being aware that a GS is similar but not identical to the CS (Dunsmoor & Paz, 2015; Shepard, 1987). This dominant notion, that generalization reflects a post-perceptual decision process, has been challenged in the past (Lashley & Wade, 1946; Mostofsky, 1965; Prokasy et al., 1963) and more recently (Struyf et al., 2015). As an alternative, it has been proposed that generalization occurs to the extent that perceptual discrimination fails [i.e., the inverse hypothesis (Honig & Urcuioli, 1981; Lashley & Wade, 1946; Mostofsky, 1965; Prokasy et al., 1963)] with the bell-shaped generalization gradient merely reflecting a probability gradient of perceptual errors (Struyf et al., 2015). Following this *perceptual account*, the more a GS resembles the CS, the more likely it is to be incorrectly identified as the CS and trigger a CR. The work of Guttman and Kalish (1956), in which similar gradients were found across a light wave spectrum, despite the fact that identical physical changes in hue were assumed to result in different levels of perceived change in color (Guttman & Kalish, 1956), has been used to contradict the perceptual account. However, follow-up work has demonstrated that Guttman and

---

Kalish selected stimuli based on incorrect psychometric functions and that the obtained gradients in fact did relate to variations in perceptual sensitivity (Blough, 1972; Honig & Urcuioli, 1981). At present, it therefore remains unclear whether variation in generalized responding between and within individuals can be attributed to differences in stimulus perception, given its stochastic nature (Petzschner, Glasauer, & Stephan, 2015).

We recently developed a generalization protocol that simultaneously evaluates generalization behavior (CR) and stimulus perception, in an attempt to identify the mechanisms underlying generalization and to account for differences between individuals (Struyf et al., 2015). In a one-trial generalization procedure tested in a large sample, we found that in a considerable portion of GS trials, GSs are indeed mistaken for the CS, and these perceptual errors determined conditioned responding to considerable extent (Struyf, Zaman, Hermans, & Vervliet, 2017). Moreover, in a differential fear conditioning procedure, individual variability in perceptual errors during CS+ trials led to different learning trajectories between individuals (Zaman et al., 2017), suggesting that variations in stimulus perception account for inter-individual differences in learning. In this experiment, during a perceptual categorization task participants categorized the CS+ and CS- based on their perceptual features. The authors found that many errors on CS+ trials during the acquisition phase (categorization of the CS+ as the CS-) led to increased conditioned responding to both the CS+ and the CS-, whereas participants making few errors only showed increased conditioned responding to the CS+. These findings were related to differences in perceived CS-US contingencies between the groups. For participants that made many errors on CS+ trials, both the CS+ and the CS- were perceived as predictive of the US (given that CS+ trials were reinforced regardless of whether the stimulus was misidentified as CS- or not). Despite those preliminary findings, it remains unclear to what extent generalization gradients are indeed driven by perceptual misidentification, and more importantly, whether inter-individual differences in generalization can be attributed in whole or in part to differences in perceptual accuracy.

In order to assess differences between participants in the distribution of perceptual errors across the stimulus spectrum and to evaluate their impact on the generalization gradient, we grouped participants based on their differences in perceptual accuracy (using statistical clustering methods) and compared the obtained generalization gradients between groups. Cluster analysis is an explorative analytic tool (i.e., data driven) that enables researchers to group individuals based on similar data patterns (Hofmans, Ceulemans, Steinley, & Van Mechelen, 2015). As such, the current study aims to more thoroughly investigate the impact of perceptual errors on generalization.

In a human simple fear conditioning procedure, we tested the extent of generalization after an acquisition phase where a CS was reinforced in 80% of the trials by the presentation of an aversive picture (US). During the generalization phase, on each trial one of six GSs or the CS was presented and participants had to indicate (in a binary categorization task) whether the stimulus was the same as the one presented during the acquisition phase or not, after which US-expectancy ratings were collected and the US was either presented or not. In participants making few perceptual errors, we expected to observe a bell-shaped generalization gradient driven by the probability distribution of perceptual errors and a binary response strategy (high US-expectancy for stimuli identified as the CS and low US-expectancy for other stimuli). A flat gradient was expected for participants making many perceptual errors during CS trials (CS identified as GS) as it would lead to similar perceived CS-US and GS-US contingencies. Thus, perceptual errors (perception of a GS as CS) were expected to heighten US-expectancy only for those with few errors on CS trials.

2. Material and methods

2.1. Participants

A total of 134 participants took part in the study in exchange for course credits, of which 133 (117 females, mean age 18.43, SD 1.38)

were included in the final analyses (one subject did not produce any responses). This number is sufficient given benchmarking studies on cluster procedures (Steinley & Brusco, 2011). Participants were all 1st year bachelor students in psychology at KU Leuven and received a course credit as compensation. The study was approved by the local university ethics committee (G-201610641).

2.2. Stimuli and apparatus

Seven differently sized white circles were created (varying in diameter from 7.37 to 11.94 cm in steps of 0.762 cm) against a black background (similar to *13*). In line with standard practices to assess generalized responding at both sides of the CS across the stimulus spectrum (Ghirlanda & Enquist, 2003), the middle circle served as CS, the others as GSs. Participants individually selected one of three aversive pictures of the International Affective Picture System [IAPS (Lang, Bradley, & Cuthbert, 2008);] as US, based on written content descriptions. 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 too aversive. Three levels of aversiveness were adopted similar to (Struyf et al., 2017). The most strongly aversive US [selected by 57.9% of the participants (n = 77)] was an image of a bloodied corpse [in young adults rated as a 7.77 on arousal (1 = calm, 9 = excited) and as a 1.77 on valence (1 = unpleasant, 9 pleasant) (Grühn & Scheibe, 2008)], the moderately aversive US [selected by 39.8% of the participants (n = 53)] was a picture of a disabled child [arousal rating of 5.86 and valence rating of 3.14 in young adults (Grühn & Scheibe, 2008)], and the mildly aversive US [selected by 2.3% of the participants (n = 3)] was a picture of a holstered gun [arousal rating of 4.92 and valence rating of 3.65 in young adults (Grühn & Scheibe, 2008)]. In our study, the moderate US received an average unpleasantness score of 3.57 (SD 2.07) and the strongly aversive US was rated 7.63 (SD 1.41) on a visual analogue scale (VAS) ranging from 0 ('not unpleasant') to 10 ('very unpleasant'); no scores were obtained for the mild US.[1] The experiment was programmed in Affect 4 (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010).

2.3. 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 own pace. An experimenter was present to provide assistance and to make sure that no communication occurred among participants.

Participants were acquainted with the experimental task during 5 practice trials. For this purpose, differently sized squares were used instead of the CS or GSs. On every trial one of two squares was presented, which participants had to categorize as same or different (one was presented prior to the discrimination task), after which a US-expectancy rating was obtained. The actual experiment comprised a habituation (6 trials), acquisition (12 trials) and generalization phase (48 trials). The habituation and acquisition phase were presented successively without break, and only comprised CS trials. During each habituation trial, the CS was presented on screen for 8000 ms. No perceptual categorization was prompted during these trials. A US-expectancy scale ranging from 1 ('certainly no picture') to 10 ('certainly picture') appeared at the bottom of the screen 3000 ms after CS onset and remained present until the CS disappeared. Trials were separated by a 2000 ms

---

[1] These averages are based on only 15% of the participants, as the majority failed to provide an unpleasantness rating due to a relatively short response window.
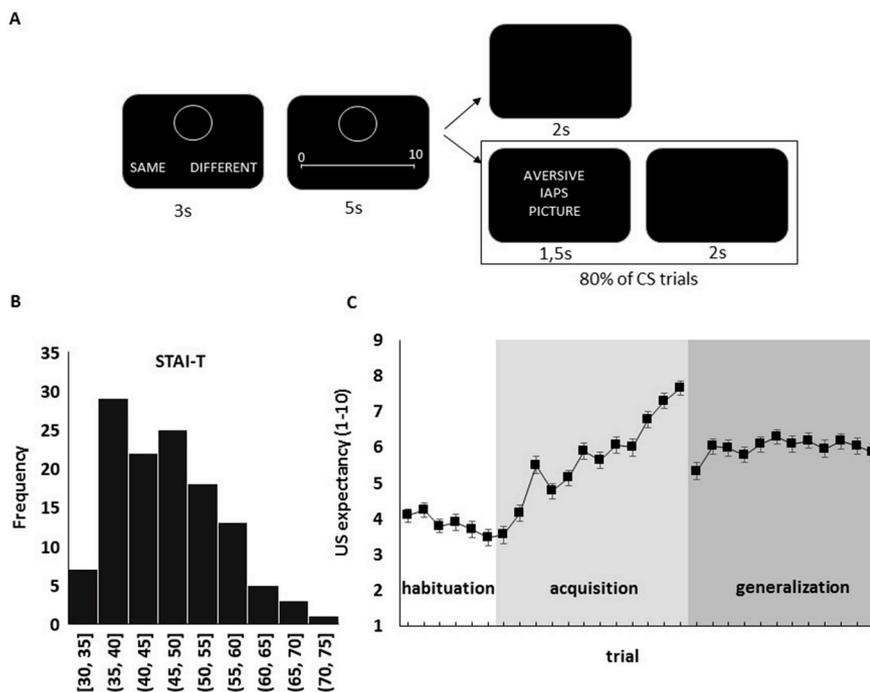
Fig. 1. (A) Trial structure during the generalization phase. Upon presentation of the CS or GS stimulus, participants indicated whether the stimulus was the same or different as the one presented during the acquisition phase. They then indicated their expectancy to be presented with the US, after which either the US (on 80% of the CS trials) or a blank screen (on GS trials and 20% of the CS trials) followed. **(B)** Distribution of STAI-T scores. **(C)** Mean US expectancies during CS trials throughout the experiment. Error bars represent standard errors of the means.

intertrial interval (ITI) during which the screen remained black. Acquisition trials had an identical trial structure apart from the fact that 80% of the trials included 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 'identical' or 'different' from the circle presented during the preceding phase in addition to their US-expectancy ratings. On every generalization trial one of the GS or the CS was presented for 8000 ms. Categorization response buttons appeared at the start of each trial (simultaneous with circle onset) at the bottom of the screen for 3000 ms after which the US-expectancy scale appeared for 5000 ms (Fig. 1A). 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 contained 12 CS trials and 36 GS trials (6 trials per GS). Trials were presented in a pseudorandom order composed of six consecutive blocks containing 6 GS trials (one of each GS) and 2 CS trials each, randomized within each block. 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 (less the 6% of the data).

Interindividual differences in trait anxiety were measured using the STAI-T questionnaire (range: 20–80, mean: 46.81, SD: 8.87, see Fig. 1B), 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 20 participants due to a programming glitch.

2.4. Data analysis

All analyses have been reported and the study has been pre-registered on the Open Science Framework (OSF) (osf.io/u3fcs). All data is publicly available at OSF (osf.io/u3fcs). Data from the habituation and acquisition phase (combined) were analyzed with a mixed model that had Trial as a fixed effect and a subject-dependent intercept (random effect). Mixed models provide powerful and flexible approaches to analyzing repeated-measures data (Blackwell, de Leon, &

Miller, 2006) and use the Satterthwaite approximation to estimate the degrees of freedom (Satterthwaite, 1946).

Categorization data per stimulus were transformed in a stimulus' probability of being categorized as the CS ('same' response) (calculated as the number of 'same' responses for a given stimulus divided by the total number of recorded responses ('same' or 'different' response) for that stimulus), resulting in a probability distribution across the stimulus dimension, and were then analyzed using mixed models with Stimulus (continuous; 4–10) and Stimulus^2 (the product of Stimulus with itself, to model the bell-shaped curve) as fixed effects and a subject-dependent intercept (random effect). Post-hoc tests applied the adjusted Bonferroni correction. Next, a cluster analysis using the k-means algorithm, as implemented in MATLAB (Hofmans et al., 2015), was used to identify groups of participants (called clusters) with similar perceptual performance across the stimulus dimension (i.e., the probability distributions of perceptual errors). The k-means algorithm identifies clusters through iterative minimization of the sum of point-to-centroid distances (using the squared Euclidean distance measure), where the centroids are computed as the mean score profile per cluster. Each subject is allocated to the cluster for which its squared Euclidean distance is minimal. The maximum number of centroids was set to 10. We ran the analysis 10000 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 an automated scree test (Ceulemans & Kiers, 2006; Wilderjans, Ceulemans, & Meers, 2013), the solution with four clusters and thus four centroids was preferred.

The US-expectancy data (of the generalization phase) were analyzed with a mixed model. Model 1 comprised Trial, Stimulus and Stimulus^2 (to model the bell-shaped curve) as fixed effects. In Model 2, Categorization and its interaction with Stimulus and Stimulus^2 were additionally included. Explorative, we tested for differences in overall gradients – irrespective of categorization – between the different clusters in a third model (fixed effects: Trial, Stimulus and Stimulus^2, Cluster, Cluster × Stimulus and Cluster × Stimulus^2). In a final model (model 4), the model was further extended with Categorization as well as its interactions with Stimulus, Stimulus^2, and Cluster. In all models a random intercept was included. All performed analyses and (computed) measures have been reported.

3. Results

3.1. Acquisition

As expected, US-expectancy ratings increased across trials in the first phase of the experiment (habituation and acquisition) [Trial effect: $F(1, 2151.31) = 554.56$, $p < .001$], suggesting that participants learned the CS-US association. Exploratory analyses revealed that US-expectancy ratings decreased slightly over the course of habituation ($\beta = -0.13$, 95% CI [-0.20 to -0.07], $p < .001$), whereas from acquisition onwards expectancy ratings increased ($\beta = 0.32$, 95% CI [0.28–0.35], $p < .001$) (Fig. 1C).

3.2. Generalization

3.2.1. Categorization

Across the entire sample, the CS was correctly identified on 65.48% of the CS trials during the generalization phase, with large variations between individuals (SD: 29%). On almost half of the GS trials (41.35%, SD: 17%), the presented stimulus was incorrectly identified as the CS. The probability that a stimulus was identified as the CS increased as the size difference between the GS and CS decreased, resulting in a bell-shaped perceptual categorization gradient [Stimulus effect: $F(1, 791.52) = 146.88$, $p < .001$; Stimulus^2 effect: $F(1, 791.51) = 207.16$, $p < .001$] (Fig. 2A). We exploratory tested for differences between the CS and each GS and found significant differences for all GSs (all p's < 0.001) except GS$_3$ ($p = .61$).

Cluster analyses revealed 4 clusters in the classification data (Fig. 2A). In Cluster 1 (called the 'CS bias group'), CS classifications were overall high for all stimuli, with a slight decrease for the most extreme GSs. In Cluster 2 (called the 'high accuracy group'), CS classifications peaked around the CS and declined strongly as GSs differed more from the CS. In Cluster 3, CS classifications were overall low and slightly skewed to GSs larger than the CS (called the 'GS bias group'). In Cluster 4 (called 'the skewed group'), CS classifications were also low and skewed with a peak towards the smallest circles (GS1 and GS2). 32.6% of the participants (n = 43) were allocated to the CS bias group (Cluster 1), 17.4% (n = 23) to high accuracy group (Cluster 2), 24.2%

(n = 32) to the GS bias group (Cluster 3) and 25.8% (n = 34) to the skewed group (Cluster 4).

Explorative, two indices were calculated to test whether perceptual errors affect experienced CS-US and GS-US contingencies: (1) total number of reinforced CS trials with a correct categorization (*same* response on CS trials) relative to the total number of trials on which the stimulus was identified as CS (*same* response on all trials). This parameter was used as a proxy for the perceived CS-US relationship. (2) Total number of reinforced CS trials with an incorrect categorization (perceived as different from the CS) relative to the total number of trials on which the stimulus was identified as different from the CS. This parameter was used as a proxy for the perceived GS-US relationship. A one-way ANOVA revealed significant differences between the four clusters in both perceived CS-US and GS-US contingencies [CS-US contingency: $F(2, 131) = 35.73$, $p < .001$; GS-US contingency: $F(2, 131) = 14.00$, $p < .001$]. Post hoc testing (Table 1) revealed no differences between the CS bias group (Cluster 1) and the GS bias group (Cluster 3) or the skewed group (Cluster 4) for either contingencies. The high accuracy group (Cluster 2) compared to all other groups (clusters 1, 3 and 4) had significantly higher CS-US contingencies and significantly lower GS-US contingencies (see Table 1). The GS bias group (Cluster 3) had a higher perceived CS-US contingency compared to the skewed group (Cluster 4) but did not differ on perceived GS-US contingency. Both in the CS bias group (Cluster 1) and the skewed group (Cluster 4) the CS-US contingency did not differ from the GS-US contingency [Cluster 1: $t(22) = 0.98$, $p = .34$; cluster 4: $t(33) = 0.20$, $p = .84$] whereas in the high accuracy group (Cluster 2) and the GS bias group (Cluster 3) the CS-US contingency was significant larger compared to the GS-US contingency [cluster 2: $t(42) = 20.41$, $p < .001$; Cluster 3: $t(31) = 4.60$, $p < .001$].

3.2.2. Relationship with STAI-T

We found no relationship between the STAI-T score and the probability of being allocated to one of the four identified clusters in the classification data [$\chi^2_{(3)} = 1.29$, $p = .73$].

3.2.3. US-expectancy

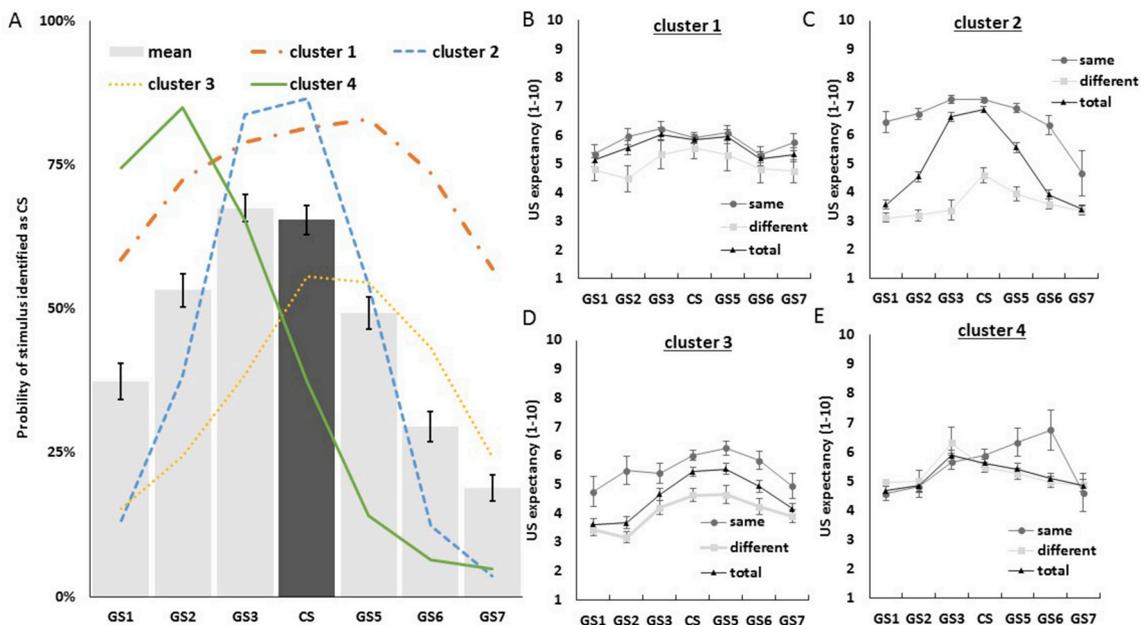As expected, we found a bell-shaped US-expectancy gradient across

Fig. 2. (A) Average probability per stimulus to be categorized as CS during the generalization phase (bars). Note that the light grey bars reflect perceptual errors as the presented stimulus was different from the CS. The different lines represent the four identified clusters within the categorization data. **(B–E)** Mean gradients of US expectancy per cluster with (trials on which the stimulus was categorized as CS = same; trials on which the stimulus was categorized as different stimulus = different) and without accounting for CS categorizations (all trials = total). Error bars represent standard errors of the means.

Table 1
Means and standard deviations of perceptual errors during CS trials, GS trials and mean perceived CS-US and GS-US contingencies for each of the identified clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
% errors CS	18.61% (17%)[a]	13.39% (12%)[a]	44.33% (28%)[b]	62.75% (21%)[c]
% errors GSs	70.64% (12%)[a]	34.27% (9%)[b]	33.39% (11%)[b]	41.73% (8%)[c]
CS-US cont.	23.25% (6%)[a,b]	38.98% (8%)[c]	28.78% (12%)[a]	19.03% (8%)[b]
GS-US cont.	17.99% (22%)[a]	5.11% (4%)[b]	13.68% (7%)[a]	19.53% (7%)[a]

Mean % errors during stimulus categorization for CS and GSs, and mean perceived contingencies between the CS and US (CS-US cont.) and between the GSs and the US (GS-US cont.). Standard deviation in brackets. Different superscripts ([a,b,c]) denote differences at $p < .05$, corrected for multiple testing by a factor 5 (i.e., adjusted Bonferroni correction).

the stimulus dimension with its maximum located at the position of the CS [Stimulus effect: $F(1, 6106.70) = 264.73$, $p < .001$ and Stimulus^2 effect: $F(1, 6106.69) = 265.21$, $p < .001$]. In a second model, we explored whether the categorization of a stimulus as CS or as GS affected the shape of the generalization gradients. We found no main effect of Categorization ($F(1, 5822.99) = 0.436$, $p = .51$), but differences in gradients depending on the categorization [Categorization × Stimulus effect: $F(1, 5822.33) = 5.14$, $p = .023$ and Categorization × Stimulus^2 effect: $F(1, 5820.54) = 4.99$, $p = .025$]. Post hoc we ran another model where the interaction terms Categorization × Stimulus and Categorization × Stimulus^2 were omitted as interactions can potentially affect true main effects in regression models (Aiken, West, & Reno, 1991). We found a very strong effect of Categorization [$F(1, 5814.36) = 463.65$, $p < .001$] demonstrating that US-expectancy ratings were overall higher when the stimulus was categorized as the CS (Fig. 3). Explorative, we tested for differences between clusters on the overall gradient without taking categorization into account (model 3). We found a significant main effect of Cluster [$F(3, 6139.87) = 44.92$, $p < .001$], as well as different shapes in generalization gradients among clusters [Cluster × Stimulus effect: $F(3, 6058.00) = 51.63$, $p < .001$; Cluster × Stimulus^2 effect: $F(3, 6058.03) = 56.66$, $p < .001$]. The different gradients in Fig. 2B–E demonstrate a relatively flat overall gradient in the CS bias group (cluster 1) and the skewed group (cluster 4), whereas for the high accuracy group (cluster 2) US expectancy decreases as a function of stimulus similarity, and likewise but to a much lesser extent for the GS bias group (cluster 3).
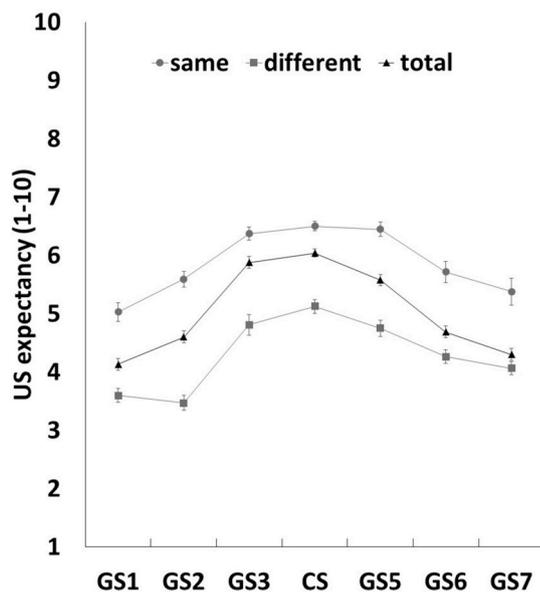

Fig. 3. Mean US expectancy regardless of stimulus categorization (all trials = total), when categorized as CS (trials on which the stimulus was categorized as CS = same) or categorized as different from the CS (trials on which the stimulus was categorized as different stimulus = different). Error bars represent standard errors of the means.

Next, we extended the model and included Categorization as well as its interactions (model 4). We found no main effect of Categorization [$F(1, 5761.17) = 1.75$, $p = .19$], but this effect was dependent upon Cluster [$F(3, 5756.97) = 74.69$, $p < .001$]. For Cluster 1, 2 and 3 the classification of a stimulus as CS increased US expectancy with the most profound increase for the high accuracy group (Cluster 2) (cluster 1: $\beta = 1.07$, 95% CI [0.71–1.43], $p < .001$; Cluster 2: $\beta = 3.51$, 95% CI [3.32–3.70], $p < .001$; Cluster 3: $\beta = 1.81$, 95% CI [1.53–2.09], $p < .001$). In the skewed group (Cluster 4) no effect of Categorization was found on US expectancy ($\beta = -0.06$, 95% CI [-0.33–0.21], $p = .64$) (see Fig. 2B–E). Furthermore, we found differences in the shape of the gradient between clusters [Cluster × Stimulus effect: $F(3, 5690.70) = 2.91$, $p = .033$ and Cluster × Stimulus^2 effect: $F(3, 5690.76) = 2.61$, $p = .050$]. The significant trial effect in model 1 and 3 [model 1: $F(1, 6106.36) = 3.98$, $p = .046$, model 3: $F(1, 5685.16) = 4.08$, $p = .044$] disappeared in models 2 and 4 [model 2: $F(1, 5703.05) = 1.61$, $p = .20$; model 4: $F(1, 5681.275) = 0.162$, $p = .68$]. Model parameters can be found in the supplemental information.

3.2.4. Relationship with STAI-T

Next, we repeated these four models but included STAI-T and its interactions. For Model 1, we found no main effect of STAI-T [$F(1, 5289.54) = 0.95$, $p = .33$] nor an effect on the shape of the gradient [STAI-T × Stimulus effect: $F(1, 5198.74) = 1.81$, $p = .18$; STAI-T × Stimulus^2 effect: $F(1, 5198.76) = 1.61$, $p = .21$]. Similarly for models 2–4, STAI-T scores did not affect the shape of the gradient, interact with the effect of cluster allocation (Model 2 all p's > 0.18), or with the effect of Categorization (Model 3: all p's > 0.25) or both (Model 4: all p's > 0.27).

4. Discussion

The inspection of gradients in conditioned responding has unequivocally demonstrated the robustness of generalization, yet failed to identify the mechanisms underlying this cross-species phenomenon. The inclusion of perception in addition to conditioned responses as sole outcome enables researches to investigate the extent to which gradients relate to perceptual errors. We found large variations in the amount of perceptual errors both for the CS and the GSs. This affected generalization gradients in a two-fold manner. First, stimuli during the generalization phase were associated with increased conditioned responding (i.e., US expectancy ratings) when they were identified as a CS compared to as a GS. Second, it led to differences in perceived contingencies between individuals that substantially differed from the objective CS-US and GS-no US probabilities for many individuals. The impact of a perceptual error on subsequent conditioned responding was dependent on the perceived contingencies. Furthermore, we found no influence of trait anxiety on the probability of perceptual errors nor on the gradient of conditioned responding. Most importantly, our data provide strong evidence that behavioral variations within and between individuals in a context of generalization are associated with variations in perceptual errors reflecting an inability to discriminate a GS from a CS.

Many recent studies on the role of perception in a context of generalization challenge the notion of *the perfect observer* (i.e., a one-on-one relationship between small physical manipulations and their perception), an assumption implicitly embedded in contemporary generalization protocols and analyses (Struyf et al., 2015). Associative learning, for instance, has been found to impair the ability to perceptually discriminate a CS from other stimuli (Laufer, Israeli, & Paz, 2016; Laufer & Paz, 2012; Resnik, Sobel, & Paz, 2011; Schechtman, Laufer, & Paz, 2010). These findings suggest perception as a potential key mechanism in generalization. Yet, the impact of these perceptual errors on behavior remains unclear.

We identified distinct patterns of perceptual errors that were associated with different generalization gradients. In the high accuracy group (Cluster 2) the distribution of perceptual errors was characterized by relatively few mistakes for the CS and the GSs (although they increased as GSs more closely resembled the CS). In this group, US expectancy ratings followed the typical bell-shaped gradient, with a peak in responding around the CS and a decline as CS-GS similarity decreases. However, the same behavioral data plotted based on its precursor (i.e., whether the circle was perceived as CS or as GS) revealed a very different pattern. Stimuli perceived as CS elicit a strong response, whereas the same stimuli identified as different from the CS result in very low US expectancy ratings. It was the combination of this (almost) binary response strategy with the probability gradient of perceptual errors that yield the typical shape of the generalization gradient in this cluster. In the GS bias group (Cluster 3), CS classifications where overall low, peaked around the CS and were skewed to GSs on the right of the CS. Similar to the high accuracy group, this pattern led to significantly larger perceived CS-US contingencies [as relative more trials on which the stimulus was perceived as CS (*same* response) were followed by the US compared to trials on which the stimulus was perceived as a GS (*different* response)] compared to the GS-US contingencies although CS-US contingencies were overall lower and GS-US contingences overall higher compared to those in the high accuracy group. In the GS bias group, we found that stimuli perceived as CS elicit higher US expectancy levels compared to the same stimuli identified as GS. The gradients obtained by plotting US expectancy based on the perceptual categorizations (same vs different) revealed similar gradients across the stimulus dimension, that differed strongly however in their overall US expectancy levels. The finding that US expectancy still slightly decreased across the different stimuli despite all being categorized as the same, we speculate, might indicate a hierarchical decision-making process where both the outcome of the perceptual decision (Struyf et al., 2015) and stimulus features (i.e., sensory input) (McLaren & Mackintosh, 2002) are used as determinants of the conditioned response. Our findings suggest however a proportionally much larger impact of the former compared to the latter on the strength of responding.

Despite distinct probability distributions in perceptual errors, similarly flat gradients for US expectancy are found for clusters 1 and 4. In the CS bias group (Cluster 1) relatively few errors were made during the CS, but many GS were incorrectly identified as CS, whereas in the skewed group (Cluster 4) many errors were made for the CS and for GSs that were smaller in size as the CS. Yet, US-expectancy gradients in both clusters are characterized by a flat slope. One plausible explanation could be that in both groups perceived CS-US and GS-US contingencies did not differ. According to (Hearst, 1968), generalization gradients emerge due to interacting excitatory and inhibitory properties which depend on the associative strength of a stimulus with the US or its absence. However, as perceived CS-US and GS-US contingencies become more similar, the inhibitory properties of the GSs will switch to excitatory properties. Consequently, the gradient is subject to excitatory forces across the entire stimulus spectrum albeit the net excitatory strength will be lower given the overall low perceived CS-US and GS-US contingencies (Rescorla & Wagener, 1972). Thus, as perceived CS-US and GS-US contingencies converge, their opposing effects on the

gradient (pulling it down or pushing it upward) dissolve towards an overall slight upward pushing force. Another interesting finding is that within this context the impact of a perceptual error on a CR is rather limited. More precisely, US expectancies were not (Cluster 4) or only slightly higher (Cluster 1) when a GS was perceived as CS compared to when the same stimulus was perceived as GS. In the eyes of the perceiver, both the CS and GSs can be predictive for the onset of the US.

We found no effect of trait anxiety scores on the extent of generalization. Using a differential fear conditioning procedure, Torrents-Rodas et al. (2013) also did not find an effect of trait anxiety on the shape of the generalization gradient. In their study the US was (most likely) more threatening as it was an uncomfortable electrical shock. In line with our findings, Struyf et al. (2017) found no effect of trait anxiety on the probability that a stimulus was perceived as either GS or CS nor on the overall generalization gradient. They did find, however, a relationship between trait anxiety and the probability that a stimulus during the generalization phase was identified as the safety cue from the acquisition phase (CS-). Furthermore, an overall weak effect of trait anxiety on fear responding was reported when these CS- categorizations were included in the generalization data. However, their findings are preliminary and replication seems warranted. Together these findings suggest that differences in trait anxiety are a poor predictor for variations in generalization gradients.

Despite these interesting findings some limitations should be acknowledged. First, we only measured US-expectancy due to practical reasons (i.e., testing location). Yet, similar gradients have been found for both psychophysiological and self-reported measures (Lissek et al., 2010; Torrents-Rodas et al., 2013) and US-expectancy ratings have been demonstrated to be a valid proxy for fear learning (Boddez et al., 2013). Similar to previous studies (Lenaert et al., 2014; Struyf et al., 2017), participants controlled the US aversiveness level as they selected one out of three IASP pictures of different levels of aversiveness, such that the US was aversive but tolerable. This procedure has been developed as an analogue to calibration protocols with electrocutaneous stimuli as US. Unfortunately, comparison of valence ratings between the different USs was not possible due to the small number of participants that rated US valence prior to the experiment. Furthermore, the explicit assessment of perception through the categorization task most likely promoted attention to distinctive stimulus features which should result in sharper gradients. Therefore, it would be interesting to replicate the current paradigm with physiological measures and a control group without a categorization task to investigate the effects of a perceptual categorization task on generalization gradients. No control group was included in the current study as the focus was on inter-individual differences in perceptual errors and their relationship to generalized responding. Finally, the decision made during the categorization task might bias US expectancy. For example, one could argue that when a stimulus is categorized as being different from the CS, one would consequently also not expect the US. However, the gradients in response strength per perceptual categorization contradict this assumption. In a previous study we found similar findings despite US expectancy being assessed prior to the categorization task (Struyf et al., 2017). Furthermore, this comparison process is inherent to contemporary generalization models (Lissek, 2012) and its explicit assessment most likely only affected the level of awareness. Finally, reasons for the large interindividual differences in the distribution of perceptual errors remain speculative but they may relate to differences in spatial tuning at primary visual brain regions (Moutsiana et al., 2016), biased perceptual decision-making, or be indicative of a memory bias (Dunsmoor, Mitroff, & LaBar, 2009). Future research should investigate the role of these distinct mechanism in more depth.

In sum, accounting for perception enables us to explain variations in behavior that otherwise would have been considered *noise*. Our data provide strong evidence that generalization is mainly driven by perceptual errors that affect behavior directly but also indirectly as they lead to different learning experiences between individuals. The large

variation in both perceptual and fear generalization patterns and their close relationship illustrate the need to further investigate inter-individual differences in perception, memory and decision-making in a context of fear generalization. Better insight in the mechanism underlying fear generalization could not only lead to vast theoretical improvements but also to more effective and person-tailored clinical practices.

Author contributions

JZ developed the paradigm. JZ collected the data. JZ analyzed the data. JZ wrote the manuscript. JZ, EC, DH and TB revised the manuscript.

Additional information

The authors declare that they have no competing interests.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2019.01.006.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions.* Thousand Oaks, Californiav: Sage Publishing.
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, P. M. (1972). Wavelength generalization and discrimination in the pigeon. *Perception & Psychophysics, 12*(4), 342–348. https://doi.org/10.3758/BF03207219.
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.
Dunsmoor, J. E., Mitroff, S. R., & LaBar, K. S. (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning & Memory, 16*(7), 460–469. https://doi.org/10.1101/lm.1431609.
Dunsmoor, J. E., & Paz, R. (2015). Fear generalization and anxiety: Behavioral and neural mechanisms. *Biological Psychiatry, 78*(5), 336–343. https://doi.org/10.1016/j.biopsych.2015.04.010.
Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour, 66,* 15–36.
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.
Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51*(1), 79–88. https://doi.org/10.1037/h0046219.
Hearst, E. (1968). Discrimination learning as the summation of excitation and inhibition. *Science, 162*(3859), 1303–1306. https://doi.org/10.1126/science.162.3859.1303.
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.
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.
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. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 32*(18), 6304–6311. https://doi.org/10.1523/JNEUROSCI.6281-11.2012.
Lenaert, B., Boddez, Y., Griffith, J. W., Vervliet, B., Schruers, K., & Hermans, D. (2014). Aversive learning and generalization predict subclinical levels of anxiety: A six-month longitudinal study. *Journal of Anxiety Disorders, 28*(8), 747–753. https://doi.org/10.1016/j.janxdis.2014.09.006.
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., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., et al. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. *American Journal of Psychiatry, 167*(1), 47–55. https://doi.org/10.1176/appi.ajp.2009.09030410.
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.
Mostofsky, D. I. (1965). *Stimulus generalization.* Staford University press.
Moutsiana, C., De Haas, B., Papageorgiou, A., Van Dijk, J. A., Balraj, A., Greenwood, J. A., et al. (2016). Cortical idiosyncrasies predict the perception of object size. *Nature Communications, 7,* 1–12. https://doi.org/10.1038/ncomms12110.
Onat, S., & Büchel, C. (2015). The neuronal basis of fear generalization in humans. *Nature Neuroscience, 18*(12), 1811–1818. https://doi.org/10.1038/nn.4166.
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.
Prokasy, W. F., Hall, J. F., & Ball, J. F. (1963). Primary stimulus generalization. *Psychological Review, 70*(4), 310–322. https://doi.org/10.1037/h0049354.
Rescorla, R. A., & Wagener, A. R. (1972). A theory of pavlovian conditionng: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.). *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
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.
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. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(31), 10460–10464. https://doi.org/10.1523/JNEUROSCI.2377-10.2010.
Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243.
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., 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 & Biobehavioral Reviews, 59*, 201–207. https://doi.org/10.1016/j.neubiorev.2015.11.004.
Torrents-Rodas, D., Fullana, M. A., Bonillo, A., Caseras, X., Andión, O., & Torrubia, R. (2013). No effect of trait anxiety on differential fear conditioning or fear generalization. *Biological Psychology, 92*(2), 185–190. https://doi.org/10.1016/j.biopsycho.2012.10.006.
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]. The Netherlands: Lisse.*
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., Madden, V. J., Iven, J., Wiech, K., Weltens, N., Ly, H. G. H. G., et al. (2017). Biased intensity judgements of visceral sensations after learning to fear visceral stimuli: A drift diffusion approach. *The Journal of Pain, 18*(10), 1197–1208. https://doi.org/10.1016/j.jpain.2017.04.011.