

# Rule-based generalization of threat without similarity

Lars Marstaller<sup>a,b,\*</sup>, Rizah Al-Jiboury<sup>b</sup>, Andrew H. Kemp<sup>b</sup>, Simon Dymond<sup>b,c</sup>

<sup>a</sup>*Department of Psychology, Bournemouth University, UK*

<sup>b</sup>*Department of Psychology, Swansea University, UK*

<sup>c</sup>*Department of Psychology, Reykjavik University, Iceland*

---

## Abstract

Threat generalization to novel instances is central to adaptive behavior. Most previous work has investigated threat generalization based on the perceptual similarity between past and novel stimuli. Few studies have explored generalization based on abstract, non-perceptual relations despite their importance for cognitive flexibility. In order to measure such rule-based generalization of threat without perceptual similarity, we developed a novel paradigm that prevents perceptual features from gaining predictive value. Our results demonstrate that participants responded according to the correct abstract rule and used it to successfully generalize their anticipatory behavioural threat responses (expectancy ratings, sudomotor nerve activity, and heart rate responses). Our results further show that participants flexibly adapted their responses to an unsigned mid-session contingency reversal. We interpret our results in the context of other rule-based generalization tasks and argue that variations of our paradigm make possible a wide range of investigations into the conceptual aspects of threat generalization.

*Keywords:* conditioning; rule learning; reversal; psychophysiology

---

---

\*Corresponding Author

*Email address:* [lars.marstaller@protonmail.com](mailto:lars.marstaller@protonmail.com) (Lars Marstaller)

## **Acknowledgements**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 663830.

## **Open Science Statement**

All experimental materials, data, and analysis code used in the production of this paper are freely accessible at <https://osf.io/b47ax/>.

## Introduction

Generalization is the expression of a previously learned behavior in response to a novel stimulus. In a dynamically changing world where conditions never truly replicate, generalization of behavioral responses to novel stimuli is necessary for any organism to establish beneficial interactions with its environment. Responding to a novel generalization stimulus (GS) in terms of a previously learned conditioned stimulus (CS) presupposes that GS and CS are equated with respect to their reinforcement contingencies even though the GS has not been reinforced. The behavioral equation of GS and CS with respect to their reinforcement contingencies, can be motivated by perceptual similarity (*e.g.*, Pavlov 1927; Spence 1937; Guttman and Kalish 1956; Honig and Urcuioli 1981) or by an abstract pattern or rule (*e.g.*, Gershman and Niv 2015; Lange et al. 2017; Meulders, Vandael, and Vlaeyen 2017; Lei et al. 2019; Lovibond, Lee, and Hayes 2019; Maes et al. 2017; Vervoort et al. 2014).

Perceptual similarity can be defined as the spatial distance between the GS and CS on a continuously varying psychological dimension, such as, for example, size or wavelength (Shepard 1987), or as the degree of overlap between the feature sets of the GS and the CS (Tversky 1977; for a more general Bayesian account, see Tenenbaum and Griffiths 2001). Experiments have shown that behavioral and physiological responses are sensitive to the degree of similarity and generally decrease as the GS increasingly differs perceptually from the CS, thus forming a generalization gradient (Spence 1937; Lissek et al. 2008).

In contrast to similarity-based generalization, rule-based generalization requires the organism to infer an abstract pattern that specifies when a GS should be behaviorally equated with a CS. In a paradigmatic study on rule-based generalization, Shanks and Darby (1998) showed that humans who have been trained on negative (A+, B+, AB-) and positive (C-, D-, CD+) association patterns, generalized the learned rule of opposites (reinforcement for combined stimuli is opposite to singular stimuli) to novel stimulus patterns of the same type (E+, F+, EF?, correctly predicting EF-; GH+, G?, H? correctly predicting G- and H-).

The main difference between similarity and rule-based generalization is that the latter requires some degree of abstraction from the identity of the stimulus and its concrete perceptual features, that is, inductive inference (Dunsmoor and Murphy 2015). The abstraction required for rule-based generalization in the patterning task makes it possible to compare similarity-based and rule-based generalization (Shanks and Darby 1998). However, the patterning task is focused exclusively on the rule of opposites, which might be a particular type of abstraction that requires other specific cognitive abilities, such as working memory or attention (Maes et al. 2015, 2017).

A simpler task to investigate rule-based generalization is based on the same/different distinction. In the match-to-sample (MTS) task, participants are first presented with a sample stimulus and then have to select the target stimulus that matches the sample stimulus from a range of comparison stimuli. Generalization in MTS is constituted by the selection of a target stimulus based

on a previously reinforced matching relation between sample and target (*e.g.*, color of elemental stimuli, or visual variance of compound stimuli; Urcuioli and Nevin 1975). In its simpler form, MTS is problematic because successful performance in MTS generalization can be achieved through perceptual, that is, similarity-based, generalization using stimulus features such as visual entropy to assess visual variance (Young, Wasserman, and Garner 1997).

To address this issue with simple MTS tasks and to investigate rule-based generalization without similarity beyond the rule of opposites, we designed a novel paradigm that uses a relative - and therefore minimally abstract - property of compound stimuli to define sameness. The basic structure of our paradigm is as follows: on each trial, participants are consecutively presented with four different images, which together constitute a complex relational stimulus that acts as the CS. The first three images are all instances of the same semantic category, for example, three images of islands. The fourth image is either another member of that category (*e.g.*, another image of an island) or not. The rule in this experiment is defined as a mapping from the relative perceptual properties of the four images to an abstract binary property of the CS (same/different). Reinforcement is contingent upon this abstract matching property of the CS, that is, depending on the CS-US contingency assignment, the CS is reinforced if the fourth stimulus is the same (CS+) but not otherwise (CS-). Importantly, all stimuli used in the experiment are trial-unique and each stimulus is presented only once during the experiment.

We argue that this paradigm requires rule-based rather than similarity-based generalization because the GS-CS relation is dependent on the abstract matching property but not any perceptual features of the stimuli. Perceptual similarity is only relevant to derive the matching property but is otherwise ‘encapsulated’ within each trial. The transfer of a previously learned CS-US association onto a novel GS is independent of the perceptual similarity between the GS and any other CS. There are no perceptual properties, for example, associated with islands, that will allow the learner to generalize to novel stimuli, which depict, for example, airports.

One feature of abstraction is that it affords a high degree of flexibility. Commonly, flexibility is experimentally assessed using reversal tasks, where reinforcement contingencies are reversed after a certain number of trials and whatever stimulus acted as CS+ now becomes the CS- and vice versa (Clark, Cools, and Robbins 2004; Izquierdo et al. 2017). Rules, such as the one used in our paradigm, afford flexibility because the relation between the matching property of the CS and the US can simply be reversed upon encountering new evidence without the need to learn a novel abstraction. To test whether rule-based generalization in our paradigm is associated with behavioral flexibility, participants undergo a mid-session reversal of reinforcement contingencies.

As behavioral measures of learning, we recorded skin conductance responses (modeled as sudomotor nerve activity, SNA), stimulus-related changes in heart rate (heart rate responses, HRR), and online US expectancy ratings. Our main hypothesis is that participants will demonstrate differential, rule-based generalization across all three measures, that is, stronger responses to CS+ than

CS- irrespective of experimental phase. We further predicted that participants would flexibly adapt their behavioral responses after the contingency reversal, that is, differential responding to CS+ and CS- would be maintained post contingency reversal.

## Materials and Methods

### *Participants*

53 right-handed young adults (age  $M = 22.00$  years,  $SD = 3.41$  years, range = 19-39 years; 24 females) took part in the experiment after giving written consent in exchange for course credits or £10 in Amazon vouchers. The study was approved by the Research Ethics Committee at Swansea University's Department of Psychology.

### *Procedure and Materials*

Participants took part in a novel differential threat generalization task. During each trial, participants were consecutively presented with four images. The first three images showed exemplars from a single semantic category (*e.g.*, pictures of islands). The fourth image either did match the semantic category (*e.g.*, was also a picture of an island) or did not match the category (*e.g.*, was a picture of a fire station).

The first three images were each presented for 1.5 sec followed by a fixation cross for 0.5 sec. The fourth image was presented for 7 sec and followed by a blank screen for 10 sec. During presentation of the fourth image, participants were asked to predict the likelihood that the stimulus would be followed by the US. To record these online expectancy ratings, a sliding scale was presented for 5 sec together with the fourth image and participants used the left and right keyboard buttons to move an indicator, which was initially centered on the scale. The scale from 0 to 100 was marked 'certain no shock' at point 0 via 'uncertain' at point 50 to 'certain shock' at point 100 and rating proceeded in 20 steps of 5 (arbitrary) unmarked units (see Figure 1).

Participants completed a total of 80 trials and were not instructed about the CS-US contingencies or the mid-session reversal. Participants were instructed that some of the sequences of four pictures would be followed by a brief shock and that it would be their task to predict at the end of each sequence how likely the shock would occur. The assignment of matching or non-matching stimulus sequences as CS+ and CS- was randomized and counter-balanced across participants and reversed after half the trials when participants were offered a short break. That is, participants saw 20 CS1+ and 20 CS2- in the first and 20 CS1- and 20 CS2+ in the second half of the experiment. The assignment of pictures to trials was randomized within subjects as was the order of trial types with the constraint that there were not more than two consecutive trials of the same type. During each phase of the experiment, 15 of the 20 CS+ presentations, that is, 75%, were reinforced at stimulus offset. The order of CS+ reinforcements

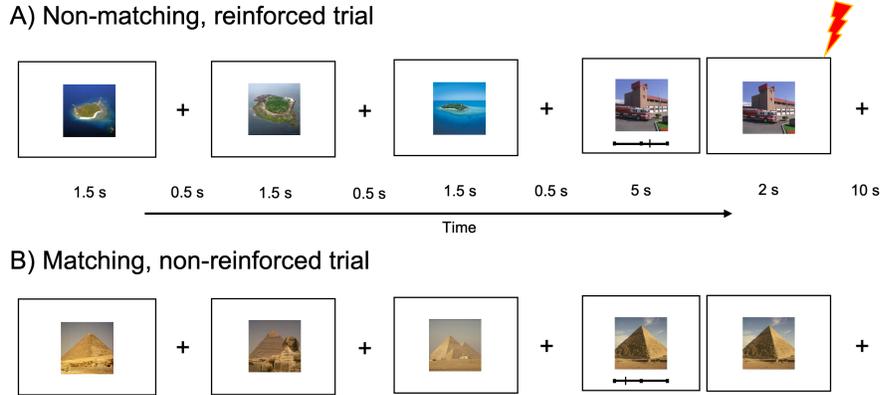


Figure 1: Experimental Paradigm: In each trial, three images of the same category are presented and followed by a fourth image that either matches that category or not. During the first 5 sec of the fourth image, a US expectancy scale is presented. The fourth image co-terminates with the US depending on its category membership.

was randomized within subjects but the first CS+ presentation of each phase was always reinforced.

Pictures of real-life scenes were used as stimuli. Eighty sets of four pictures as well as a set of 40 unrelated pictures were selected from the ‘Massive Memory’ scene categories stimulus set (Konkle et al. 2010). Non-matching stimuli were randomly selected pictures from the set of unrelated pictures. The experiment was programmed and presented using Presentation (v.20, Neurobehavioural Systems Ltd, <https://www.neurobs.com/>).

Transcutaneous electrical stimulation to the outside wrist of the dominant hand served as US. Stimulation consisted of five 20 msec bursts of electrical current each spaced 20 msec apart for a total US duration of 180 msec. Stimulation was delivered using a bar electrode (Biopac EL351) and a constant voltage stimulator (Biopac STM200). The voltage of the electrical stimulation was set prior to the experiment using a stepwise procedure, which allows the voltage to be adjusted to each participant’s individual tolerance level so it is perceived as “uncomfortable but not painful” (LaBar et al. 1998). On average, US strength was 40.85 V (SD = 7.8 V). The average aversiveness rating of the US strength was 4.91 (SD = 2.2) on a scale from 0 (‘not at all painful’) to 10 (‘extremely painful’).

Missing individual mean estimates were imputed on the group level using multivariate imputation by chained equations as implemented in the ‘mice’ package (v.3.4) for R using predictive mean matching with 10 iterations of 10 imputations (Buuren and Groothuis-Oudshoorn 2011; Azur et al. 2011). This imputation procedure is based on the assumption that values are missing at random and preserves the relations in the data as well as the uncertainty about

these relations (Buuren and Groothuis-Oudshoorn 2011). In total 3.77% of expectancy ratings, 11.32% of electrodermal data, and 22.64% of electrocardiographic data were imputed. All statistical analyses were conducted using R statistical language (v.4.0.0; <https://www.r-project.org/>) in RStudio (v.1.2.1335; <https://rstudio.com/>).

### *Psychophysiological Recording and Analysis*

During the task, participants' electrodermal activity and electrocardiogram were recorded and digitized at a sampling rate of 2 kHz with 16 bit resolution using a Biopac MP160 system and Acknowledge 5 software (<https://www.biopac.com/>). Electrodermal activity was measured on the distal phalanges of the index and middle fingers of the participant's non-dominant hand using a pair of 6 mm Ag-AgCl electrodes (Biopac TSD203), amplified using a Biopac EDA100C amplifier, and filtered at acquisition using a 0.1 Hz low-pass hardware filter. The electrocardiogram was recorded in a three-electrode lead-2 configuration with two shielded 8 mm Ag-AgCl electrodes (Biopac EL258S) attached to the participant's inner upper arm (positive left, negative right) and one unshielded 8 mm Ag-AgCl electrode (Biopac EL258) attached to the participant's right ankle bone (ground). Electrocardiographic data were amplified using a Biopac ECG100C amplifier and filtered at acquisition using a 0.05 Hz high-pass and a 35 Hz low-pass hardware filter.

All psychophysiological responses were computed for the fourth image stimulus in each trial. Psychophysiological data were downsampled to 1 kHz and analyzed using a custom python (v.3.6, <https://www.python.org/>) script using the Neurokit module (<https://neurokit.readthedocs.io>). The electrodermal data were analyzed using the cvxEDA module (<https://github.com/lciti/cvxEDA>). cvxEDA uses a convex optimization approach to fit a model of the skin conductance response to the data and to derive estimates for phasic and tonic components of sudomotor nerve activity (SNA) as well as noise (Greco et al. 2016). The advantage of model-based analysis approach is that it extends purely operational approaches and defines an invertible causal relationship between the underlying psychological state (arousal) and the physiological response (electrodermal activity) beyond correlations (Bach and Friston 2013). The inversion of a causal model then gives an estimate of the psychological state rather than its physiological correlate. As an indicator of arousal, we calculated the maximum phasic SNA model estimate in the time window 1-4 sec post stimulus onset.

Cardiac cycles were detected using the approach described in Gamboa (2008). This algorithm first normalizes the electrocardiographic signal and then identifies the R peaks as zeros in the first derivative of the signal. Finally, the algorithm recursively removes R peaks with an interbeat interval smaller than 0.3 sec (200 bpm) or larger than 3 sec (20 bpm; Gamboa 2008). For the analysis of event-related cardiac responses, the difference between the mean R-R intervals two seconds pre stimulus onset (overlapping with the third stimulus) and four seconds post stimulus onset was calculated and converted to heart rate to give the heart rate response (HRR).

While SNA is an indicator of sympathetic arousal, HRRs are the result of innervation from both branches of the autonomic nervous system and therefore provide additional evidence about the levels of parasympathetic, vagus nerve activity. As such, HRR can provide positive evidence for safety learning whereas SNA can only provide negative evidence, that is, the absence of threat responding as indicated by lower arousal levels. The combination of SNA and HRR therefore allows one to assess threat as well as safety learning and to assess the effects of contingency awareness on threat generalization.

## Results

### *Expectancy Ratings*

Mean expectancy ratings were analyzed using 2x2 repeated measures analyses of variance (ANOVA) with the factors stimulus type (CS+, CS-) and experimental phase (pre, post reversal). The results demonstrate a significant interaction between stimulus type and experimental phase ( $F(1,52) = 4.12, p = 0.048, \eta_G^2 = 0.02$ ) as well as a main effect of stimulus type ( $F(1,52) = 96.21, p < 0.001, \eta_G^2 = 0.53$ ; see Figures 2 and 3A). The main effect of experimental phase was not significant ( $F(1,52) = 1.13, p = 0.292$ ). The interaction showed that expectancy ratings increased post reversal for the CS+ ( $t(52) = -2.22, p = 0.031, d = 0.33$ ) but not the CS- ( $t(52) = 1.50, p = 0.139, d = 0.20$ ). The main effect of stimulus type shows that participants' expectancy ratings were significantly stronger in response to CS+ ( $M = 68.8, SD = 9.4$ ) than in response to CS- ( $M = 28.6, SD = 8.0$ ). The results of this ANOVA of the imputed data are consistent with the results from the same analysis of 51 non-imputed data, which resulted in the same effects, *i.e.*, a significant interaction between stimulus type and experimental phase as well as a significant main effect of stimulus type.

### *Psychophysiology*

In order to investigate learning-related arousal, SNA and HRR were analyzed using a 2x2 repeated measures ANOVA with the factors stimulus type (CS+, CS-) and experimental phase (pre, post reversal). For SNA, the results demonstrate a significant main effect of stimulus type ( $F(1,52) = 13.67, p < 0.001, \eta_G^2 = 0.04$ ; see Figure 3B). The main effect of experimental phase ( $F(1,52) = 4.75, p = 0.034$ ) and the interaction between stimulus type and experimental phase were not significant ( $F(1,52) = 0.03, p = 0.872$ ). The main effect of stimulus type shows that participants' phasic SNA was significantly stronger in response to the CS+ ( $M = 0.27, SD = 0.14$ ) than in response to the CS- ( $M = 0.22, SD = 0.13$ ) irrespective of experimental phase. The results of this ANOVA are consistent with the results from the same ANOVA of 47 non-imputed data points, which resulted in a significant main effect of stimulus type. However, in addition to the main effect of stimulus type, the ANOVA of non-imputed data additionally resulted in a significant main effect of phase ( $F(1,46) = 4.44, p = 0.041, \eta_G^2 = 0.01$ ), which demonstrated larger responses pre ( $M = 0.26, SD = 0.14$ ) than post reversal ( $M = 0.23, SD = 0.14$ ).

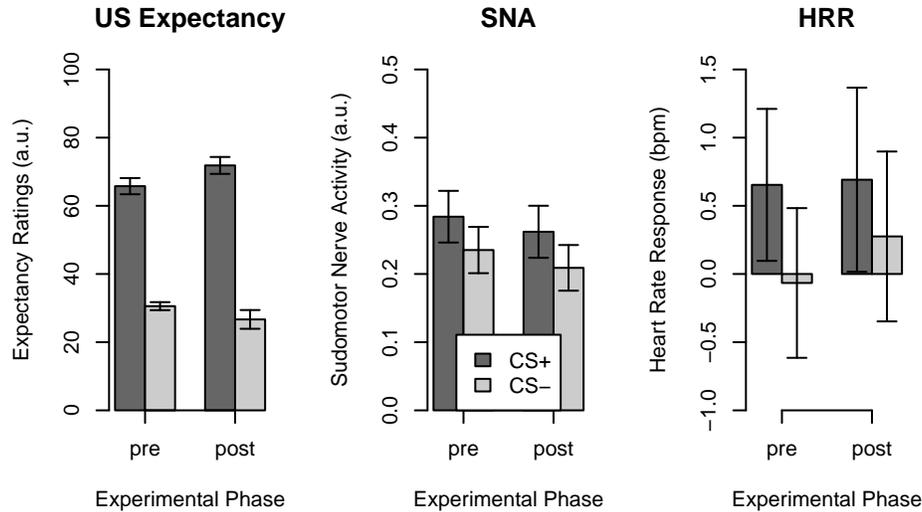


Figure 2: Mean (+/- 95% CI) behavioural and psychophysiological responses to conditioned stimuli pre and post contingency reversal (means of imputed data). Note that the CS+ pre reversal is the CS- post reversal and vice versa. Left: US Expectancy ratings. Center: Skin conductance responses (sudomotor nerve activity). Right: Changes in heart rate compared to pre-stimulus baseline (heart rate responses). [SNA = sudomotor nerve activity; HRR = heart rate response]

For HRR, the results show a significant main effect of stimulus type ( $F(1,52) = 8.32, p = 0.006, \eta_G^2 = 0.02$ ; see Figure 3C). The the main effect of experimental phase ( $F(1,52) = 0.89, p = 0.349$ ) and interaction between stimulus type and experimental phase ( $F(1,52) = 0.67, p = 0.418$ ) were not significant. The main effect of stimulus type shows that participants' heart rates generally increased more strongly in response to CS+ ( $M = 0.67, SD = 2.29$ ) than in response to CS- ( $M = 0.10, SD = 2.18$ ). The results of this ANOVA are not consistent with the results of the same ANOVA of 38 non-imputed data points, which showed no significant effects (all  $p > 0.1$ ). However, the non-imputed data do show the same pattern with stronger responses to CS+ before ( $M = 0.64, SD = 2.06$ ) and after reversal ( $M = 0.68, SD = 2.67$ ) than to CS- before ( $M = 0.11, SD = 2.11$ ) and after reversal ( $M = 0.49, SD = 2.26$ ).

## Discussion

The primary aim of this study was to investigate anticipatory behavioral responses (US expectancy ratings, sudomotor nerve activity, and heart rate responses) during a novel rule-based threat generalization paradigm. All three measures indicated successful learning by showing significantly stronger responses to CS+ than CS-. The secondary aim of this study was to test the effects of mid-session contingency reversal on behavior. The results demonstrated no effect of phase on US expectancy ratings, sudomotor nerve activity, or heart

rate responses and hence show successful behavioral adaptation to contingency reversal. For US expectancy ratings, the results even provided evidence for a post reversal increase in differential responding.

While all three measures converged towards the same general pattern of results, there were subtle differences. The main effect (generalized  $\eta^2$ ) of stimulus type was more than ten times larger for expectancy ratings compared to psychophysiological responses and mostly due to differences in CS- responses. Given that the task required rule-learning, this difference in effect size might possibly indicate that participants were consciously aware when a stimulus pattern signaled safety but emotionally perceived CS- as potential threat and unconsciously mobilized defensive responses (Ohman and Mineka 2001; Hamm and Weike 2005). However, additional research is required to confirm this speculation. Within psychophysiological responses, sudomotor nerve activity and heart rate responses also differed from each other. Sudomotor nerve activity was overall lower post reversal, which might be indicative of habituation (Boucsein et al. 2012). In contrast, the results indicate that heart rate responses were higher in response to CS- post compared to pre reversal, which might suggest that heart rate responses were resistant to contingency reversal as the post reversal CS- is the same as the pre reversal CS+.

The results of the present study further demonstrate threat conditioned tachycardia, which stands in contrast to widely reported threat conditioned bradycardia (*e.g.*, Hugdahl 1979; Castegnetti et al. 2016). Several studies have shown that tachycardia but not bradycardia is associated with emotional threat responding whereas both, tachycardia and bradycardia have been associated with contingency awareness (Hodes, Cook, and Lang 1985; Hamm and Vaitl 1996; López et al. 2009). Increased tachycardia in response to CS- post reversal therefore suggests a resistance of emotional responding to reversal. Together, the differences between the three measures suggest that the task might be able to differentiate between the two aspects of threat responding, contingency awareness and emotional response. While further evidence is required to investigate this aspect of our paradigm, our results highlight the benefits of using different outcome measures in threat conditioning research.

The more general purpose of this experiment was to test whether this novel paradigm would assess rule-based generalization without similarity beyond the rule of opposites. While the results clearly demonstrate the effectiveness of the paradigm, we argue that the paradigm measures rule-based generalization because several features of its design make it impossible to explain the behavior observed in our experiment within an associative framework. First, generalization is based on a relative rather than absolute property of the compound stimulus and no perceptual feature by itself can gain predictive value. Second, the use of complex scenes as trial-unique stimuli increases the number and variability of perceptual features and reduces the probability that any complex feature (*e.g.*, entropy) can acquire predictive value (for a threat generalization study based on perceptual similarity between natural scene stimuli, see Park, Lee, and Lee 2018). Third, prototypical perceptual-semantic features (Dunsmoor and Murphy 2014) are prevented from acquiring predictive value because categories for each

CS are trial-unique.

We assume that these three design features of our paradigm ensure that participants have to engage in rule-based generalization. By maintaining these design features and defining different relative patterns, a large number of variations of the paradigm can be constructed. For example, if A, B, and C are exemplars drawn from different semantic categories, then these patterns can be used to define the abstract rule: ABAB vs ABBA, or ABCC and AABC vs ABAC and ABBC, or ABC vs ABB, instead of AAAA vs AAAB used in the current experiment. In addition, it is possible to contrast rule-based and similarity-based generalization in our paradigm similar to the patterning task (Shanks and Darby 1998) by controlling the degree of similarity between two consecutive CS (*e.g.*, using perceptually similar stimuli A and C in two opposing patterns, such as AAAB+ followed by CCCC-).

It is important to note that our paradigm did not contain separate conditioning and generalization phases in contrast to many other studies (*e.g.*, Dunsmoor et al. 2017; Lissek et al. 2008; Vervliet et al. 2010; Wong and Lovibond 2017). The effect is that each new CS constitutes a GS and reinforcement of the CS can be understood as feedback on generalization. Continual reinforcement during generalization reduces extinction of generalization (Vervliet and Geens 2014) and might even have enhanced the acquisition of the abstract pattern, yet additional studies are necessary to investigate this possibility. Intriguingly, however, the absence of reinforcement during generalization in other paradigms could possibly explain why not all participants in such experiments generalized based on the correct abstract pattern (see, *e.g.*, Shanks and Darby 1998; Maes et al. 2017).

Finally, in contrast to other generalization studies, our paradigm included an un signaled mid-session contingency reversal, such that the CS+ became the CS- and vice versa after half the trials. The results do not show any significant main effects of phase, which suggests that participants reversed their anticipatory behavioral responses and hence, implies that rule-based generalization affords high behavioral flexibility. In line with this interpretation, the results suggest that participants switched their anticipatory behavioral responses without having to extinguish existing and acquire novel CS-US associations. However, additional evidence directly contrasting the effects of contingency reversal on rule-based and similarity-based generalization is necessary to further interpret this finding and its implications.

One limitation of this study is related to the data imputation procedure, which is based on the assumption that data are missing at random (Buuren and Groothuis-Oudshoorn 2011). This assumption has not been validated by the authors and as a consequence the possibility cannot be excluded that non-random factors affected the psychophysiological data. Given that the analysis results of the imputed and non-imputed data are generally consistent, it seems highly likely that the general pattern of results reported here are due to any such non-random factors. We follow recommendations from the recent literature on how to deal with missing data in fear conditioning research and have published all data and scripts used in this study (Lonsdorf et al. 2019).

Taken together, the results provide evidence that participants in our experi-

ment were able to infer the correct abstract matching rule, use it to generalize their threat responses, and to flexibly adapt their responses following mid-session contingency reversal. The results support the view that the evidence base for implicit threat conditioning in humans is considerably weaker and explicit processes much more important than previously assumed (De Houwer 2019; Mertens and Engelhard 2020). Overall, our findings support the view that learning of an abstract rule through inductive inference plays an important role in threat generalization and our paradigm provides a starting point to further explore rule-based threat generalization (Dunsmoor, Martin, and LaBar 2012; Dunsmoor and Murphy 2014; Lovibond, Lee, and Hayes 2019).

## References

- Azur, Melissa J, Elizabeth A Stuart, Constantine Frangakis, and Philip J Leaf. 2011. "Multiple Imputation by Chained Equations: What Is It and How Does It Work?" *Int J Methods Psychiatr Res* 20 (1): 40–49. <https://doi.org/10.1002/mpr.329>.
- Bach, Dominik R, and Karl J Friston. 2013. "Model-Based Analysis of Skin Conductance Responses: Towards Causal Models in Psychophysiology." *Psychophysiology* 50 (1): 15–22. <https://doi.org/10.1111/j.1469-8986.2012.01483.x>.
- Boucsein, Wolfram, Don C Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T roth, Michael E Dawson, Diane L Filion, and Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. 2012. "Publication Recommendations for Electrodermal Measurements." *Psychophysiology* 49 (8): 1017–34. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>.
- Buuren, S van, and K Groothuis-Oudshoorn. 2011. "MICE: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (3): 1–67.
- Castegnetti, Giuseppe, Athina Tzovara, Matthias Staib, Philipp C Paulus, Nicolas Hofer, and Dominik R Bach. 2016. "Modeling Fear-Conditioned Bradycardia in Humans." *Psychophysiology* 53 (6): 930–9. <https://doi.org/10.1111/psyp.12637>.
- Clark, L, R Cools, and T W Robbins. 2004. "The Neuropsychology of Ventral Prefrontal Cortex: Decision-Making and Reversal Learning." *Brain Cogn* 55 (1): 41–53. [https://doi.org/10.1016/S0278-2626\(03\)00284-7](https://doi.org/10.1016/S0278-2626(03)00284-7).
- De Houwer, Jan. 2019. "Moving Beyond System 1 and System 2." *Exp Psychol* 66 (4): 257–65. <https://doi.org/10.1027/1618-3169/a000450>.
- Dunsmoor, Joseph E, Marijn C W Kroes, Stephen H Braren, and Elizabeth A Phelps. 2017. "Threat Intensity Widens Fear Generalization Gradients." *Behav Neurosci* 131 (2): 168–75. <https://doi.org/10.1037/bne0000186>.
- Dunsmoor, Joseph E, Alex Martin, and Kevin S LaBar. 2012. "Role of Conceptual Knowledge in Learning and Retention of Conditioned Fear." *Biol Psychol* 89 (2): 300–305. <https://doi.org/10.1016/j.biopsycho.2011.11.002>.

- Dunsmoor, Joseph E, and Gregory L Murphy. 2014. "Stimulus Typicality Determines How Broadly Fear Is Generalized." *Psychol Sci* 25 (9): 1816–21. <https://doi.org/10.1177/0956797614535401>.
- . 2015. "Categories, Concepts, and Conditioning: How Humans Generalize Fear." *Trends Cogn Sci* 19 (2): 73–77. <https://doi.org/10.1016/j.tics.2014.12.003>.
- Gamboa, Hugo. 2008. "Multi-Modal Behaviour Biometrics Based on Hci and Electrophysiology." PhD thesis, Universidade Tecnica de Lisboa.
- Gershman, Samuel J, and Yael Niv. 2015. "Novelty and Inductive Generalization in Human Reinforcement Learning." *Top Cogn Sci* 7 (3): 391–415. <https://doi.org/10.1111/tops.12138>.
- Greco, Alberto, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. 2016. "CvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing." *IEEE Trans Biomed Eng* 63 (4): 797–804. <https://doi.org/10.1109/TBME.2015.2474131>.
- Guttman, Norman, and Harry I. Kalish. 1956. "Discriminability and Stimulus Generalization." *Journal of Experimental Psychology* 51 (1): 79–88.
- Hamm, Alfons O, and Almut I Weike. 2005. "The Neuropsychology of Fear Learning and Fear Regulation." *Int J Psychophysiol* 57 (1): 5–14. <https://doi.org/10.1016/j.ijpsycho.2005.01.006>.
- Hamm, A O, and D Vaitl. 1996. "Affective Learning: Awareness and Aversion." *Psychophysiology* 33 (6): 698–710. <https://doi.org/10.1111/j.1469-8986.1996.tb02366.x>.
- Hodes, R L, E W Cook 3rd, and P J Lang. 1985. "Individual Differences in Autonomic Response: Conditioned Association or Conditioned Fear?" *Psychophysiology* 22 (5): 545–60. <https://doi.org/10.1111/j.1469-8986.1985.tb01649.x>.
- Honig, W K, and P J Urcuioli. 1981. "The Legacy of Guttman and Kalish (1956): Twenty-Five Years of Research on Stimulus Generalization." *J Exp Anal Behav* 36 (3): 405–45. <https://doi.org/10.1901/jeab.1981.36-405>.
- Hugdahl, K. 1979. "Direction of Changes in the Cardiac Component in a Pavlovian Conditioning Paradigm with Variations in Cs- and Ucs-Contents." *Biol Psychol* 9 (2): 91–102. [https://doi.org/10.1016/0301-0511\(79\)90056-5](https://doi.org/10.1016/0301-0511(79)90056-5).
- Izquierdo, A, J L Brigman, A K Radke, P H Rudebeck, and A Holmes. 2017. "The Neural Basis of Reversal Learning: An Updated Perspective." *Neuroscience* 345 (March): 12–26. <https://doi.org/10.1016/j.neuroscience.2016.03.021>.
- Konkle, Talia, Timothy F Brady, George A Alvarez, and Aude Oliva. 2010. "Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects." *J Exp Psychol Gen* 139 (3): 558–78. <https://doi.org/10.1037/a0019165>.
- LaBar, K S, J C Gatenby, J C Gore, J E LeDoux, and E A Phelps. 1998. "Human Amygdala Activation During Conditioned Fear Acquisition and Extinction: A Mixed-Trial fMRI Study." *Neuron* 20 (5): 937–45. [https://doi.org/10.1016/s0896-6273\(00\)80475-4](https://doi.org/10.1016/s0896-6273(00)80475-4).
- Lange, Iris, Liesbet Goossens, Stijn Michiels, Jindra Bakker, Shmuel Lissek, Silvia Papalini, Simone Verhagen, et al. 2017. "Behavioral Pattern Separation

and Its Link to the Neural Mechanisms of Fear Generalization.” *Soc Cogn Affect Neurosci* 12 (11): 1720–9. <https://doi.org/10.1093/scan/nsx104>.

Lei, Yi, Jinxia Wang, Haoran Dou, Yiwen Qiu, and Hong Li. 2019. “Influence of Typicality in Category-Based Fear Generalization: Diverging Evidence from the P2 and N400 Effect.” *Int J Psychophysiol* 135 (January): 12–20. <https://doi.org/10.1016/j.ijpsycho.2018.11.002>.

Lissek, Shmuel, Arter L Biggs, Stephanie J Rabin, Brian R Cornwell, Ruben P Alvarez, Daniel S Pine, and Christian Grillon. 2008. “Generalization of Conditioned Fear-Potentiated Startle in Humans: Experimental Validation and Clinical Relevance.” *Behav Res Ther* 46 (5): 678–87. <https://doi.org/10.1016/j.brat.2008.02.005>.

Lonsdorf, Tina B, Maren Klingelhöfer-Jens, Marta Andreatta, Tom Beckers, Anastasia Chalkia, Anna Gerlicher, Valerie L Jentsch, et al. 2019. “Navigating the Garden of Forking Paths for Data Exclusions in Fear Conditioning Research.” *Elife* 8 (December). <https://doi.org/10.7554/eLife.52465>.

Lovibond, Peter F, Jessica C Lee, and Brett K Hayes. 2019. “Stimulus Discriminability and Induction as Independent Components of Generalization.” *J Exp Psychol Learn Mem Cogn*, October. <https://doi.org/10.1037/xlm0000779>.

López, Raúl, Rosario Poy, M Carmen Pastor, Pilar Segarra, and Javier Moltó. 2009. “Cardiac Defense Response as a Predictor of Fear Learning.” *Int J Psychophysiol* 74 (3): 229–35. <https://doi.org/10.1016/j.ijpsycho.2009.09.006>.

Maes, Elisa, Guido De Filippo, Angus B Inkster, Stephen E G Lea, Jan De Houwer, Rudi D’Hooge, Tom Beckers, and Andy J Wills. 2015. “Feature- Versus Rule-Based Generalization in Rats, Pigeons and Humans.” *Anim Cogn* 18 (6): 1267–84. <https://doi.org/10.1007/s10071-015-0895-8>.

Maes, Elisa, Elias Vanderoost, Rudi D’Hooge, Jan De Houwer, and Tom Beckers. 2017. “Individual Difference Factors in the Learning and Transfer of Patterning Discriminations.” *Front Psychol* 8: 1262. <https://doi.org/10.3389/fpsyg.2017.01262>.

Mertens, Gaëtan, and Iris M Engelhard. 2020. “A Systematic Review and Meta-Analysis of the Evidence for Unaware Fear Conditioning.” *Neurosci Biobehav Rev* 108 (January): 254–68. <https://doi.org/10.1016/j.neubiorev.2019.11.012>.

Meulders, Ann, Kristof Vandaël, and Johan W S Vlaeyen. 2017. “Generalization of Pain-Related Fear Based on Conceptual Knowledge.” *Behav Ther* 48 (3): 295–310. <https://doi.org/10.1016/j.beth.2016.11.014>.

Ohman, A, and S Mineka. 2001. “Fears, Phobias, and Preparedness: Toward an Evolved Module of Fear and Fear Learning.” *Psychol Rev* 108 (3): 483–522. <https://doi.org/10.1037/0033-295x.108.3.483>.

Park, Doyoung, Hwa-Jin Lee, and Sue-Hyun Lee. 2018. “Generalization of Conscious Fear Is Positively Correlated with Anxiety, but Not with Depression.” *Exp Neurobiol* 27 (1): 34–44.

Pavlov, Ivan. 1927. *Conditioned Reflexes*. London: Oxford Press.

Shanks, D. R., and R. J. Darby. 1998. “Feature- and Rule-Based Generalization in Human Associative Learning.” *Journal of Experimental Psychology: Animal Behavior Processes* 24 (4): 405–15.

- Shepard, R. N. 1987. "Toward a Universal Law of Generalization for Psychological Science." *Science* 237: 1317–23.
- Spence, Kenneth W. 1937. "The Differential Response in Animals to Stimuli Varying Within a Single Dimension." *Psychological Review* 44: 430–44.
- Tenenbaum, J B, and T L Griffiths. 2001. "Generalization, Similarity, and Bayesian Inference." *Behav Brain Sci* 24 (4): 629–40; discussion 652–791. <https://doi.org/10.1017/s0140525x01000061>.
- Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84 (4): 327–52.
- Urcuioli, P J, and J A Nevin. 1975. "Transfer of Hue Matching in Pigeons." *J Exp Anal Behav* 24 (2): 149–55. <https://doi.org/10.1901/jeab.1975.24-149>.
- Vervliet, Bram, and Maarten Geens. 2014. "Fear Generalization in Humans: Impact of Feature Learning on Conditioning and Extinction." *Neurobiol Learn Mem* 113 (September): 143–8. <https://doi.org/10.1016/j.nlm.2013.10.002>.
- Vervliet, Bram, Merel Kindt, Debora Vansteenwegen, and Dirk Hermans. 2010. "Fear Generalization in Humans: Impact of Prior Non-Fearful Experiences." *Behav Res Ther* 48 (11): 1078–84. <https://doi.org/10.1016/j.brat.2010.07.002>.
- Vervoort, Ellen, Bram Vervliet, Marc Bennett, and Frank Baeyens. 2014. "Generalization of Human Fear Acquisition and Extinction Within a Novel Arbitrary Stimulus Category." *PLoS One* 9 (5): e96569. <https://doi.org/10.1371/journal.pone.0096569>.
- Wong, Alex H K, and Peter F Lovibond. 2017. "Rule-Based Generalisation in Single-Cue and Differential Fear Conditioning in Humans." *Biol Psychol* 129 (October): 111–20. <https://doi.org/10.1016/j.biopsycho.2017.08.056>.
- Young, M E, E A Wasserman, and K L Garner. 1997. "Effects of Number of Items on the Pigeon's Discrimination of Same from Different Visual Displays." *J Exp Psychol Anim Behav Process* 23 (4): 491–501. <https://doi.org/10.1037/0097-7403.23.4.491>.