UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Zaman, J., Wester, A. Z. L., Yu, K., & Verheyen, S. (2026). Bitten by the Dog, Afraid of the Hedgehog: A Registered Report of the Asymmetry Effect in Conceptual Fear Generalization. *Collabra: Psychology*, *12*(1). https://doi.org/10.1525/collabra.155963

Cognitive Psychology

Bitten by the Dog, Afraid of the Hedgehog: A Registered Report of the Asymmetry Effect in Conceptual Fear Generalization

Jonas Zaman[1,2,3] [a†], Anna Zara Louise Wester[4†], Kenny Yu[2], Steven Verheyen[3]

[1] REVAL Rehabilitation Research, Hasselt University, Diepenbeek, Belgium, [2] Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium, [3] Brain & Cognition, Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, Netherlands, [4] Laboratory for Biological Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium

Keywords: fear, typicality, categories, conceptual generalization, replication

https://doi.org/10.1525/collabra.155963

Collabra: Psychology

Vol. 12, Issue 1, 2026

Dunsmoor and Murphy (2014) found that conditioned fear spreads from typical to atypical category exemplars but not vice versa, paralleling the principles of category-based induction. Their work was foundational in establishing that fear can generalize within conceptual structures along a dimension of category typicality and offers insight into how higher-order reasoning might be integrated in classical conditioning. This registered report aims to replicate and extend the findings of Dunsmoor and Murphy by investigating how idiosyncratic category structures and trait anxiety affect the asymmetry effect in conceptual fear generalization. It includes several methodological improvements and a comparison of different methods for processing skin conductance responses to identify the source of contradictory findings on the effect in the literature. Results from this multiverse approach revealed no asymmetry effect in explicit expectancy ratings or in skin conductance responses, contrary to the original study's findings. These findings suggest that while conceptual fear generalization relates to typicality gradients, the asymmetry effect is neither as robust nor as universal as originally proposed. The original premise of a direct parallel between associative learning and category-based induction warrants reconsideration, with additional research needed to clarify the conditions under which these cognitive domains interact.

Introduction

Associative learning, once thought of as a simple learning process, has been repeatedly demonstrated to rely not only on low-level stimulus features but also on conceptual knowledge during the transfer of learning (e.g., Bennett et al., 2015; Boyle et al., 2016; Damasio et al., 1991; Dunsmoor & Murphy, 2015; Dymond et al., 2011; Gerdes et al., 2020; Scheveneels et al., 2017). For instance, when category exemplars (conditioned stimulus, CS) are repeatedly paired with an aversive shock (unconditioned stimulus, US), conditioned fears generalize to other category exemplars (Dunsmoor et al., 2011, 2012, 2014). Moreover, Dunsmoor and Murphy (2014) were the first to demonstrate that the conceptual generalization of fear followed the principles of category-based induction (Heit, 2000; Osherson et al., 1990). They found that fear spread from typical to atypical category exemplars but not vice versa (called the asymmetry effect; Dunsmoor & Murphy, 2014), paralleling humans' reliance on category typicality to induce unknown properties of exemplars on the basis of other exemplars (i.e., the phenomenon of premise typicality; Osherson et al., 1990; Rips, 1975). The implications of this finding are potentially far-reaching. It brings together constructs from the fields of classical fear conditioning, semantic cognition, and higher-level reasoning. As such, it carries the promise that these three important but hitherto distinct areas of psychology might be integrated or at least mutually inform each other. However, for this cross-pollination to reach its full poten-

† Joint first authors

a Corresponding author: Jonas Zaman, PhD
REVAL Rehabilitation Research, Faculty of Rehabilitation Sciences, UHasselt
Universiteit Hasselt - Campus Diepenbeek, Belgium
Wetenschapspark 5-7 - B-3590 Diepenbeek
Tel: (+32) 011 / 26 85 31
E-mail: jonas.zaman@hasselt.be
Institutional URL : https://www.uhasselt.be/en/onderzoeksgroepen-en/reval

tial, further work into the exact nature of how conceptual representations impact category-based generalization is needed (Lei et al., 2019).

The asymmetry effect that Dunsmoor and Murphy (2014) identified has so far gone unreplicated, nor is it possible to reproduce or explore the original findings because the data are no longer available (personal communication Joseph E. Dunsmoor, 19/11/2021). The results from the original study were not straightforward either. The asymmetry effect of item typicality was only found for the psychophysiological measure (skin conductance response, SCR) but not for the self-report data (US expectancy ratings). Wong and Beckers (2021) used the same paradigm in a recent study. In individuals scoring low on trait anxiety, they found no fear generalization from typical to atypical exemplars based on SCR data but did so for the self-report data, while neither measure yielded an asymmetry effect in high trait anxiety individuals (due to no SCR generalization and similar levels of generalized US expectancy ratings in both groups). Another recent study failed to observe an effect of typicality, although here, category labels instead of exemplars were used during conditioning (Lei et al., 2019). In light of these conflicting findings and the observation that the premise typicality effect itself is also not uncontested (Hampton & Cannon, 2004; Hayes & Heit, 2018), a well-powered replication seems warranted before concluding that the transfer of learning occurs along a dimension that represents the typicality of category instances, as it supposedly does in higher-order reasoning.

If fear were indeed to spread asymmetrically depending on items' category typicality, crosstalk between classical fear conditioning, semantic cognition, and higher-level reasoning could not only be fruitful, but would be necessary. A one-therapy-fits-all approach would no longer be viable for the treatment of anxieties and phobias, since the basis for fear generalization would no longer be shared perceptual dimensions that are grounded in the environment, but rather conceptual representations that are known to be subject to individual differences (Hampton, 2020; Pecher & Zwaan, 2017). Work from the domain of semantic cognition that has documented the nature of the various ways in which individuals' internal category structures can differ (e.g., Verheyen et al., 2019; Verheyen & Storms, 2013) should then inform the assessment of idiosyncratic dimensions of generalization for therapy purposes. For instance, large differences between individuals can exist in category typicality ratings of exemplars (Barsalou, 1987, 1989, 1993; Rosch, 1999, 2011), which should theoretically affect generalization patterns if responses spread along a typicality dimension as Dunsmoor and Murphy (2014) have argued.

Here, we propose to replicate and extend the findings of Dunsmoor and Murphy (2014) by conditioning participants to fear three exemplars of a category and to investigate the extent to which this fear generalizes to other category members. The asymmetry effect will be assessed by comparing generalization from typical to atypical category exemplars with generalization from atypical to typical category exemplars in a between-subjects design. In addition to a fear learning and generalization phase, the experiment will include a typicality rating task, which was not administered in Dunsmoor and Murphy (2014). It is intended to map individuals' category structure to investigate whether insight into participants' idiosyncratic category structures allows one to predict inter-individual differences in generalization patterns. Finally, analogous to Wong and Beckers (2021), trait anxiety levels will be assessed using the short version of the Depression Anxiety Stress Scale (DASS-21: de Beurs et al., 2001; Lovibond & Lovibond, 1995). The additional measures are intended to identify the source of the discrepancies in the existing literature regarding the asymmetry effect.

Open Practices Statement

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2011). After being granted "in-principle acceptance", this Registered Report was preregistered using the dedicated Registered Report Protocol Preregistration option at the Open Science Framework (https://osf.io/hpd3s), where we also host all stimulus materials, the experiment, the data, and the analysis code (https://osf.io/48rbp/).

Method

Participants

The sample size estimation comprised the following steps. First, we calculated the partial eta squared (η_p^2 = 0.194, N = 37) for the Stimulus × Group interaction on the SCR data of Dunsmoor and Murphy (2014). This was based on the reported F-values and degrees of freedom using the formula provided in Lakens (2013)[1]. Second, a reanalysis of the SCR data of Wong and Beckers (2021) across their entire sample yielded a η_p^2 = 1.5e^{-4} for the Stimulus × Group interaction (N = 96) and a repeated measures correlation of r = 0.23. Next, we calculated the average effect size weighted by sample size (η_p^2 = 0.054). Finally, a sample size estimate of N = 74 [critical F = 3.97, dfs = (1, 72)], was obtained using GPower (v3.1) for a repeated measures ANOVA with a within-between interaction, two within-subject measurement points, two groups, a power of .90, a repeated measures correlation of .23, a nonsphericity correction of ε = 1, and an effect size f = .239. Individuals who self-reported one of the following were not admitted to the study: (1) a history of chronic pain or breathing, car-

1 These values were derived from a follow-up ANOVA with Trial and its interaction as additional predictors, as the authors do not report the F-value for the Stimulus × Group interaction in their main analysis. Furthermore, as only p-values and no F-values were reported for the US-expectancy data, we were unable to calculate an effect size for this dependent variable.

diac, cardiovascular, neurological, or psychiatric disorders, (2) pregnancy, acute pain, use of recreational drugs, ongoing recovery from severe trauma, advice from a general practitioner to avoid stress, any electronic implant (e.g., pacemaker). In case of technical problems, psychophysiology measurement difficulties, or failure to meet the data inclusion criteria (see below), additional participants (N = 6). were recruited to attain the target sample size for analysis. Adult healthy volunteers were recruited through the university's on-site and online advertisement boards. We a priori expected our sample to have similar demographics (i.e., predominantly Caucasian and female) as in Dunsmoor and Murphy (2014) and Wong and Beckers (2021) and this proved to be the case (see Table 1 for demographics). Informed consent was obtained before participation. Monetary compensation of 12 EUR or course credits were provided for participation. The study proposal has been approved by the Medical Ethical Committee of the University Hospitals Leuven, Belgium (approval ID: G-2022-5666-R2(MAR)).

Stimulus and Apparatus

Visual Stimuli

Twenty-four different black-and-white pictures of various animals belonging to one of two categories (birds and mammals) were used. The procedure to select typical, atypical, and category exemplars of intermediate typicality was the same as in Dunsmoor and Murphy (2014). Because of cultural differences in typicality perception (Gabora et al., 2008; Malt, 1995), the stimuli that were ultimately selected were different than in the original article. These were based on Wong and Beckers (2021), who obtained typicality norming data in the student population from which we recruited participants. The typical category items are hummingbird, pigeon, and sparrow (birds); bear, cow, and gorilla (mammals). The atypical items are cassowary, emu, and penguin (birds); bat, platypus, and seal (mammals). Intermediate typicality items are duck, flamingo, kiwi, peahen, swan, and turkey (birds); alpaca, camel, dolphin, otter, rat, and sloth (mammals). Which animal category served as CS+ and CS- was counterbalanced across participants. For the CS+ category, depending on the group, participants were either conditioned to fear the typical exemplars and tested on the atypical exemplars or vice versa. For the CS- category, 3 of the 6 intermediate items were used during conditioning and the three remaining items during the test. It was randomly determined for every new participant which three intermediate items were used during acquisition and which ones during testing.

Electrocutaneous Stimulus

The electrocutaneous stimulus consisted of a 200ms monopolar square waveform pulse. Via two surface electrodes, it was presented at the distal end of the humerus of the non-dominant arm using a commercially available electrocutaneous stimulation device (Constant Current Stimulator, model DS7; Digitimer©, Hertfordshire, UK). The intensity of the electrocutaneous stimulus, serving as the unconditioned stimulus (US), was individually tailored to a highly annoying but not painful sensation. As detailed information regarding US calibration was lacking in Dunsmoor and Murphy (2014), we decided to implement the threshold of "highly annoying but not painful" as follows: An intensity VAS scale with the labels 0 = no sensation, 50 = painful, 100 = most extreme pain was adopted. US intensity would increase using the Ascending Methods of Limits approach (Yarnitsky et al., 1995), until the pain threshold was reached. Hereafter, intensities would decrease until they were no longer rated as painful. Finally, participants were asked to confirm whether this intensity level was "annoying but not painful" and informed that this would be used for the rest of the experiment.

Skin Conductance Responses

Electrodermal activity was recorded with two disposable Ag/AgCl electrodes (Biopac EL507) attached to the hypothenar eminence of the palm of the non-dominant hand, which was cleaned with tap water before the start of the procedure. The inter-electrode distance was 2.5 cm. A Coulbourn skin conductance coupler (LabLinc v71-23) provided a constant 0.5 V across electrodes. The signal was digitized at 1000 Hz throughout the experiment.

Procedure

The experiment comprised fear acquisition, a generalization test, a typicality rating task, and the administration of the DASS-21. As participant instructions were not included in the original paper, we opted for the following instructions at the start of the experimental procedure: "Different pictures will appear on the screen, which might or might not be followed by an electrocutaneous stimulus. Your task will be to indicate the extent to which you expect the electrocutaneous stimulus at each picture."During fear acquisition, an exemplar was presented for 6 seconds on every trial, and US expectancy ratings were obtained. Instead of the three response levels in the original study, we adopted a VAS scale ranging from 0 = certain US absence to 100 = certain US occurrence, to allow for more response differentiation (Peterson, 1997; Weijters et al., 2010). We adopted the scale labels from Wong and Beckers (2021) instead of those by Dunsmoor and Murphy (2014) (i.e., no risk, moderate risk, high risk) as they are more commonly used. On CS+ trials, picture offset was co-terminated with the US. A US was presented on two of the three repetitions per CS+ exemplar (66% reinforcement rate). Instead of the 10s fixation cross in the original study, a 20s fixation cross separated trials to ensure that SCR returned to baseline before the subsequent trial started (Breska et al., 2011). Depending on the group, either all three typical or all three atypical exemplars (from the CS+ category) were used as CS+. In both groups, three intermediate typicality exemplars of the CS- category served as controls. CS- trials were never followed by a US. The fear acquisition phase consisted of 3 blocks, totaling 18 trials, as the three different exemplars per CS category (CS+ vs. CS-) were each presented once per block. There was no break between blocks. Immediately following the fear acquisition phase, the generalization test commenced. Participants were not informed about this transition, nor that electrocutaneous stimuli would no longer occur. Depending on which exemplars were presented during fear acquisition, either all three typical or all three atypical exemplars served as novel test items for the CS+ category (GEN+). If the typical items were presented during acquisition, the atypical items were presented during generalization, and vice versa. For the CS-, the three remaining intermediate exemplars that were not presented during acquisition served as novel test items (GEN-). During the generalization test, participants continued to make US expectancy ratings, but electrocutaneous stimuli (USs) were no longer presented. This phase also consisted of 3 blocks, with each exemplar presented once per block (2 × 3 × 3 = 18 trials).

After the generalization test, participants were asked to provide typicality ratings for all 12 exemplars of the CS+ category using a 7-point Likert scale. Following Rosch and Mervis (1975), participants were told: "*to select 1 if they think the stimulus fits very badly with the image or idea they have of the* CS+ *category; select 4 if they think the stimulus fits fairly well with the image or idea they have of the* CS+ *category; select 7 if they think the stimulus fits very well with the image or idea they have of the* CS+ *category*." Participants were further encouraged to employ the typicality scale's full range to provide nuance to their responses. No time restrictions were imposed on the typicality rating task. To conclude the experiment, participants filled out the DASS-21 (de Beurs et al., 2001; Lovibond & Lovibond, 1995), which is the short version of the Depression Anxiety Stress Scale, after which they were debriefed. The DASS-21 comprises of three subscales for which the internal consistency (Cronbach alpha) in the current sample were the following: DASS (stress) = .79, DASS (anxiety) = .58, and DASS (depression) = .80.

Measures

SCR data were visually inspected for problematic recordings (e.g., hairy caterpillar-like signals, physiologically impossible jumps in the signal due to electrode disconnection) by somebody unfamiliar with the conditions from which the data originated, and excluded from analyses, if the case. Based on recent work on the effects of different SCR processing methods (Kuhn et al., 2022; Sjouwerman et al., 2022) and the inconsistent SCR findings between Dunsmoor and Murphy (2014) and Wong and Beckers (2021), we decided to analyze skin conductance twice: using a Trough-to-peak (TTP) scoring method similar to Dunsmoor and Murphy (2014) and using a Peak scoring method as adopted in Wong and Beckers (2021). For the TTP analysis, we used LedaLab's (MATLAB ©) Continuous Decomposition Analysis (CDA), which should be noted as a model-based decomposition method rather than a traditional TTP approach (Kuhn et al., 2022). While a non-model-based TTP method would more precisely replicate Dunsmoor and Murphy's original methodology, the original authors confirmed during the registered report review that our Ledalab-based approach provided an acceptable approximation of their analysis. For this implementation, we downsampled the data to 100 Hz, applied a 5 Hz low-pass Butterworth filter, and optimized the CDA analysis parameters twice. A SCR was considered related to stimulus presentation if the trough-to-peak response began between 1 and 4s after stimulus onset and was >.02 microsiemens (S). Responses that did not fit these criteria were scored as zero. For the peak scoring method, SCRs were calculated by subtracting the mean skin conductance level (SCL) during the 2s interval preceding picture onset from the maximum value in the 6s interval post-stimulus onset. Negative values or values < .02 microsiemens were set to 0 (Dawson et al., 2007). Both TTP and SCR values were square root transformed to reduce the distribution skewness of non-zero responses.

Data Inclusion Criteria

The same quantifiable inclusion criterion as reported in Wong and Beckers (2021) was used since, in the target paper, this criterion was not clearly described (i.e., failure to follow instructions). Successful differential learning of CS-US contingencies was defined as higher averaged expectancy ratings for the CS+ than for the CS- in the last acquisition block (i.e., the last three trials of CS+ and the last three trials of CS-, N = 6 failed to meet this criterion and were replaced). In addition, to ensure that all typical exemplars were effectively rated as (numerically) more typical than atypical exemplars, participants were excluded if this was not the case (N = 0).

Analyses

We adopted a multiverse approach (Kuhn et al., 2022; Sjouwerman et al., 2022) to investigate the robustness of the findings across different preprocessing pipelines and alternative ways of representing the data. We analyzed the data (US expectancy, SCR TTP, and SCR Peak) twice: once using the mean responses per stimulus type per phase [as done in Dunsmoor and Murphy (2014)], and once with means calculated per block instead of across the entire phase [as done in Wong and Beckers (2021)]. Mean CS+ and CS- responses in the acquisition phase were based on trials 2-9 of the acquisition phase [as done in Dunsmoor and Murphy (2014)], as no learning had occurred yet on the first trial. For mean GEN+ and GEN- responses, all generalization trials were included. The rmANOVA included the following predictors: Stimulus (for acq: CS+/ CS-; for gen: GEN+ / GEN-) as within-subject factors and Group (CS+: typ / CS+: atyp) as a between-subjects factor. All main and interaction effects were included. In the second analysis, the factor Block (1 / 2 / 3) and its interactions were added. Furthermore, two sensitivity analysis were conducted: 1) on the data of the first generalization block only, as the previous studies (Dunsmoor & Murphy, 2014; Wong & Beckers, 2021) found the strongest effects there, and 2) on the difference scores between CS+ and CS- (or GEN+ and GEN-) as done in Dunsmoor and Murphy (2014) (as this analysis led to identical conclusions we did not report it, see SI). As effect size, we report Partial Eta Squared (η_p^2) and corresponding two-sided 90% confidence intervals (90% CI's). An α = .05 was employed in all analyses.

In light of the varied outcomes in the literature and in order to remain unbiased, we reserved predictions as to the outcome of the experiment. A significant stimulus x group interaction in the generalization data was considered evidence for the asymmetry effect for the dependent variable under consideration, provided that the interaction takes the theoretically predicted shape (stronger generalization from typical to atypical exemplars than vice versa) and that providing group does not exert a significant effect (on its own or in an interaction) in the acquisition data. In the case of a non-significant stimulus × group interaction in the generalization data, Bayes factors were computed to quantify the strength of evidence supporting the absence of an effect[2]. Bayes factors (BF_{10}) below 1 indicate evidence in favor of the null hypothesis, with values between 1/3 and 1 generally interpreted as weak evidence, between 1/10 and 1/3 as moderate evidence, and below 1/10 as strong evidence for the absence of an effect (Jeffreys, 1961; Lee & Wagenmakers, 2013).[3]

Finally, the influence of idiosyncratic exemplar typicality was assessed on US expectancy ratings, which was not done by Dunsmoor and Murphy (2014) or Wong and Beckers (2021), by comparison of two Generalized Linear Mixed Models (GLMM). Trial-by-trial generalized responses were modeled using typicality (entered as a continuous predictor rather than a binary factor) and trial number as linear predictors. In the first model, typicality was based on group-averaged typicality ratings; in the second, on individual participants' ratings. Both models further comprised a random subject-dependent intercept. Model selection was made based on BIC comparison, with lower BIC values indicating better model fit. Deviating from the pre-registration, we tested for the effect of individual typicality ratings per group separately, so as not to confound this variable since both groups were exposed to items of different typicality levels during generalization testing.

All analyses were conducted in R (version 4.4.1) using the following packages: We used the afex package (Singmann et al., 2016) for conducting repeated-measures ANOVAs, ggplot2 (Wickham, 2016) for data visualization, dplyr (Wickham et al., 2023) for data manipulation, lme4 (Bates et al., 2015) for linear mixed-effects model analysis, effectsize (Ben-Shachar et al., 2020) for effect size estimations, and brms for Bayesian analysis (Bürkner, 2017).

Deviations from Replication Target

In sum, the experiment is a constructive replication of the study by Dunsmoor and Murphy (2014) with the following exceptions:

- We recruited Dutch-speaking undergraduate students in Belgium as participants as opposed to American participants in the original study. The study was therefore conducted in Dutch instead of English and the stimuli were chosen to accommodate the cultural backgrounds of Belgian undergraduates.
- During the acquisition and generalization phase, the stimuli were preceded by a 20s fixation cross, instead of a 10s one (Breska et al., 2011), to ensure that SCR returns to baseline before the next trial starts. This was done to rule out the possibility that lower SCRs to the CS- compared to the CS+ were an artefact of the presentation order and the short ISI. That is, CS- trials are often preceded by reinforced CS+ trials. Hence, lower SCR on CS- trials may merely be due to insufficient SCR recovery from the US (on the previous CS+ trial). In SCR analyses as adopted in the original paper, this could yield many negative SCR amplitudes for CS- trials, which are by custom reset to 0 and, and on average will artificially lead to lower SCR to the CS-. Therefore, the increment of the ISI should prevent this possibility without altering learning.
- US expectancy ratings were obtained using a VAS scale (0 = no risk, 100 = high risk) instead of using three levels to allow for more differentiation (Peterson, 1997; Weijters et al., 2010).
- The experiment included a newly added task: a typicality rating task, which was presented after acquisition and generalization so that it could not affect these phases. It allows for a check on the intended manipulation and for the investigation of idiosyncratic typicality effects.
- The DASS-21 (de Beurs et al., 2001; Lovibond & Lovibond, 1995) was administered to assess participants' trait anxiety and compare our sample levels to those reported in Wong and Beckers (2021). This allows us to explore the potential role of trait anxiety in any discrepancies between our results and those of Dunsmoor and Murphy (2014) and Wong and Becker (2021). The inclusion of the DASS-21 was to further establish to what extent conceptual fear generalization is a general phenomenon as was suggested by Dunsmoor and Murphy (2014) or pertains to specific subgroups of participants.

We consider the proposed changes to the materials necessary to accommodate any cultural differences in conceptual structure and argue that the remaining changes constitute methodological improvements. We believe that if these changes would be responsible for a different outcome, the circumstances under which the asymmetry effect in con-

2 We did not use the package's default (standard) priors; instead, we specified the following priors for the Bayes factor calculations: prior intercept = normal(50, 20), prior fixed effect regression coefficients = normal(0, 20), and prior random intercept standard deviation = exponential(λ = 1). These priors allow a wide range of plausible effect sizes while ensuring stable model estimation.

3 The Bayesian analyses were not pre-registered but suggested during review by Yoann Stussi.

Table 1. Demographics

	ATYP group *Mean (SD)*	TYP group *Mean (SD)*	Group difference tests
Age	22.41 (5.30)	21.46 (3.01)	$t_{(72)}$ = -0.94, *p* = .35
Gender	F = 30, M = 7	F = 27, M = 10	$X_{(1)}$ = 0.31, *p* = .58
US int (mA)	14.75 (4.72)	13.72 (5.90)	$t_{(72)}$ = -0.83, *p* = .41
DASS (stress)	11.26 (7.87)	11.28 (6.52)	$t_{(72)}$ = 0.01, *p* = .99
DASS (anxiety)	7.68 (6.30)	7.44 (5.53)	$t_{(71)}$ = -0.17, *p* = .87
DASS (depression)	7.50 (5.17)	5.28 (6.92)	$t_{(70)}$ = 1.54, *p* = .13

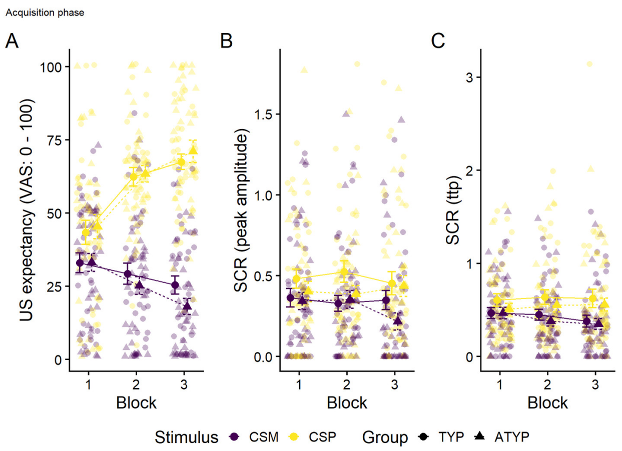

Figure 1. US expectancy ratings (A) and skin conductance responses (B-C) from the acquisition phase per block.

Error bars denote standard errors of the mean. SCR = Skin Conductance Response, VAS = Visual Analogue Scale), ttp = Through-to-Peak

ceptual fear generalization presents are too narrow to warrant further investigation of theoretical or clinical interest. The lead author of the target article, Dr. Joseph E. Dunsmoor, approved the proposed protocol.

Results

Demographics

In total, we ran 80 participants, but 6 were excluded due to failing the aforementioned data inclusion criteria. The demographics of the 74 participants whose data were retained for analysis are included in Table 1.

There were no significant group differences for any of the listed variables, as determined by independent samples t-tests. For gender, a chi-square test was used. Based on the cut-off criteria used in Wong and Beckers (2021), 31% of our sample fell in the low anxiety category (DASS-anxiety < 4, 23 participants) and 9% into the high anxious category (DASS-anxiety > 14) with the majority (60%) of our sample scoring between these cut-offs.

Acquisition – US Expectancy

Participants gave higher US expectancy ratings during CS+ trials compared to CS- trials [Stimulus effect: $F(1, 72)$ = 177.72, $p < .001$, η_p^2 = .71, *90% CI* = (.60, .79)] with no significant group differences herein [Group effect: $F(1, 72)$ = 0.17, p = .678, η_p^2 = .00, *90% CI* = (.00, .07); Stimulus × Group effect: $F(1, 72)$ = 1.69, p = .198, η_p^2 = .02, *90% CI* = (.00, .13)]. Next, we included block to investigate changes throughout the acquisition phase (see Fig. 1A). As expected, US expectancy rating evolved differently for the CS+ and CS- across blocks [Stimulus × Block effect: $F(1.8, 129.6)$ = 59.60, $p < .001$, η_p^2 = .45, *90% CI* = (.34, .55)] with increases in US expectancy across CS+ trials (β = 24.94, SE = 3.36, $p < .001$) and decreases across CS- trials (β = -11.30, SE = 2.65, $p < .001$). There were no significant differences in learning patterns between both groups [Group × Block: $F(1.6, 116)$ = 0.31, p = .685, η_p^2 = .004, *90% CI* = (.00, .04); Block × Stimulus × Group effect: $F(1.8, 129.6)$ = 0.93, p = .389, η_p^2 = .01, *90% CI* = (.00, .06)]. For full model output see SI.

Acquisition – SCR

Participants showed higher SCR responses to CS+ compared to CS- stimuli (Fig. 1B-C) using both scoring methods [Stimulus effect, SCR peak: $F(1, 72)$ = 17.99, $p < .001$, η_p^2 = .28, *90% CI* = (.12, .43); SCR TTP: $F(1, 72)$ = 31.64, $p < .001$, η_p^2 = .31, *90% CI* = (.14, .46)] with no significant group differences in overall responding [Group effect, SCR peak: $F(1, 72)$ = 0.75, p = .390, η_p^2 = .01, *90% CI* = (.00, .10); SCR TTP: $F(1, 72)$ = 0.52, p = .475 η_p^2 = .007, *90% CI* = (.00, .09)] or in their pattern of differential responding [Stimulus × Group effect, SCR peak: $F(1, 72)$ = 0.48, p = .491, η_p^2 = .006, *90% CI* = (.00, .09); SCR TTP: $F(1, 72)$ = 0.44, p = .508, η_p^2 = .006, *90% CI* = (.00, .09)]. When we reran the model with block and its interactions as additional predictors, we found different patterns between groups for the peak scoring method, while this failed to reach significance in the TTP scoring method [Stimulus × Group × Block, SCR peak: $F(1.96, 141.36)$ = 5.66, p = .005, η_p^2 = .07, *90% CI* = (.01, .16); TTP peak: Stimulus × Group × Block: $F(1.88, 135.69)$ = 0.42, p = .644, η_p^2 = .006, *90% CI* = (.00, .04)]. Uncorrected follow-up testing for the peak scoring method revealed only a significant decrease for the CS- in the Atypical group (β = -0.13, SE = 0.05, p = .015) but not in the Typical group (β = -0.02, SE = 0.05, p = .752), while for the CS+ there was no

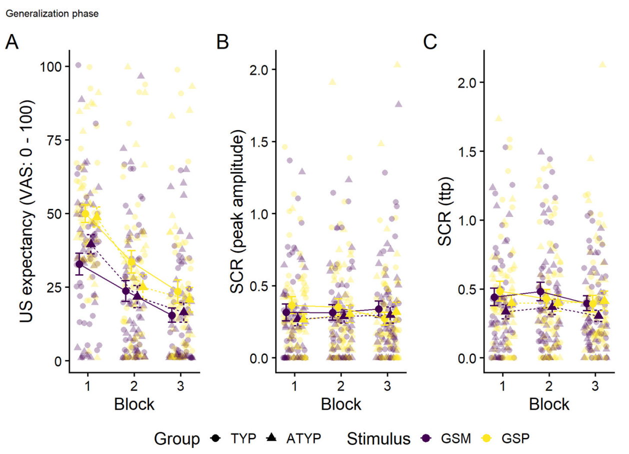

Figure 2. US expectancy ratings (A) and skin conductance responses (B-C) from the generalization phase per block. Error bars denote standard errors of the mean.

SCR = Skin Conductance Response, VAS = Visual Analogue Scale), ttp = Through-to-Peak

significant trend in either group (TYP: β = -0.03, *SE* = 0.05, *p* = .570, ATYP: β = 0.04, *SE* = 0.05, *p* = .464). For full model output, see SI.

Generalization – US Expectancy

US expectancy ratings were higher on GEN+ trials compared to GEN- trials [Stimulus effect: $F(1, 72) = 18.09$, $p < .001$, $\eta_p^2 = .20$, *90% CI* = (.06, .36)] but did not differ between groups [Group effect: $F(1, 72) = 0.08$, $p = .773$, $\eta_p^2 = .001$, *90% CI* = (.00, .06)]; Stimulus × Group effect: $F(1, 72) = 2.08$, $p = .153$, $\eta_p^2 = 0.03$, *90% CI* = (.00, .14), $BF_{10} = 0.46$]. When we reran the model with block and its interactions as additional predictors, no group differences emerged either [Block × Stimulus × Group effect: $F(1.9, 137.2) = 0.37$, $p = .678$, $\eta_p^2 = 0.005$, *90% CI* = (.00, .04), $BF_{10} = 0.18$], nor when we only included the data from the first block [$F(1, 72) = 2.44$, $p = .122$, $\eta_p^2 = 0.03$, *90% CI* = (.00, 0.15), $BF_{10} = 0.50$] (Fig. 2A). Exploratorily, we investigated the moderating role of individuals' anxiety scores (as a continuous predictor) but did not find evidence thereof [Stimulus × Group × Anxiety effect: $F(1, 69) = 0.64$, $p = .427$, $\eta_p^2 = .009$, *90% CI* = (.00, .10); Stimulus × Anxiety effect: $F(1, 69) = 0.35$, $p = .555$, $\eta_p^2 = .005$, *90% CI* = (.00, .09); Group × Anxiety effect: $F(1, 69) = 0.04$, $p = .834$, $\eta_p^2 = .000$, *90% CI* = (.00, .05); Anxiety effect: $F(1, 69) = 0.77$, $p = .384$, $\eta_p^2 = .01$, *90% CI* = (.00, .10)]. For full model output, see SI.

Generalization – SCR

We found no differential SCR responses on GEN+ trials compared to GEN- trials when averaging across all trials of the generalization phase [Stimulus effect, SCR peak: $F(1, 72) = 0.34$, $p = .559$ $\eta_p^2 = .005$, *90% CI* = (.00, .08); SCR TTP: $F(1, 72) = 2.24$, $p = .139$, $\eta_p^2 = .03$, *90% CI* = (.00, .14)] nor were there group differences [SCR peak: Group effect: $F(1, 72) = 0.28$, $p = .596$, $\eta_p^2 = .004$, *90% CI* = (.00, .08); TTP scoring: $F(1, 72) = 0.83$, $p = .365$, $\eta_p^2 = .01$, *90% CI* = (.00, .10)], or did we find evidence for the asymmetry effect [Stimulus × Group effect, SCR peak: $F(1, 72) = 0.03$, $p = .857$, $\eta_p^2 = .005$, *90% CI* = (.00, .05), $BF_{10} = 0.006$; SCR TTP: $F(1, 72) = 2.88$, $p = .094$, $\eta_p^2 = .04$, *90% CI* = (.00, .16), $BF_{10} = 0.033$]. When we reran the model with block and its interactions as additional predictors, no group differences emerged either [Block × Stimulus × Group effect, SCR peak: $F(1.77, 127.57) = 1.46$, $p = .236$, $\eta_p^2 = .02$, *90% CI* = (.00, .08), $BF_{10} = 0.36$; SCR TTP: $F(1.92, 138.30) = 0.58$, $p = .557$, $\eta_p^2 = .008$, *90% CI* = (.00, .05), $BF_{10} = 0.015$], nor when we only included the data from the first block [Stimulus × Group effect, SCR peak: $F(1, 72) = 1.17$, $p = .283$, $\eta_p^2 = .02$, *90% CI* = (.00, .11), $BF_{10} = 0.017$; SCR TTP: $F(1, 72) = 0.09$, $p = .768$, $\eta_p^2 = .001$, *90% CI* = (.00, .06), $BF_{10} = 0.71$] (Fig. 2B-C). Exploratorily, we examined the moderating role of individual anxiety scores. Evidence was mixed: a significant three-way interaction emerged in the SCR peak data but not in the SCR TTP data [Stimulus × Group × Anxiety: SCR peak, $F(1, 69) = 6.58$, $p = .012$, $\eta_p^2 = .09$, *90% CI* = (.00, .23); SCR TTP, $F(1, 69) = 0.06$, $p = .807$, $\eta_p^2 = .00$, *90% CI* = (.00, .06)]. Uncorrected follow-up contrasts tested for group differences in estimated marginal means for the scenarios of ± 1 SD (5.70) to the average anxiety score (7.52). This revealed no group differences at + 1 SD [$t(69) = 1.59$, $p = .116$], but a significant difference at -1 SD ('low anxiety') [$t(69) = -2.01$, $p = .049$]. This effect appeared to be driven by differences in SCR peak responses during GEN– trials. For full model output, see SI.

Idiosyncratic Exemplar Typicality Effects

Despite considerable variability in typicality ratings (see Fig. 3), individual typicality ratings did not account better for participants' US expectancy during the generalization phase than group-averaged ratings did (TYP group: $BIC_{individual} = 3100$ vs. $BIC_{group} = 3082$, ATYP group: $BIC_{individual} = 3031$ vs. $BIC_{group} = 3025$). Surprisingly, in both groups there was a negative relationship between group-averaged typicality ratings and generalized US expectancy ratings [TYP group: β = -14.00, SE = 3.67, ATYP group: β = -8.55, SE = 3.07]. For full model output, see SI.

Discussion

This study sought to replicate and extend Dunsmoor and Murphy's (2014) work on conceptual fear generalization and the asymmetry effect—the phenomenon whereby fear is thought to generalize more readily from typical to atypical category members than vice versa. Our results showed that participants conditioned to fear typical exemplars (TYP group) did not expect the aversive stimulus more when presented with atypical exemplars than participants conditioned to atypical exemplars (ATYP group) expected it when presented with typical exemplars (supported by both frequentist and Bayesian analysis). Neither did we observe the asymmetry effect when block was included, nor when focusing only on the data of the first generalization block. The effect also did not appear in skin conductance response (SCR) measurements, regardless of whether a Trough-to-

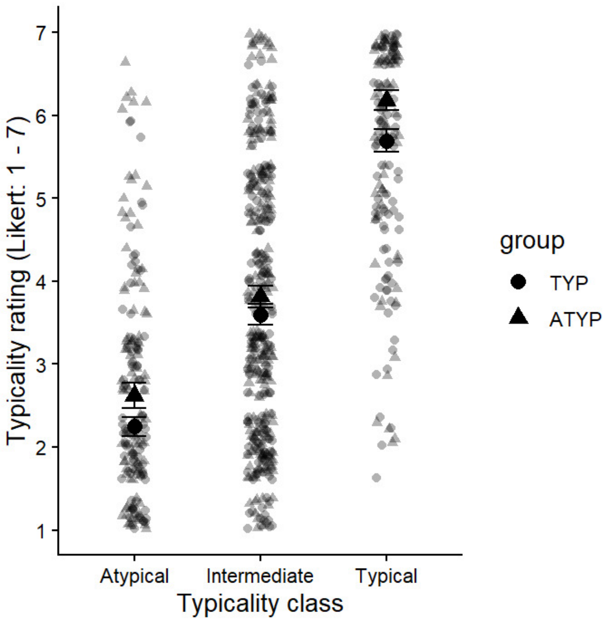

Figure 3. Distribution of the typicality ratings per class.

There was a strong effect of the typicality class with ratings corresponding to the created classes [Class effect: $F(1.69, 115.26) = 338.58$, $p < .001$*, $\eta_p^2 = 0.83$, *90% CI = (.79, .87)] and no group differences [Group effect: F(1, 68) = 2.45, p = .122* , $\eta_p^2 = 0.03$, *90% CI =* (.00, 1.16)*; Group × Class effect: F(1.69, 115.26) = 0.55, p = .549*, $\eta_p^2 = 0.008$, *90% CI =* (.00, .05)*]. Error bars denote standard errors of the mean.

peak or a Peak scoring method was used. Our findings reveal a distinct pattern when compared to previous research on the asymmetry effect. Dunsmoor and Murphy (2014) found evidence in SCR data but not in expectancy ratings, while Wong and Beckers (2021) detected the effect only in the US expectancy ratings (and not in SCR data) of low-anxious but not high-anxious participants. The failure to replicate the asymmetry effect raises questions about its robustness. Our findings suggest that the asymmetry effect in conceptual fear generalization may not be as reliable or generalizable as previously thought.

This replication helps rule out several methodological explanations for the contradictory findings in previous work on conceptual fear generalization. First, the absence of an asymmetry effect in US expectancy ratings in Dunsmoor and Murphy (2014) cannot be attributed to their use of a limited 3-point response scale. In our study, we adopted the same continuous VAS scale as Wong and Beckers (2021), who did observe the asymmetry effect in their low-anxiety subgroup—yet we did not replicate the effect, even with this improved sensitivity. Second, differences in SCR scoring methods cannot account for the discrepancy across studies. We applied a similar trough-to-peak method to Dunsmoor and Murphy (2014) as well as the peak method used by Wong and Beckers (2021), and neither analysis yielded evidence for the typicality-based asymmetry. Additionally, we tested different data aggregation approaches (e.g., means per phase, difference scores, means per block) as adopted by the previous studies, yet none altered the overall null result. Third, it is possible that previous contradictory findings of the asymmetry effect were the result of underpowered designs. Neither Dunsmoor and Murphy (2014) nor Wong and Beckers (2021) relied on a priori sample size calculations for their studies. Dunsmoor and Murphy (2014) tested a total of 37 participants (TYP group = 19, ATYP group = 18), while Wong and Beckers (2021) included a larger sample (N = 96). However, the asymmetry effect was observed only in the low-anxiety subgroup (n = 49; resulting in n = 24 per group). In contrast, our study included 37 participants per group and was explicitly powered based on a weighted effect size from both previous studies. Despite this improved power, we failed to observe the asymmetry effect. Finally, the observation that 60% of our sample had anxiety scores outside the cut-offs used by Wong and Beckers (2021), combined with the absence of anxiety data in the Dunsmoor and Murphy (2014) study, leaves the possibility open that differences in the distribution of anxiety scores across samples may explain the inconsistent findings in the literature. While this hypothesis warrants further investigation, we remain cautious about assuming large differences in trait anxiety between our sample and that of Dunsmoor and Murphy, given that both likely relied on WEIRD (Western, Educated, Industrialized, Rich, and Democratic) student populations. Although not powered to test the interaction with anxiety scores, we exploratively assessed the impact of anxiety scores on the potential asymmetry effect. The mixed findings are again difficult to reconcile with the findings of Wong and Beckers (2021). Whereas they found differences between anxiety groups only for the US expectancy ratings but not for SCR responses, we found no effect of anxiety scores in the US expectancy ratings and SCR ttp data but did find an effect for the SCR peak analysis. However, the effect originated from different responses to the control stimulus (GS-) rather than the stimulus of interest the GS+. These findings underscore the importance of systematically assessing individual differences, such as trait anxiety, as potential boundary conditions for conceptual fear generalization effects.

A key aspect of the study involved exploring how individual differences in category representation influence fear generalization processes. Despite observing considerable variability in how participants assessed the typicality of the items (Fig. 3), inclusion of these individual differences as a predictor did not better account for generalization patterns. While we should not discard the possibility that individual differences in concept representation may have little effect on fear generalization patterns, limitations inherent to the typicality rating task could also be responsible for the absence of a relationship between individual typicality judgments and fear generalization patterns. While typicality ratings are a widely used measure of conceptual structure, they are too simplistic to fully capture the richness and complexity of how individuals mentally represent categories. In typicality rating tasks, participants must collapse their multidimensional concept representations into a single dimension of variation (Verheyen et al., 2019; Verheyen & Storms, 2013). It is possible that a more comprehensive account of how conceptual knowledge shapes fear generalization requires methods that are better at capturing the former's multifaceted nature. Future research should consider incorporating measures such as similarity judgments, word associations, or feature listings (De Deyne et al., 2008, 2013; Dry & Storms, 2009) that allow one to build multi-dimensional concept representations that may provide more comprehensive insight into the mechanisms underlying conceptual fear learning.

A surprising finding was the negative relationship between group-averaged typicality ratings and US expectancy across both groups. This pattern runs counter to established models of conceptual fear generalization, which typically predict that conditioned responding increases with exemplar typicality (Dunsmoor & Murphy, 2014, 2015; Visser et al., 2021). In the present study, stimuli perceived as more typical of the conditioned category elicited weaker, rather than stronger, expectancy responses. One possible explanation is that this inverse relationship may reflect processes specific to fear learning, such as the engagement of inhibitory or safety learning mechanisms once highly typical exemplars were identified as non-threatening. Future studies should further investigate the boundary conditions under which conceptual typicality influences fear generalization and examine whether these mechanisms differ from those underlying generalization in non-aversive conceptual learning domains (e.g., Murphy, 2002; Rogers & McClelland, 2004). Beyond methodological refinements, several complementary research directions would further elucidate the nature and boundaries of conceptual fear generalization. Examining conceptual fear generalization in clinically relevant contexts would enhance ecological validity and translational potential. Using more naturalistic stimuli, such as virtual reality environments or narrative contexts, could provide a richer framework for studying how conceptual knowledge influences fear learning in complex situations (Diemer et al., 2015).

In sum, our findings question the purported relationship between conceptual fear generalization and premise typicality effects in category-based induction. Contrary to Dunsmoor and Murphy's (2014) proposal that these phenomena share underlying principles, our results reveal that they do not necessarily align. Our failure to replicate the asymmetry effect aligns with Hampton and Cannon's (2004) skepticism regarding premise typicality's robustness (see also Hayes & Heit, 2018). Despite this humbling outcome, cross-talk between the fear conditioning, semantic cognition, and induction literatures remains essential to determine whether theoretical integration across these domains is warranted.

Author Contributions

JZ and SV conceptualized the study. JZ, KY, and SV drafted the manuscript. AW provided critical revisions. JZ and AW processed the data. JZ and KY analyzed the data. JZ, AW, KY, and SV read and approved the final version of the manuscript.

Data Accessibility Statement

The study is pre-registered on the open science framework (https://osf.io/hpd3s), and all data, analysis code, and the experiment can be found at https://osf.io/48rbp/.

Competing Interests

The authors declare no competing interests.

Editors: Sara Garofalo (Associate Editor)

Submitted: June 04, 2025 PST. Accepted: January 14, 2026 PST. Published: February 06, 2026 PST.

References

Barsalou, L. W. (1987). The Instability of Graded Structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In *Similarity and Analogical Reasoning* (pp. 76–121). Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.006

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). Lawrence Erlbaum Associates.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bennett, M. P., Meulders, A., Baeyens, F., & Vlaeyen, J. W. S. S. (2015). Words putting pain in motion: the generalization of pain-related fear within an artificial stimulus category. *Frontiers in Psychology*, *6*(APR), 1–12. https://doi.org/10.3389/fpsyg.2015.00520

Boyle, S., Roche, B., Dymond, S., & Hermans, D. (2016). Generalisation of fear and avoidance along a semantic continuum. *Cognition and Emotion*, *30*(2), 340–352. https://doi.org/10.1080/02699931.2014.1000831

Breska, A., Maoz, K., & Ben-Shakhar, G. (2011). Interstimulus intervals for skin conductance response measurement. *Psychophysiology*, *48*(4), 437–440. https://doi.org/10.1111/j.1469-8986.2010.01084.x

Damasio, A. R., Tranel, D., & Damasio, H. (1991). Somatic markers and the guidance of behaviour: theory and preliminary testing. In A. L. Levin, H. S. Eisenberg, & H. M. Benton (Eds.), *Frontal Lobe Function and Dysfunction* (pp. 217–229). Oxford University Press. https://doi.org/10.1093/oso/9780195062847.003.0011

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 159–182). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

de Beurs, E., Van Dyck, R., Marquenie, L. A., Lange, A., & Blonk, R. W. B. (2001). De DASS: Een vragenlijst voor het meten van depressie, angst en stress. [The DASS: A questionnaire for the measurement of depression, anxiety, and stress.]. *Gedragstherapie*, *34*(1), 35–53.

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. https://doi.org/10.3758/s13428-012-0260-7

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, *40*, 1030–1048. https://doi.org/10.3758/brm.40.4.1030

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, *6*, Article 26. https://doi.org/10.3389/fpsyg.2015.00026

Dry, M. J., & Storms, G. (2009). Similar but not the same: A comparison of the utility of directly rated and feature-based similarity measures for generating spatial models of conceptual data. *Behavior Research Methods*, *41*(3), 889–900. https://doi.org/10.3758/brm.41.3.889

Dunsmoor, J. E., Kragel, P. A., Martin, A., Labar, K. S., La Bar, K. S., & Labar, K. S. (2014). Aversive learning modulates cortical representations of object categories. *Cerebral Cortex*, *24*(11), 2859–2872. https://doi.org/10.1093/cercor/bht138

Dunsmoor, J. E., Martin, A., & LaBar, K. S. (2012). Role of conceptual knowledge in learning and retention of conditioned fear. *Biological Psychology*, *89*(2), 300–305. https://doi.org/10.1016/j.biopsycho.2011.11.002

Dunsmoor, J. E., & Murphy, G. L. (2014). Stimulus Typicality Determines How Broadly Fear Is Generalized. *Psychological Science*, *25*(9), 1816–1821. https://doi.org/10.1177/0956797614535401

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 and Memory*, *18*(3), 156–160. https://doi.org/10.1101/lm.2016411

Dymond, S., Schlund, M. W., Roche, B., Whelan, R., Richards, J., & Davies, C. (2011). Inferred threat and safety: Symbolic generalization of human avoidance learning. *Behaviour Research and Therapy*, *49*(10), 614–621. https://doi.org/10.1016/j.brat.2011.06.007

Gabora, L., Rosch, E., & Aerts, D. (2008). Toward an Ecological Theory of Concepts. *Ecological Psychology*, *20*(1), 84–116. https://doi.org/10.1080/10407410701766676

Gerdes, A. B. M. M., Fraunfelter, L., & Alpers, G. W. (2020). Hear it, fear it: Fear generalizes from conditioned pictures to semantically related sounds. *Journal of Anxiety Disorders*, *69*(November 2019), 102174. https://doi.org/10.1016/j.janxdis.2019.102174

Hampton, J. A. (2020). Investigating Differences in People's Concept Representations. In *Shifting Concepts* (pp. 67–82). Oxford University Press. https://doi.org/10.1093/oso/9780198803331.003.0005

Hampton, J. A., & Cannon, I. (2004). Category-based induction: An effect of conclusion typicality. *Memory & Cognition*, *32*(2), 235–243. https://doi.org/10.3758/BF03196855

Hayes, B. K., & Heit, E. (2018). Inductive reasoning 2.0. *Wiley Interdisciplinary Reviews: Cognitive Science*, *9*(3), e1459. https://doi.org/10.1002/wcs.1459

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, *7*(4), 569–592. https://doi.org/10.3758/BF03212996

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Kuhn, M., Gerlicher, A. M. V., & Lonsdorf, T. B. (2022). Navigating the manyverse of skin conductance response quantification approaches – A direct comparison of trough-to-peak, baseline correction, and model-based approaches in Ledalab and PsPM. *Psychophysiology*, *59*(9). https://doi.org/10.1111/psyp.14058

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course* (pp. 1–264). Bayesian Cognitive Modeling: A Practical Course. https://doi.org/10.1017/CBO9781139087759

Lei, Y., Wang, J., Dou, H., Qiu, Y., & Li, H. (2019). Influence of typicality in category-based fear generalization: Diverging evidence from the P2 and N400 effect. *International Journal of Psychophysiology*, *135*(May 2018), 12–20. https://doi.org/10.1016/j.ijpsycho.2018.11.002

Lovibond, S. H., & Lovibond, P. H. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney Psychology Foundation.

Malt, B. C. (1995). Category Coherence in Cross-Cultural Perspective. *Cognitive Psychology*, *29*(2), 85–148. https://doi.org/10.1006/cogp.1995.1013

Murphy, G. L. (2002). *The Big Book of Concepts*. The MIT Press. https://doi.org/10.7551/mitpress/1602.001.0001

Osherson, D. N., Smith, E. E., Wilkie, O., ... López, A. (1990). Category-based induction. *Psychological Review*, *97*(2), 185–200. https://doi.org/10.1037/0033-295X.97.2.185

Pecher, D., & Zwaan, R. A. (2017). Flexible concepts: a commentary on Kemmerer (2016). *Language, Cognition and Neuroscience*, *32*(4), 444–446. https://doi.org/10.1080/23273798.2016.1274413

Peterson, R. (1997). A Quantitative Analysis of Rating-scale Response Variability. *Marketing Letters*, *8*, 9–21. https://doi.org/10.1023/A:1007909424993

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 665–681. https://doi.org/10.1016/S0022-5371(75)80055-7

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition*. The MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Rosch, E. (1999). Reclaiming concepts. *Journal of Consciousness Studies*, *6*, 61–77.

Rosch, E. (2011). "Slow lettuce": Categories, concepts, fuzzy sets, and logical deduction. In R. Belohlavek & G. J. Klir (Eds.), *Concepts and Fuzzy Logic* (pp. 89–120). The MIT Press. https://doi.org/10.7551/mitpress/8842.003.0006

Scheveneels, S., Boddez, Y., Bennett, M. P., & Hermans, D. (2017). One for all: The effect of extinction stimulus typicality on return of fear. *Journal of Behavior Therapy and Experimental Psychiatry*, *57*, 37–44. https://doi.org/10.1016/j.jbtep.2017.03.002

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments*. R package version 0.16-1.

Sjouwerman, R., Illius, S., Kuhn, M., & Lonsdorf, T. B. (2022). A data multiverse analysis investigating non-model based SCR quantification approaches. *Psychophysiology*, *August 2021*, 1–27. https://doi.org/10.1111/psyp.14130

Verheyen, S., & Storms, G. (2013). A Mixture Approach to Vagueness and Ambiguity. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0063507

Verheyen, S., White, A., & Égré, P. (2019). Revealing Criterial Vagueness in Inconsistencies. *Open Mind*, *3*, 41–51. https://doi.org/10.1162/opmi_a_00025

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. https://dplyr.tidyverse.org

Wong, A. H. K., & Beckers, T. (2021). Trait anxiety is associated with reduced typicality asymmetry in fear generalization. *Behaviour Research and Therapy*, *138*(December 2020), 103802. https://doi.org/10.1016/j.brat.2021.103802

Yarnitsky, D., Sprecher, E., Zaslansky, R., & Hemli, J. A. (1995). Heat pain thresholds: normative data and repeatability. *Pain*, *60*(3), 329–332. https://doi.org/10.1016/0304-3959(94)00132-X

Supplementary Materials

Supplemental Information

Download: https://collabra.scholasticahq.com/article/155963-bitten-by-the-dog-afraid-of-the-hedgehog-a-registered-report-of-the-asymmetry-effect-in-conceptual-fear-generalization/attachment/327106.docx?auth_token=_c97pbpbk9m-3AkJpvaw

---

Peer Review Communication

Download: https://collabra.scholasticahq.com/article/155963-bitten-by-the-dog-afraid-of-the-hedgehog-a-registered-report-of-the-asymmetry-effect-in-conceptual-fear-generalization/attachment/327107.docx?auth_token=_c97pbpbk9m-3AkJpvaw

---