Model-Based Fixation-Pattern Similarity Analysis Reveals Adaptive Changes in Face-Viewing Strategies Following Aversive Learning

Lea Kampermann[1], Niklas Wilming[2], Arjen Alink[1], Christian Büchel[1], Selim Onat[1, *]

[1] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
[2] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

* Corresponding Author
Email: s.onat@uke.de

Short Title:

Model-based Fixation-Pattern Similarity Analysis

Abstract

Learning to associate an event with an aversive outcome typically leads to generalization when similar situations are encountered. In real-world situations, generalization must be based on the sensory evidence collected through active exploration, which in turn can also be influenced by aversive learning. However, we currently do not know how far exploration strategies can be shaped by learning and whether or not learning results in adaptive changes during the course of ensuing generalization. Here, we investigated learning-induced changes in eye-movement patterns using a similarity-based multivariate fixation-pattern analysis together with a set of parametrically controlled stimuli. Humans learnt to associate an aversive outcome (a mild electric shock) with one face along a circular perceptual continuum, whereas the most dissimilar face on this continuum was kept neutral. Before learning, eye-movement patterns mirrored the similarity characteristics of the stimulus continuum, indicating that exploration was mainly guided by subtle physical differences between the faces. Aversive learning resulted in a global increase in dissimilarity of eye movement patterns during generalization. Model-based analysis of the similarity geometry indicated that this was specifically driven by a separation of patterns along the adversity gradient, defined between the reinforced and neutral face. These findings show that aversive learning can introduce substantial remodeling of exploration patterns in an adaptive manner during viewing of faces. We suggest that separation of patterns for harmful and safe prototypes results from an internal categorization process operating along the perceptual continuum following learning.

Authors Summary

Eye movements can shed light on the global objectives of the nervous system, as they represent the final behavioral outcome of complex neuronal processes. They can therefore provide important insights into systems level alterations induced by aversive learning, which is important to elucidate as many anxiety disorders are believed to result from an inability to form optimal aversive representations. Participants associated an aversive outcome with a given face positioned along a similarity continuum, thereby learning facial prototypes for adversity and safety. We examined eye-movement patterns during viewing of these faces by characterizing their similarity relationships. Before learning, the known similarity relationships between the stimuli could be estimated based on eye-movement patterns recorded during viewing of these faces. This indicates that exploration of neutral faces was mainly driven by their physical characteristics. Aversive learning gave rise to a decrease in similarity of viewing patterns specifically along the adversity gradient, indicating the presence of a new exploration strategy for the newly learnt adversity and safety prototypes. Our results provide evidence for adaptive changes in viewing strategies of faces with learning, and are compatible with the view that the nervous system achieves categorization to distinguish safety and adversity following aversive experiences.

Introduction

To avoid costly situations, animals must be able to rapidly predict future adversity based on previously learnt aversive associations [1], as well as actively sampled information from the environment. However, sensory samples are noisy and the environment is complex, consequently newly encountered situations are never exactly the same as previously experienced ones [2,3]. Therefore, for aversive learning to be effective a careful balance between stimulus generalization and selectivity is needed [4,5]. While generalization makes it possible to promptly deploy defensive behavior when similar situations are encountered anew [6–8], selectivity ensures that only truly aversive stimuli are recognized as aversive [9,10], thus avoiding costly false alarms. In real-world situations adversity predictions are based on sensory samples collected through active exploration [11,12]. A central part of active exploration are eye-movements [12–16] which can rapidly determine what information is available in a scene for recognizing adversity [17]. Yet, it is not known in how far representations of adversity interact with active exploration during viewing of complex visual information. Here we investigated this question by comparing exploration strategies during viewing of faces before and after aversive learning.

Face viewing behavior offers an ideal test bed for investigating changes in active exploration strategies through learning. First, active viewing of faces is a key ability during daily social interactions [18,19] where detecting minute differences in the configuration of facial elements is crucial for inferring the identity or emotional content of a face [20–22]. For humans, it is therefore a natural choice of stimuli to investigate how exploration strategies change with learning. Second, the universal spatial configuration of facial elements makes it easily possible to generate faces with subtle differences that globally form a perceptual similarity continuum [10,23]. These key features make it possible to use a task that mimics a real-world exploration context, and therefore offers the possibility to probe changes in exploration strategies with aversive learning along a parametrically controlled stimulus continuum. For aversive learning, one randomly chosen face along this continuum (CS+) was paired with a mild electric shock (UCS) through a simple Pavlovian procedure, which introduced an adversity gradient based on physical similarity to the CS+ face. The most dissimilar face (CS–) separated by 180° on the circular continuum was not reinforced and thus stayed neutral. Using this approach, we investigated how exploration strategies were modified by both the physical similarity relationships between faces, as well as the adversity gradient introduced through the aversive learning.

Necessity for model-based fixation-pattern similarity analysis

Exploration strategies for faces are typically investigated by counting the number of fixations within predefined regions of interest, such as regions centered on the eyes, mouth and nose elements. Modifications of exploration strategies with aversive learning can therefore be characterized by relative changes in the number of fixations within these regions. However, this approach can detect changes only when modifications in fixation locations are spatially consistent across participants. For example, this approach might fail finding a true effect of aversive learning when one group of participants focuses more on the right eye after learning and another on the left eye. In such cases it is possible that the net change in both regions of interest is small. Additionally, aversive learning might influence exploration behavior by optimizing fixation locations in order to collect diagnostic information about adversity in a more precise manner. This could lead to small shifts in location of fixation points that may not necessarily result in major differences in fixation counts across regions of interest.

Because of these drawbacks, we complemented count-based analyses with a variant of representational similarity analysis [24] that we term "fixation-pattern similarity analysis" (FPSA, Fig 1A). FPSA considers exploration patterns as multivariate entities [25–31] and assesses the between-condition dissimilarity of the entire fixation pattern for individual participants (Fig 1A). FPSA thereby eliminates the requirement for arbitrarily defined regions of interest, while at the same time being sensitive for fine-grained changes in exploration patterns. Furthermore, FPSA has the added benefit that it can cope with large inter-individual differences on facial exploration patterns that occur naturally [29,31–33]. While this inter-individual variability in exploration patterns can dilute the sensitivity of count-based approaches, FPSA would only require a consistent change in the similarity relationships of exploration patterns with aversive learning.

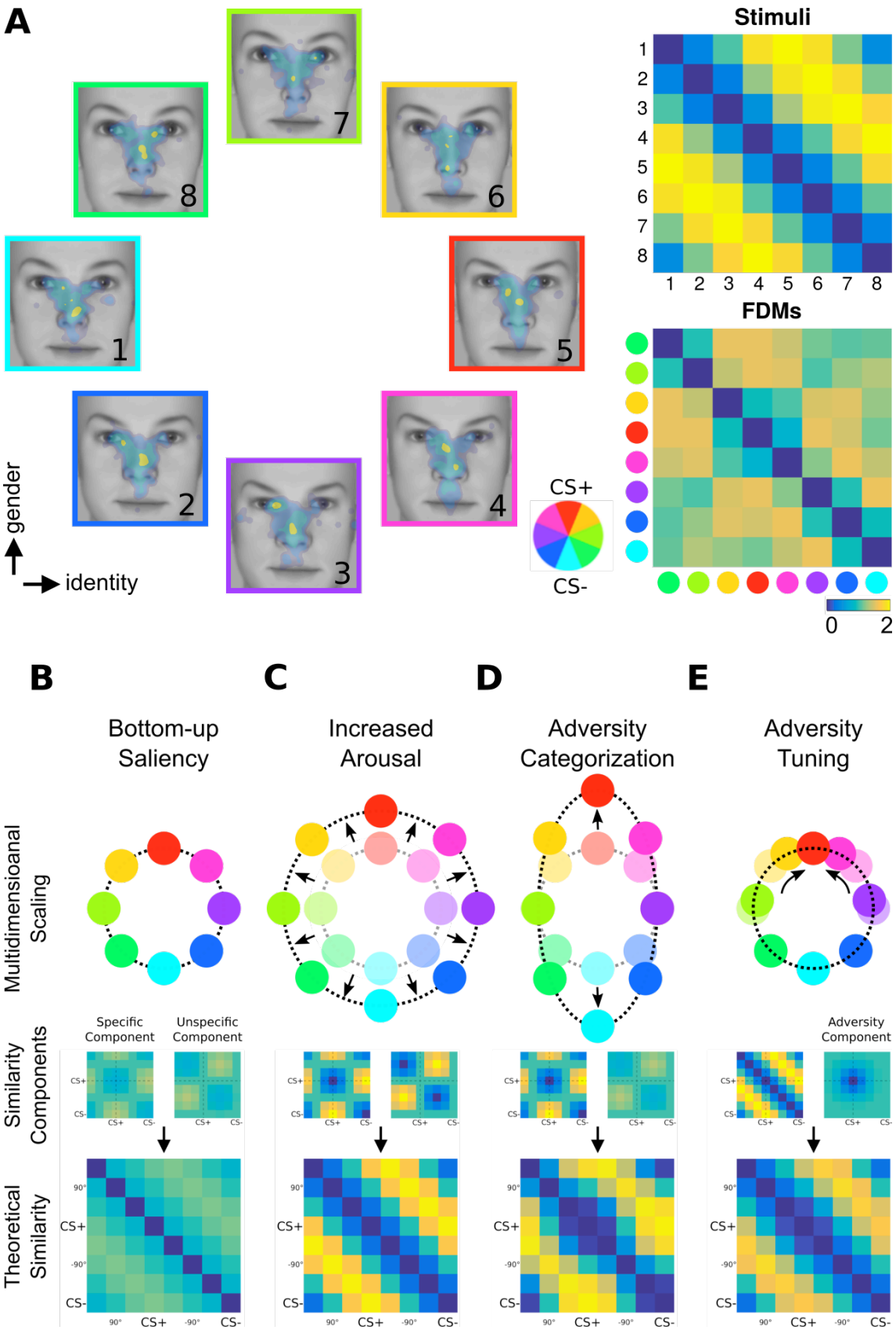


Fig 1. Model-Based Fixation-Pattern Similarity Analysis.

(A) 8 exploration patterns (*colored frames*) from a representative individual overlaid on 8 face stimuli (numbered 1 to 8) calibrated to span a circular similarity continuum across two dimensions (gender and identity;

see also SFig 1 for original stimuli). A pair of maximally dissimilar faces was randomly selected as CS+ (*red border*) and CS– (*cyan border*; see color wheel for color code). The similarity relationships among the 8 faces and the resulting exploration patterns are depicted as two 8×8 matrices. Physical similarity (*top right panel*) between all pair-wise combination of faces were calibrated (see methods and SFig 2A) to have a perfect circular similarity, characterized by highest similarity (*blue*) between neighbors, and lowest similarity (*yellow*) for opposing pairs (see also SFig 2 for calibration). FPSA summarizes the similarity relationship between the 8 exploration patterns as a symmetric 8×8 matrix (*bottom right panel*). Here and in the following, 4th and 8th columns (and rows) are aligned with the CS+ and CS–, respectively. (B-E) Multidimensional scaling representation of four theoretical similarity relationships between exploration maps (*top row*). Each colored node represents one exploration pattern (same color scheme; *red:* CS+; *cyan:* CS–), where internode distances are proportional to dissimilarity between exploration patterns, depicted as 8×8 matrices (*bottom row*). *Shaded nodes* in (C-E) depict the first hypothesis shown in (B). These matrices are further decomposed onto basic similarity components (*middle row*) centered either on the CS+/CS– (specific component) or +90°/–90° faces (unspecific component). A third component (middle row, leftmost panel) is uniquely centered on the CS+ face (adversity component). In (B), equal contribution of individual components results in circularly similar exploration patterns. In (C), a stronger equal contribution results in a better global separation of all exploration patterns (denoted by *radial arrows* second column). In (D), a stronger contribution of the specific component results in a biased separation of exploration patterns specifically along the adversity gradient defined between the CS+ and CS– nodes. In (E), the adversity component centered on the CS+ face can specifically decrease the dissimilarity of exploration patterns for faces similar to the CS+, resulting in circularly shifted nodes (*circular arrows*) while preserving the global circularity of the similarity relationships.

Hypotheses on learning-induced changes in the similarity of exploration patterns.

Using model-based FPSA we formulated parametric hypotheses on how aversive learning might alter the similarity relationships between exploration patterns when one face on the continuum started to predict adversity (Fig 1B-E, top and bottom panels). Importantly, the circular organization of stimuli allowed us to examine the similarity structure along radial and circular directions by decomposing the similarity structure into three basic components [34] (Fig 1B-E, middle panels). The *specific* and *unspecific* components (Fig 1B, middle panel) model radial changes along two orthogonal axes, separating on the one hand the adversity and safety predicting faces, and on the other faces located ±90° in relation to the CS+, serving as control independent of adversity. Hence, a stronger contribution of the specific component can capture increased pattern dissimilarity along the adversity gradient. The third *adversity* component (Fig 1E, middle panel) models local changes in dissimilarity of exploration patterns only around the CS+ in a symmetrical manner. We formulated four mutually non-exclusive hypotheses about the dissimilarity of fixation patterns across stimuli that differentially affect these components.

First, if the fixation selection mechanism during face viewing is based on salient low-level features [11,35], we would expect exploration patterns to track the circular similarity relationships between faces irrespective of aversive learning (Fig 1B). Therefore, the *bottom-up saliency* hypothesis predicts that a circular relationship between exploration patterns is already present before aversive learning has taken place. This would be characterized by low dissimilarity between neighboring faces (1st off-diagonal) and high dissimilarity between opposing faces separated by 180° (4th off-diagonal, Fig 1B bottom panel), i.e. the adversity specific and unspecific components would have about equal weight. Second, aversive learning might lead to heightened arousal, resulting in an increased contribution of low-level image features to the selection of fixation locations with the objective of collecting increased sensory evidence [36]. This would result in exploration strategies that more strongly mirror the physical similarity relationships between faces (Fig 1C) and leads to a globally

increased dissimilarity between all exploration patterns. The *increased arousal* hypothesis therefore predicts an equal but stronger contribution of the underlying specific and unspecific components (Fig 1C, middle panel), leading to a better separation of all exploration patterns globally in comparison to pre-learning period. Third, eye-movements may reflect a categorization process for faces as aversive vs. safe [37–41]. To achieve this, exploration strategies can be tailored to collect relevant information to predict adversity and safety. Such a fixation strategy would preferentially target locations that are maximally discriminative of the CS+ and CS– faces. This would lead to exploration patterns becoming more similar for faces sharing similar features with the CS+ and CS– faces, while simultaneously predicting an increased dissimilarity between these two sets of exploration patterns. Therefore, the *adversity categorization* hypothesis would lead to an increase of the adversity specific component without influencing the unspecific component (Fig 1D, middle panel). As a fourth possible scenario, aversive learning might result in the deployment of a new sensorimotor strategy only for the adversity predicting face, thereby leading to a localized change in the similarity relationships around the adversity-predicting CS+ face (Fig 1E). This is supported by evidence from univariate behavioral readouts such as autonomic skin-conductance responses [23], subjective ratings of subjective adversity [42] and startle responses [43] that show canonical generalization profiles consisting of gradually decaying responses with increasing dissimilarity to the CS+ stimulus. In a similar line, the *adversity tuning* hypothesis predicts an increased similarity of exploration patterns for faces closely neighboring the CS+ face, decaying proportionally with increasing dissimilarity to the CS+ face. This strategy would selectively increase the weight of the CS+ centered component without changing either the adversity specific or unspecific components.

In sum, using FPSA we analyzed the similarity relationships between exploration patterns during viewing of faces. We provide first evidence that exploration patterns during viewing of faces can be adaptively tailored during generalization following an aversive learning. First, aversive learning changed exploration patterns in subtle ways that were not captured by fixation counts. Second, before learning, exploration patterns showed an approximately circular similarity structure that followed the physical stimulus similarity structure. Third, after learning the similarity structure changed specifically along the adversity gradient, indicating that CS+ and CS– exploration patterns were jointly modified, while the similarity between other faces remained largely unchanged.

Results

Univariate generalization profiles in ratings, autonomic responses and fixation counts

We created 8 face stimuli that were organized along a circular similarity continuum characterized by subtle physical differences in facial elements across two dimensions (gender and identity; see SFig 1 for stimuli). We carefully calibrated the degree of similarity between all pairwise combinations of these faces using a simple model of the primary visual cortex known to mirror human similarity judgments [44] (see SFig 2 for calibration). The similarity relationship between all pair-wise faces conformed to a near perfect circular organization (Fig 1A, top right panel), such that dissimilarity varied with angular difference between faces (lowest for left and right neighbors and highest for opposing faces) with equidistant angular steps. Participants (n = 61) viewed these faces before and after an aversive associative learning procedure (Fig 2A) while we measured their eye-movements. During the conditioning phase, only the CS+ and CS– faces were presented and the CS+ face was partially reinforced with an aversive outcome (UCS, mild electric shock in ~30% trials). The CS– was the face most dissimilar to the CS+ (separated by 180°) and was not reinforced. During the subsequent generalization phase, all faces were presented and the CS+ continued to be partially reinforced to prevent extinction of the previously learnt association. These reinforced trials were excluded from the analysis. To ensure comparable arousal states between the baseline and generalization phases, we administered UCSs also during the baseline period, however they were fully predictable as their occurrence was indicated by a shock symbol (Fig 2A). Furthermore, we inserted null trials during all phases (i.e. trials without stimulus presentation but otherwise exactly the same) in order to obtain reliable baseline levels.

To monitor aversive learning we used autonomic skin-conductance responses (SCR; during each phase; Fig 2B) and subjective ratings of UCS expectancy (at the end of each phase; Fig 2C) as univariate behavioral readouts evoked by individual faces. As expected, the aversive association had a profound effect on these univariate measurements. SCR recorded during the conditioning phase were on average 4.4 times higher for the CS+ face than CS– (Fig 2B, middle panel, paired t-test, $p < .001$). In agreement with autonomic responses, UCS expectancy ratings gathered at the end of the conditioning phase were also highest for the CS+ face (Fig 2C, middle panel). The CS+ face therefore gained an aversive quality that was stronger than the CS– face during the conditioning phase, as shown by both subjective reports as well as autonomic measures. In the subsequent generalization phase, amplitudes in both measurements decayed with increasing dissimilarity to the CS+ face (Fig 2B-C, right panel) leading to an adversity-tuned profile which was well captured by a circular Gaussian curve in both recording modalities (comparison to flat null model, $p < .001$, log-likelihood ratio test). Notably, during the generalization phase, subjective ratings for both the CS+ and CS– faces differed from their respective values in the baseline phase (Fig 2B-C). This suggests that aversive learning simultaneously modified the adversity associated with the CS+ and CS– faces in opposite directions. In line with this, arousal levels measured by SCR evoked by the CS– were indistinguishable from neutral null trials (t-test; $p > .01$; shown as gray area in Fig 2B). This suggests that CS– faces were devoid of any aversive associations during the generalization phase. Furthermore, we ruled out that aversive associations were already present before learning, i.e. during the baseline phase. Here model comparison favored the flat null model in both recording modalities (comparison of flat null model and Gaussian model $p = .54$, log-likelihood ratio test; black horizontal lines in Fig 2B-C). In summary,

these univariate measurements confirmed that aversive learning was successfully established and transferred towards other perceptually similar stimuli, providing evidence for generalization following aversive learning.

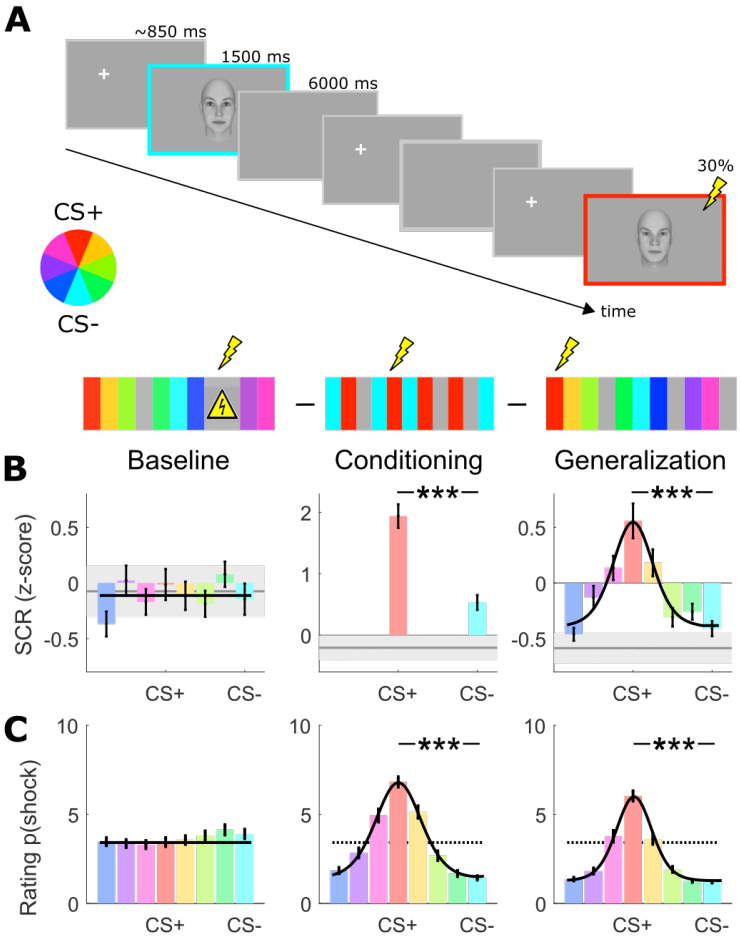


Fig 2. Univariate Characterization of Aversive Learning

(A) On every trial, one out of 8 faces was presented for 1.5 seconds preceded by a fixation cross which was randomly placed outside of the face on either the left or right side and with minimum presentation of 850ms (jittered). In some trials, no face was shown (null trial, *gray*), resulting in a SOA of 6 or 12s. For each volunteer, a pair of most dissimilar faces was randomly selected as the CS+ (*red*) and CS– (*cyan*, see color wheel). During baseline, UCSs (indicated by shock sign) were completely predictable by a triangular signboard. During conditioning and generalization, the CS+ face was paired with an aversive outcome in ~30% of CS+ trials. (B) Group-level z-scored skin-conductance responses (n = 51) and (C) subjective ratings of UCS expectancy (n = 61) for baseline, conditioning and generalization phases for individual faces (same color code). Responses are aligned to the CS+ for each volunteer separately. In (B), the gray shaded area indicates response amplitudes evoked by the null trials (mean and 95% CI). For baseline and generalization phases, the winning model (circular Gaussian vs. flat null model) is depicted as either a black line (null model) or curve (circular Gaussian model). In (C) the average ratings for the baseline period are depicted as dashed line also for conditioning and generalization phases. Asterisks depict significant differences between responses to CS+ and CS– stimuli. (***: $p < .001$, *t-test*). Error bars denote SEM.

Whether complex behavior such as eye movement patterns during viewing of faces also exhibits learning-induced changes, and if so, whether it exhibits generalization during viewing of similar faces is an open question [17]. We first investigated this using a fixation count-based approach. To this end, we computed

fixation density maps (FDMs) for every participant and face separately (Fig 3A for FDMs from two representative participants) and evaluated fixation probability within the 4 different regions of interest (left and right eyes of the face, nose and mouth [45,46]; ROIs shown in Fig 3B as insets). We reasoned that if aversive learning had a specific influence on exploration of faces, this would result in a bell-shaped modulation of fixation counts around the CS+ face, similar to SCR and subjective ratings. In line with previous reports [32], left and right eyes together with the nose region were the most salient locations across the baseline and generalization phases, and attracted ~84% of all fixation density, whereas the mouth region had only a marginal contribution with ~3%. Overall, aversive learning increased the number of fixations directed at the nose (+4%) and mouth (+0.6%) regions at the expense of left (-3.5%) and right (-2.8%) eyes (Fig 3B). Repeating the same group-level analysis as for SCR and subjective ratings, we examined the presence of adversity tuning in fixation counts by testing whether a circular Gaussian could explain fixation counts along the similarity continuum, separately for each ROI. Model comparison on percentage changes (Fig 3B, black lines and curves) favored the flat null model for all regions ($p > 0.05$, log-likelihood test), with the exception of the mouth region where the Gaussian model was marginally favored ($p = .012$ uncorrected, log-likelihood ratio test). Therefore, using the fixation count-based approach, we were able to show a weak specific effect at the mouth region, which however accounted for only a small percentage of fixations overall. Thus, facial locations that accounted for most of the fixation density showed only unspecific changes that were independent of the adversity gradient.

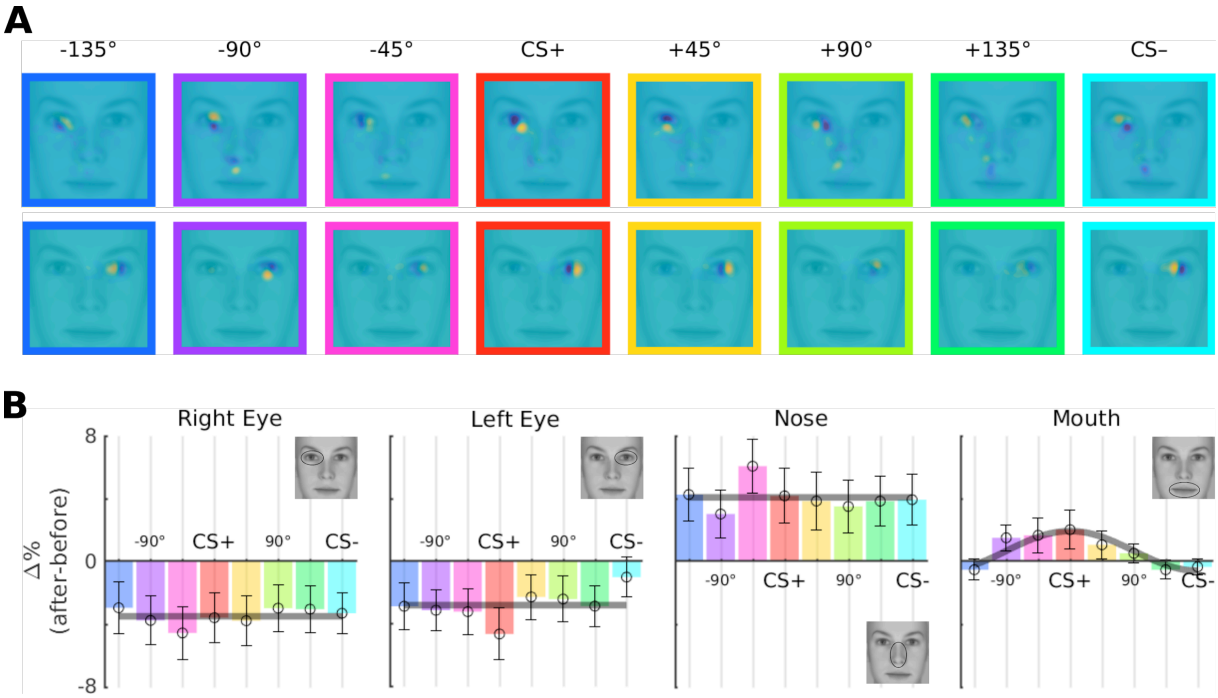


Fig 3 Impact of Aversive Learning on Fixation Counts at Four Different Regions of Interest.

(A) Fixation density maps (FDMs) of two exemplary volunteers preferentially fixating on the left (*top row*) or right eye (*bottom row*) during the generalization phase. FDMs for the 8 faces are aligned to individual CS+ face (*colored frame*), and smoothed with a Gaussian kernel of 1 visual degree. To emphasize differences between conditions, the average pattern is subtracted from single-conditions for each volunteer separately (dark blue: fewer density than average, yellow: more density than average). (B) Percentage change in fixation density for different faces (colored bars) within 4 different regions of interest (black contours in inset). Y-axis shows the difference between generalization and baseline phases (values > 0 represent more fixations during generalization). Lines or curves indicate the winning model (Flat null model vs. Gaussian model). Errorbars:

SEM.

Multivariate pattern analysis of exploration patterns

Despite the lack of an adversity-specific effect at the most salient locations, a careful examination of single-subject FDMs revealed fine-grained changes in exploration patterns that lawfully changed along the dissimilarity continuum (Fig 3A). Notably, these differences existed within the regions of interest for which fixation count analyses didn't detect any difference. This suggests that classification-based multivariate pattern analysis methods [31] might be more appropriate for investigating the impact of aversive learning on fixation patterns during viewing of faces. We thus examined multivariate information content within FDMs and tested whether eye-movements deployed for the exploration of the CS+ face could be differentiated from the CS– face beyond what could already be accounted by physical differences between the faces. We thus evaluated how accurately a cross-validated linear classifier could discriminate FDMs on these faces before and after learning, expecting decoding accuracy to increase if aversive learning led to a differentiation of exploration patterns (using a 50% holdout cross-validation with 1000 random splits of FDMs into test and training sets). We report the proportion of trials held out from training that were classified as CS+ trials, corresponding to correct classification for actual CS+ trials, but false alarms for actual CS- trials. After aversive learning, the average classification performance for CS+ (57.4 ± 1.8 %; mean ± SEM across subjects) was significantly better than before learning (53.1 ± 1.3 %; paired t-test, $p = .02$) as well as chance-level decoding (based on label permutation: 50.0 ± 0.1 %). This indicates that aversive learning introduced changes in the exploration patterns that were not present before learning. We next measured the performance of the same classifier to discriminate intermediate faces between the CS+ and CS–. The proportion of trials classified as CS+ decayed according to the typical bell-shaped curve with decreasing similarity to the CS+ face (Fig 4A middle panel, see SFig 3 for classification results for single runs), suggesting a gradual deployment of the adversity specific exploration strategy with increasing similarity to the CS+. In order to understand whether this increased decoding performance was driven by the mouth region, as it exhibited adversity-tuned changes already in fixation counts, we repeated the same analysis but this time excluding the data from this region. This yielded undistinguishable results both in terms of CS+ classification (57.7 ± 1.8 %, $p > .05$ comparison to classification including mouth ROI) as well as its generalization along the continuum (Fig 4A, rightmost panel). This excludes the possibility that decoding performance was solely driven by adversity-tuned fixation counts in the mouth region found in the previous group-level analysis. Hence, it suggests that aversive learning influenced eye-movement patterns in a distinct manner for different participants, given that group-level adversity tuning was not a requisite for above-chance level decoding performance. Altogether these results show that exploration patterns during viewing of faces were affected by aversive learning. Furthermore, these effects could not be explained by physical differences between neutral faces before learning or the adversity-tuning present in the mouth region at the group-level after learning has taken place. This corroborates the notion that aversive learning was associated with new exploration strategies that were gradually deployed during generalization.

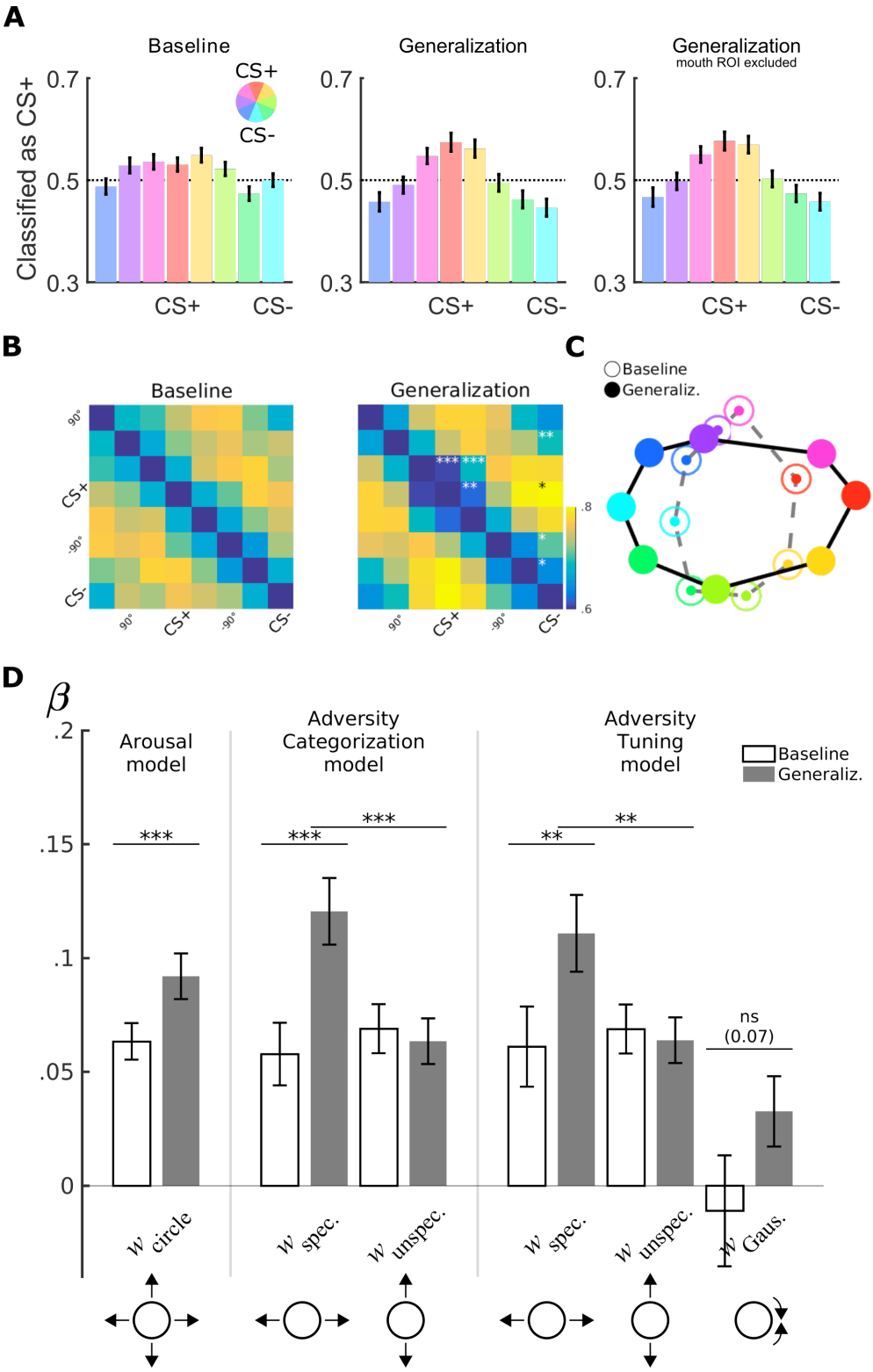


Fig 4. Classifier- and Similarity-based Multivariate Analysis of Exploration Patterns

(A) Classification accuracies for a cross-validated linear support-vector machine trained to discriminate CS+ and CS–. Bars (M ± SEM) show the mean proportion of trials classified as CS+ (*red*), i.e. correct classification for the actual CS+ condition, and false alarms for CS– trials (*cyan*). Dotted lines mark the chance level. The last panel shows the same analysis excluding fixations coming from the mouth region (Fig 3, rightmost panel) (B) Dissimilarity matrices of exploration patterns for baseline (*left panel*) and generalization phases (*right panel*).

Fourth and eight columns (and rows) are aligned with each volunteer's CS+ and CS– faces, respectively. Asterisks on the upper diagonal denote significant differences in dissimilarity values for the corresponding element between baseline and generalization phases. (C) Multidimensional representational similarity analysis conducted jointly on 16×16 dissimilarity matrix (not shown) comprising baseline and generalization phases. Distances between nodes are proportional to the dissimilarity between corresponding FDMs (*open circles*: baseline; *filled circles*: generalization phase; same color scheme). (D) Bar plots (M ± SEM) depict predictor weights estimated for single-participants before (white bars) and after (gray bars) learning for different models. (*Left:* bottom-up saliency and arousal models; *middle:* adversity categorization; *right:* adversity tuning). w_{circle}: weight for the circular component, which is the sum of equally weighted specific and unspecific components; $w_{\text{specific}}/w_{\text{unspecific}}$: weights for specific and unspecific components centered; w_{Gauss}: weight for adversity component centered uniquely on the CS+. (**: p < .01; ***: p < .001, paired *t-test*).

Model-based fixation-pattern similarity analysis

However, these results cannot disentangle different scenarios about learning-induced changes in the exploration patterns, as all the outlined scenarios predict smoothly increasing dissimilarity values between the CS+ and CS– faces. To gain further insights, we therefore used model-based FPSA, which exploits similarity relationships between all pairwise combinations of exploration patterns. We computed a dissimilarity matrix consisting of all pairwise comparisons of FDMs for individual volunteers (using 1 - Pearson correlation as a pattern distance measure) and averaged these after separately aligning them to each volunteer's CS+ face (shown always at the 4th column and row in Fig 4B). Furthermore, in order to gather an intuitive understanding of the learning-induced changes in the similarity geometry we used multidimensional scaling (jointly computed on the 16×16 matrices). Multidimensional scaling (MDS) summarizes similarity matrices by transforming observed dissimilarities as closely as possible onto distances between different nodes (Fig 4C) representing different viewing patterns, therefore making it easily understandable at a descriptive level.

Already during the baseline period the dissimilarity matrix was highly structured (Fig 4B). In agreement with a circular similarity geometry and the MDS depiction, lowest dissimilarity values (1.04 ± .01; M ± SEM) were found between FDMs of neighboring faces (i.e. first off-diagonal), whereas FDMs for faces separated by 180° exhibited significantly higher dissimilarity values (1.21 ± .01; paired t-test, $t(60) = 7.03$, $p < .001$). Using the bottom-up saliency model, we investigated the contribution of physical characteristics of the stimulus set to the observed pre-learning dissimilarity structure (Fig 1B). This model uses a theoretically circular similarity matrix (consisting of equally weighted sums of specific and unspecific components) as a linear predictor, this way estimating the global dissimilarity between exploration patterns in accordance with a circular organization. The circular bottom-up saliency model performed significantly better compared to a null model consisting of a constant similarity for all pairwise FDMs comparisons (for bottom-up model adjusted $r^2 = .09$; log-likelihood-ratio test for the alternative null model: $p < 10^{-5}$; $\text{BIC}_{\text{NullModel}} = -1529.3$, $\text{BIC}_{\text{BottomUp}} = -1650$; see S1 Table for the results of model fitting). We additionally fitted the bottom-up model for every volunteer separately (Fig 4D). Model parameters at the aggregate level were significantly different from zero (Fig 4C; $w_{\text{Circle}} = .063 \pm 0.008$, M ± SEM; $t(60) = 7.89$, $p < 10^{-5}$) indicating that exploration strategies prior to learning mirrored the physical similarity structure of the stimulus set. This provides evidence that fixation selection strategies are, at least to some extent, guided by physical stimulus properties during viewing of neutral faces.

However, we observed significant changes when comparing baseline and generalization dissimilarity values element-by-element (Fig 4B, indicated by asterisks) providing evidence for learning-induced changes in the similarity relationships. The same bottom-up saliency model was again significant (adjusted $r^2 = .33$; $p < 10^{-}$

[5], log-likelihood ratio test), but now even performed notably better compared to the baseline phase ($BIC_{BottomUp}$ = -1650 for the baseline vs. $BIC_{BottomUp}$ = -2715.5 for the generalization phase; see S2 Table for model fitting results). Critically, we found a significant increase in the model parameter from baseline to generalization phase (w_{Circle} = 0.092 ± 0.01; paired *t*-test, $t(60) = 9.13$, $p < 10^{-5}$; Fig 4D compare two leftmost bars) suggesting a global increase in dissimilarity between FDMs. Overall, these results are compatible with the view that aversive learning led to a better separation of exploration patterns globally, in agreement with the heightened arousal model (Fig 1C), which predicted an increased contribution of the bottom-up saliency to the similarity of exploration patterns as shown by larger model parameters.

However, the MDS method suggested that the separation of exploration patterns might have occurred mainly along the adversity gradient defined by the CS+ and CS– faces, whereas the separation along the orthogonal direction did not exhibit any noticeable changes (Fig 4C). We thus extended the circular bottom-up model to capture independent variance along the two orthogonal directions using the adversity categorization model (Fig 1D). Model comparison indicated that this model performed better than the bottom-up model ($BIC_{BottomUp}$ = -2715 vs. $BIC_{AdversityCateg.}$ = -2897.3; adjusted r^2 = .44; see S3/4 Table for fitting results with the adversity categorization model on baseline and generalization phases, respectively). Notably, this difference was accompanied by a nearly two times stronger contribution of the specific component ($w_{Specific}$ = 0.12 ± 0.014, $t(60) = 21.034$; $w_{Unspecific}$ = 0.063 ± 0.01, $t(60) = 11.07$; Fig 4D), which was significantly larger than the unspecific component (pair-wise *t-test,* $t(60) = -3.81$, $p = 3.2 \times 10^{-4}$,). Additionally, the weight of the unspecific similarity component did not exhibit a significant modification with learning. This provides evidence that increased overall dissimilarity with learning was driven by changes in the scanning behavior specifically along the task-relevant adversity direction.

The remodeling of the similarity geometry along the adversity gradient can also be accompanied by exploration strategies that are specifically deployed for the adversity predicting face, which would result in localized changes in the similarity geometry only around the CS+ face. We subjected this view to model comparison by augmenting the previous model with a similarity component that consisted of a two-dimensional Gaussian centered on the CS+ face. The width parameter of the Gaussian was adjusted to be around 65°, a value close to the width of adversity tuning in subjective ratings. Positive contribution of this predictor would lead to more similar exploration patterns around the CS+ (Fig 1E). It can thus capture changes in similarity relationships that are specific to the CS+ face. The model comparison procedure favored the simpler adversity categorization model over the augmented adversity tuning model ($BIC_{AdversityCateg.}$ = -2897.3 vs. $BIC_{AdversityTuning.}$= -2864.4 during the generalization phase; adjusted r^2 = .44; see S5/6 Table for fitting results with adversity tuning model in baseline and generalization phases, respectively). Hence the increase in the number of predictors did not result in a significant reduction in explained variance. In line with this result, the parameter estimates for the adversity component were not significantly different than zero neither in baseline or generalization phases ($w_{Gaussian}$ = 0.015 ± 0.04 in baseline, p = .72, $t = 0.35$; $w_{Gaussian}$ = 0.07 ± 0.04 in generalization, p = 0.12, $t = 1.56$; Fig 4C). Also, pair-wise differences between parameter estimates did not reach significance (p = 0.37, t = 0.89). We therefore conclude that further improvements of the adversity categorization model to include adversity-specific changes did not result in a better understanding of the adversity-induced changes in the similarity geometry of exploration strategies.

Discussion

We aimed to characterize the effect of aversive learning at the systems level, by examining changes in eye movement strategies during viewing of faces organized along a similarity continuum. As expected, subjective ratings of shock expectancy and autonomic recordings of arousal exhibited adversity-tuned responses that peaked on the CS+ face and decayed smoothly with increasing dissimilarity. Adversity tuning of subjective ratings emerged as a differentiation of responses to both the CS+ and CS– face, in line with the view that learning entails formation of both safe and harmful associations. However, in contrast to SCR and subjective ratings we observed only weak evidence in favor of adversity-tuned changes in the fixation counts associated with different facial elements. This could arise due to individually distinct modifications in overt behavior strategies with aversive learning, which are diluted when averaging fixation counts across large regions of interest. This view was supported by above-chance level classification between the CS+ and CS–, which was based on single-subject exploration patterns. In sum, our results provide evidence that active exploration patterns during viewing of faces are flexible and can be adaptively tailored following aversive learning to explore adversity- and safety-related facial prototypes.

We investigated the nature of these adaptive changes in exploration strategies using a similarity-based multivariate technique that we call fixation-pattern similarity analysis. In combination with a set of stimuli that were parametrically controlled, this had several advantages over typically used approaches to the phenomenon of generalization. When responses (e.g. neuronal or behavioral) are tested along a single dimension, this results in smoothly decaying response amplitudes with increasing dissimilarity to the adversity predicting stimulus. These generalization profiles can provide important clues about the selectivity of aversive representations [10] and may help understanding cognitive impairments such as anxiety disorders characterized by less selective generalization profiles [47–49]. However, the different hypotheses we could test here are difficult to be distinguished based on univariate generalization profiles, as they all predict monotonously decaying generalization profiles with increasing dissimilarity to the CS+. The exact distinction between different hypotheses would require an extensive characterization of the decay components with increasing dissimilarity [3]. For example, in comparison to the adversity-tuning hypothesis, the categorization hypothesis might result in a faster decaying generalization profile, and a thus more selective tuning. Nevertheless, the exact relationships between pattern dissimilarities and the decay rate in univariate generalization profiles cannot be established straightforwardly. Fixation-pattern similarity analysis, in the same spirit as representational similarity analysis [24] exploits information present in all pair-wise combinations between multivariate patterns. In comparison to univariate generalization profiles, FPSA can disambiguate between these hypotheses as they differ in how they predict pair-wise relationships between exploration patterns. In other words, FPSA disambiguates different models by how well they predict multivariate patterns instead of univariate response amplitudes.

We therefore characterized learning-associated changes using model-based FPSA in conjunction with a stimulus set that was parametrically controlled. Already before learning, the similarity of exploration patterns was highly structured and reflected the circularity of the face continuum. This is compatible with the view that exploration of neutral faces is, at least to some extent, guided by physical characteristics—in contradiction with a purely holistic viewing strategy for faces. Following aversive learning, we observed a significant remodeling of the similarity structure leading to an increase in dissimilarity. This was caused by an increased separation of

exploration patterns along the adversity gradient, indicated by a stronger contribution of the adversity-specific component. Modeling of the similarity relationships indicated an increase of dissimilarity between the CS+ and CS– faces, which was concomitant to a joint increase in similarity around the CS+ and CS–. This points to the fact that aversive learning jointly influenced exploration patterns at both ends of the stimulus continuum. This is compatible with the view that following an aversive learning, the nervous system achieves a categorization process along the smoothly changing perceptual continuum [37,39,41,50]. Our results show that this categorization process results also in detectable changes at the behavioral level. These changes in exploration patterns would presumably lead to an increased efficiency of information transmission downstream and help the categorization process of faces. Here, the key contribution of the FPSA was providing insights onto how active exploration strategies were remodeled with aversive learning in a way that could not have been easily predicted based on univariate generalization profiles. Furthermore, FPSA allowed us to better understand these changes as adaptive modifications in exploration strategies that were specifically tailored along the adversity gradient.

Multivariate pattern analysis methods provide a set of powerful tools for extracting information in eye-movements recordings. For example, classifier-based methods can successfully decode the identity of different observers [31,33], or task-dependent changes in eye-movements strategies within the same observers [27,28,30,31,52]. Furthermore, they might contribute to improve mental health screening when used as behavioral biomarkers [25,53]. We complemented classifier-based analyses with a similarity-based pattern analysis to gather insights on the specific ways eye-movement patterns changed during viewing of faces with learning. In this report, use of a calibrated stimulus continuum enabled us modeling of similarity relationships in a parametric manner. This provided an understanding of the increased decoding performance with aversive learning as an increase in dissimilarity along the task-relevant direction. In a similar line, Kietzmann et al. [54] parametrically modeled similarity relationships as a sanity-check for electro-encephalogram recordings, which are extremely sensitive to eye-movement induced artifacts.

While skin-conductance responses and subjective ratings inform about the aggregate cognitive evaluations of a given stimulus, eye-movements can be informative about the temporal evolution of the information sampling process [55]. In so far, they allow to reason about how changes in viewing strategies from learning might produce the observed changes in pattern similarities. In this respect an interesting question is how the fixation selection process produces changes in viewing strategies selectively along the CS+/CS– axis. Clearly, following aversive learning the selection of fixation locations must become sensitive to the adversity introduced with the learning procedure. One possibility is that participants quickly judge the risk of being shocked with the first landing fixation, and consequently that they are more aroused and attentive with higher adversity. Therefore, adversity could selectively increase the influence of salient stimulus features. Such a process would explain why exploration patterns change on the CS+ stimulus and other stimuli sharing similar features. This suggestion is also compatible with effects of aversive learning on overt attention. For example, fearful faces are more salient and attract more fixations than neutral ones [56–58]. Furthermore, elementary visual features can benefit higher priority when they predict adversity, to the point of distracting an on-going task [17]. Most importantly, their strength gradually increases with similarity to the adversity predicting features [17]. These oculomotor saliency gradients indicate that stimuli predicting adversity receive higher priority during sensorimotor processing. Yet, this view does not give an account of why exploration patterns also change on the

CS– stimulus, which presumably would be classified as safe soon after stimulus onset and should therefore not lead to an increase in saliency. Therefore, a second possibility is that sensory information provided by the first fixation is not enough to predict adversity with high certainty. In this case, it is plausible that subsequent fixations sample areas that are informative for both adversity and safety in order to resolve the remaining uncertainty. Such a strategy would result in separation of exploration patterns only along the CS+ and CS– axis. The benefit of this strategy would presumably be increased information transmission to better categorize faces as either safe or aversive.

Eye-movements patterns can provide important insights about what the nervous system tries to achieve as they summarize the final outcome of complex interactions at the neuronal level [59]. Our results demonstrate that changes induced by aversive generalization extend beyond autonomous systems or explicit subjective evaluations, but can also affect an entire sensory-motor loop at the systems level [17]. Furthermore the methodology applied here can easily be extended to neuronal recordings, where gradients of activity during generalization have been successfully used to characterize selectivity of aversive representations. Therefore, it will be highly informative to test different hypotheses we outlined here using neuronal recordings with representational similarity analysis during the emergence of aversive representations.

Materials and Methods

Participants

Participants were 74 naïve healthy males and females (n = 37 each) with normal (or corrected-to-normal) vision (age = 27 ± 4, M ± SD) and without history of psychiatric or neurological diseases, any medical condition or use of medication that would alter pain perception. Out of 74 participants, we discarded 13 participants who did not successfully associate the CS+ face with the UCS based on their subjective ratings. This exclusion criterion was based on a model comparison procedure testing whether ratings at the end of the generalization phase could be significantly better modeled with a circular Gaussian in comparison to a flat null model (see Nonlinear modeling and model comparison section). We furthermore required that the peak of the fitted circular Gaussian was within the ±45° degrees of the CS+ face. We thus conducted the analyses of eye-movements on a homogenous set of participants who were aversively conditioned (n = 61, 31 males) and could reliably detect the CS+ face. Participants had not participated in any other study using facial stimuli in combination with aversive learning before. They were paid 12 Euros per hour for their participation in the experiment and provided written informed consent. All experimental procedures were approved by the Ethics committee of the Chamber of Physicians in Hamburg.

Data sharing

The dataset used in this manuscript has been published as a dataset publication [62]. We publicly provide the stimuli as well as the code ([63] developed with Matlab 2016b, MathWorks, Natick MA) necessary to download the dataset, conduct all analyses and prepare the figures reported here.

Stimulus preparation and calibration of generalization gradient

Using a two-step procedure, we created a final set of 8 calibrated faces (Fig 1A, see also SFig 1) that were perceptually organized along a circular similarity continuum based on a model of the primary visual (V1) cortex. Using the FaceGen software (FaceGen Modeller 2.0, Singular Inversion, Ontario Canada) we created two gender-neutral facial identities and mixed these identities (0%/100% to 100%/0%) while simultaneously changing the gender parameters in two directions (more male or female). In the first step, we created a total of 160 faces by appropriately mixing the gender and identity parameters to form 5 concentric circles (see SFig 1) based on FaceGen defined parameter values for gender and identity. Using a simple model of the primary visual cortex known to closely mirror human perceptual similarity judgments [44], we computed V1 representations for each face after converting them to grayscale. The spatial frequency sensitivity of the V1 model was adjusted to match human contrast sensitivity function with bandpass characteristics between 1 and 12 cycles/degree, peaking at 6 cycles/degrees [64]. The V1 model consists of pair of Gabor filters in quadrature at five different spatial scales and eight orientations. The activity of these 40 channels were averaged in order to obtain one single V1 representation per face. We characterized the similarity relationship between the V1 representations of 160 faces using multidimensional scaling analysis with 2 dimensions (SFig 2). As expected, while two dimensions explained a large variance, the improvement with the addition of a third dimension was only minor, providing thus evidence that the physical properties of the faces were indeed organized along two-dimensions (stress

values for 1D, 2D and 3D resulting from the MDS analysis were 0.42, .04, .03, respectively). The transformation between the coordinates of the FaceGen software values (gender and identity mixing values) and coordinates returned by the MDS analysis allowed us to gather FaceGen coordinates that would correspond to a perfect circle in the V1 model. In the second step, we thus generated 8 faces that corresponded to a perfect circle. This procedure ensured that faces used in this study were organized perfectly along a circular similarity continuum according to a simple model of primary visual cortex with well-defined bandpass characteristics known to mirror human similarity judgments. Furthermore it ensured that dimensions of gender and identity introduced independent variance on the faces.

To present these stimuli we resized them to 1000x1000 pixels (originals: 400x400) using bilinear interpolation, and slightly smoothed with a Gaussian kernel of 5 pixels with full-width at half maximum of 1.4 pixels to remove any possible pixel artifacts that could potentially lead participants to identify faces. Faces were then normalized to have equal luminance and root-mean-square contrast. The gray background was set to the same luminance level ensuring equal brightness throughout of the experiment. Faces were presented on a 20" monitor (1600 x 1200 pixels, 60 Hz) using Matlab R2013a (Mathworks, Natick MA) with psychophysics toolbox [65,66]. The distance of the participants' eyes to the stimulus presentation screen was 50 cm. The center of the screen was at the same level as the participants' eyes. Faces spanned horizontally ~17° and vertically ~30°, aiming to mimic a typical face-to-face social situation. Stimuli are available in [63].

Experimental paradigm

The fear conditioning paradigm (similar to [10]) consisted of baseline, conditioning and test (or generalization) phases (Fig 2A). Four equivalent runs with exactly same number of trials were used during baseline (1 run) and generalization phases (3 runs) consisting of 120 trials per run (~10 minutes). Every run started with an eye-tracker calibration. Between runs participants took a break and continued with the next run in a self-paced manner. We avoided having more than 1 runs in the baseline period in order not to induce fatigue in participants. This consisted of a blurred unrecognizable face. At each run during the baseline and generalization phases, 8 faces were repeated 11 times, UCS trials occurred 5 times and one oddball was presented. We presented 26 null trials with no face presentation but otherwise the same trial structure (see below sequence optimization). In order to keep arousal levels comparable to the generalization phase, UCSs were also delivered during baseline, however they were fully predictable by a shock symbol therefore avoiding any face to UCS associations. During the conditioning phase, participants saw only the CS+ and the CS– faces (and null trials). These consisted of 2 maximally dissimilar faces separated by 180° on the circular similarity continuum and randomly assigned for every participant in a balanced manner. The conditioning was 124 trials long (~10 minutes) and CS+ and CS– faces were repeated 25 times. CS+ faces were additionally presented 11 times with the UCSs, resulting in a reinforcement rate of ~30 %. The same reinforcement ratio was used during the subsequent generalization phase in order to avoid extinction of the learnt associations. Participants were instructed that the delivery of UCSs during baseline would not be associated with faces, however in the following conditioning and generalization phases they were instructed that shocks would be delivered after particular faces have been presented. In all three phases, subjects were instructed to press a button when an oddball stimulus appeared on the screen.

Faces were presented using a rapid-event design with a stimulus onset asynchrony of 6 seconds and stimulus duration of 1.5 seconds. The presentation sequence was optimized using a modified m-sequence with 11 different conditions [67,68] (8 faces, UCS, oddball, null). An m-sequence is preferred as it balances all transitions from condition *n* to *m* (thus making the sequence as unpredictable as possible for the participant) while providing an optimal design efficiency (thus making deconvolution of autonomic skin conductance responses more reliable). However all conditions in an m-sequences appear equally number of times. Therefore, in order to achieve the required reinforcement ratio (~30%), we randomly pruned UCS trials and transformed them to null trials. Similarly oddball trials were pruned to have an overall rate of ~1%. This resulted in a total of 26 null trials. While this deteriorated the efficiency of the m-sequence, it was a still good compromise as the resulting sequence was much more efficient than a random sequence. Resulting from the intermittent null trials, SAOs were 6 or 12 seconds approximately exponentially distributed.

Face onsets were preceded by a fixation-cross, which appeared randomly outside of the face either on the left or right side along an imaginary circle (r = 19.6°, +/- 15° above and below the horizontal center of the image). The side of fixation-cross was balanced across conditions to avoid confounds that might occur [55].

Calibration and delivery of electric stimulation

Mild electric shocks were delivered by a direct current stimulator (Digitimer Constant Current Stimulator, Hertfordshire UK), applied by a concentric electrode (WASP type, Speciality Developments, Kent UK) that was firmly connected to the back of the right hand and fixated by a rubber glove to ensure constant contact with the skin. Shocks were trains of 5-ms pulses at 66Hz, with a total duration of 100 ms. During the experiment they were delivered right before the offset of the face stimulus. The intensity of the electric shock applied during the experiment was calibrated for each participant before the start of the experiment. Participants underwent a QUEST procedure [69] presenting UCSs with varying amplitudes selected by an adaptive algorithm and were required to report whether a given trial was “painful” or “not painful” in a binary fashion using a sliding bar. The QUEST procedure was repeated twice to account for sensitization/habituation effects, thus obtaining a reliable estimate. Each session consisted of 12 stimuli, starting at an amplitude of 1mA. The subjective pain threshold was the intensity that participants would rate as “painful” with a probability of 50%. The amplitude used during the experiment was 2 times this threshold value. Before starting the actual experiment, participants were asked to confirm whether the resulting intensity was bearable. If not then the amplitude was incrementally reduced and the final amplitude was used for the rest of the experiment.

Eye tracking and fixation density maps

Eye tracking was done using an Eyelink 1000 Desktop Mount system (SR Research, Ontario Canada) recording the right eye at 1000 Hz. Participants placed their head on a headrest supported under the chin and forehead to keep a stable position. Participants underwent a 13 point calibration / validation procedure at the beginning of each run (1 Baseline run, 1 Conditioning run and 3 runs of Generalization). The average mean-calibration error across all runs was Mean = 0.36°, Median = .34°, SD = 0.11. 91% of all runs had a calibration better than or equal to .5°.

Fixation events were identified using commonly used parameter definitions [62] (Eyelink cognitive configuration: saccade velocity threshold = 30° / second, saccade acceleration threshold = 8000° per second^2,

motion threshold = .1°). Fixation density maps (FDMs) were computed by spatially smoothing (Gaussian kernel of 1° of full width at half maximum) a 2D histogram of fixation locations, and were transformed to probability densities by normalizing to unit sum. FDMs included the center 500x500 pixels, including all facial elements where fixations were mostly concentrated (~95% of all fixations).

Shock expectancy ratings and autonomic recordings

After baseline, conditioning and generalization phases, participants rated different faces for subjective shock expectancy by answering the following question, "*How likely is it to receive a shock for this face?*". Faces were presented in a random order and rated twice. Subjects answered using a 10 steps scale ranging from "*very unlikely*" to "*very likely*" and confirmed by a button press in a self-paced manner.

Electrodermal activity evoked by individual faces was recorded throughout the three phases. Reusable Ag/AgCl electrodes filled with isotonic gel were connected to the palm of the subject's left hand using adhesive collars, placed in thenar/hypothenar configuration. Skin-conductance responses were continuously recorded using a Biopac MP100 AD converter and amplifier system at a sampling rate of 500 Hz. Using the Ledalab toolbox [70,71], we decomposed the raw data to phasic and tonic response components after downsampling it to 100 Hz. Ledalab applies a positively constrained deconvolution technique in order to obtain phasic responses for each single trial. We averaged single-trial phasic responses separately for each condition and experimental phase to obtained 21 average values (9 (8 faces + 1 null condition) from baseline and generalization and 3 (2 faces + 1 null condition) from the conditioning phase). CS+ trials with UCS were excluded from this analysis. These values were first log-transformed ($\log_{10}(1+SCR)$) and subsequently z-scored for every subject separately (across all conditions and phases), then averaged across subjects. Therefore, negative values indicate phasic responses that are smaller than the average responses recorded throughout the experiment. Due to technical problems, SCR data could only be analyzed for n = 51 out of the 61 participants.

Nonlinear modelling and model comparison

We fitted a von Mises function (circular Gaussian) to generalization profiles obtained from subjective ratings, skin-conductance responses and fixation counts at different ROIs by minimizing the following likelihood term in (1) following an initial grid-search for parameters

$$L(D(x) \mid \theta, \sigma) = \Sigma - \log [N(D(x) - G (x \mid \theta) \mid 0, \sigma)] \quad (1)$$

where x represents signed angular distances from a given volunteer's CS+ face; $G(x|\theta)$ is a von Mises-like function that was used to model the adversity tuning. It is defined by the parameter vector θ, which codes for the amplitude (difference between peak and base), location (peak position), precision and offset (base value) of the resulting generalization profile; $D(x)$ represents the observed generalization profile for different angular distances; and $N(x| 0, \sigma)$ is the normal probability density function with mean zero and standard deviation of σ. The fitting procedure consisted of finding parameters values that minimized the sum of negative log-transformed probability values. Using log-likelihood ratio test we tested whether this model performed better than a null model consisting of a horizontal line, effectively testing the significance of the additional variance explained by the model. G(x) was a scaled and shifted version of a normalized von Mises function in the form

$$G(x) = \alpha \cdot V(x \mid K, \mu) + \theta \quad (2)$$

α represents the depth of adversity tuning which corresponds to the difference between peak and baseline responses, and θ sets to the baseline level. K and μ controls the precision of the tuning and the peak position of adversity tuning, respectively. V(x) is a modified von Mises function that is scaled to fit between 0 and 1 using the following equality,

$$V(x) = [\exp(K \cdot \cos(x-\mu)) - \exp(-K)] / \exp(K) - \exp(-K) \quad (3)$$

Classification with linear support vector machine

We used single trial FDMs for validation of linear support vector machines [72] that were trained to classify exploration patterns obtained during viewing of CS+ and CS– conditions. As the generalization phase had more trials than the baseline phase (3 runs vs. 1 run), we took precautions to make a fair comparison between these phases so that differences between number of trials do not invalidate comparison of accuracies between the baseline and generalization phases. To this end, we trained and tested a linear SVM classifiers always within a given run, and averaged classification accuracy for the generalization phase across the three runs. This effectively kept signal-to-noise ratio of FDMs in comparable levels between the baseline and generalization phases, therefore avoiding any possible favor of the generalization phase. We trained a classifier on randomly drawn 50% of the CS + and CS– trials, and tested on the remaining 50% and averaged classification performance across 1000 repetitions using this procedure. To reduce the dimensionality, FDMs were first downscaled 10 times resulting in a vector of 2500 pixels. We further reduced dimensionality by projecting FDMs onto their principal components using two different approaches. We identified the number of N principal components corresponding to the elbow where the slope of eigenvalues was levelling off substantially (N = 14, 19, 17, 20 for the 4 runs). We also repeated the analysis simply using N principal components that explained 90% of total variance in each run (N = 63, 69, 74, 75). Both approaches yielded similar results, we report numbers from the first approach in the results section, but present results obtained with both approaches in SFig 3. Principal components were computed excluding null trials, UCS trials and oddballs. Loadings on these principal components were scaled using the inverse of the square root of eigenvalues, thus effectively whitening their contributions.

Fixation-pattern similarity analysis

FPSA was conducted on single participants. Condition specific FDMs (8 faces per baseline and generalization phases) were computed by collecting all fixations across trials on a single map which was then normalized to unit sum. We corrected FDMs by removing the common mean pattern (done separately for baseline and generalization phases). We used 1 - Pearson correlation as the similarity metric. This resulted in a 16x16 similarity matrix per subject. Statistical tests for element-wise comparison of the similarity values were conducted after Fisher transformation of correlation values. The multidimensional scaling was conducted on the baseline and generalization phases jointly using the 16x16 similarity matrix as input (*mdscale* in MATLAB). Importantly, as the similarity metric is extremely sensitive to the signal to noise ratio [34] present in the FDMs, we took precautions that the number of trials between generalization and baseline phases were exactly the same in order to avoid differences that would have been caused by different signal to noise ratios. To account for

unequal number of trials during the baseline (11 repetitions) and generalization (3 runs x 11 = 33 repetitions) phases, we computed a similarity matrix for each run separately in the generalization phase. These were later averaged across runs for a given participant. This ensured that FDMs of the baseline and generalization phases had comparable signal-to-noise ratios, therefore not favoring the generalization phase for having more trials.

We generated 3 different models based on a quadrature decomposition of a circular similarity matrix. A circular similarity matrix of 8x8 can be obtained using the term $\mathbf{M}\otimes\mathbf{M}$, where $\mathbf{M}$ is a 8x2 matrix in form of $[\cos(x)\ \sin(x)]$, and the operator $\otimes$ denotes the outer product. x represents angular distances from the CS+ face, is equal to 0 for CS+ and π for CS–. Therefore, while $\cos(x)$ is symmetric around the CS+ face, $\sin(x)$ is shifted by 90°. For the bottom-up saliency and increased arousal models (Fig 1B and C) we used $M\otimes M$ as a predictor together with a constant intercept. For the tuned exploration model depicted in Fig 1D, we used $\cos(x)\otimes\cos(x)$ and $\sin(x)\otimes\sin(x)$ to independently model ellipsoid expansion along the specific and unspecific directions, respectively. Together with the intercept this model comprised 3 predictors. Finally the aversive generalization model (Fig 1E) was created using the predictors of the tuned exploration model in conjunction with a two-dimensional Gaussian centered on the CS+ face (in total 4 predictors). We tested different widths for the Gaussian and took the one that resulted in the best fit. This was equal to 65° of FWHM and similar to the values we observed for univariate explicit ratings and SCR responses.

All linear modeling was conducted using non-redundant, vectorized forms of the symmetric dissimilarity matrices. For a 8x8 dissimilarity matrix this resulted in a vector of 28 entries. Different models were fitted as mixed-effects, where intercept and slope contributed both as fixed- and random-effects (*fitlme* in Matlab). We selected mixed-effect models as these performed better than models defined uniquely with fixed-effects on intercept and slope. To do model selection, we used Bayesian information criterion (BIC) as it compensates for an increase in the number of predictors between different models. Additionally, different models were also fitted to single participants (*fitlm* in Matlab) and the parameter estimates were separately tested for significance using *t-test*.

Acknowledgements

The authors wish to thank Tim Kitzmann for his input on an early version of this manuscript, Cliodhna Quigley for proof-reading, Helen Blank for comments, Patricia Billaudelle and Katrin Harland for their assistance with data collection. This research is supported by the DFG SFB TRR 58.

References

1. Pavlov I. Conditioned reflexes an investigation of the physiological activity of the cerebral cortex. [London]: Oxford University Press Humphrey Milford; 1927.

2. Tenenbaum JB, Griffiths TL. Generalization, similarity, and Bayesian inference. Behav Brain Sci. 2001;24: 629–640. doi:10.1017/S0140525X01000061

3. Ghirlanda S, Enquist M. A century of generalization. Anim Behav. 2003;66: 15–36. doi:10.1006/anbe.2003.2174

4. Guttman N, Kalish HI. Discriminability and stimulus generalization. J Exp Psychol. 1956;51: 79–88.

5. Shepard RN. Toward a universal law of generalization for psychological science. Science. 1987;237: 1317–1323.

6. Bass MJ, Hull CL. The irradiation of a tactile conditioned reflex in man. J Comp Psychol. 1934;17: 47–65.

7. Dunsmoor JE, Mitroff SR, LaBar KS. Generalization of conditioned fear along a dimension of increasing fear intensity. Learn Mem Cold Spring Harb N. 2009;16: 460–469. doi:10.1101/lm.1431609

8. Resnik J, Sobel N, Paz R. Auditory aversive learning increases discrimination thresholds. Nat Neurosci. 2011;14: 791–796. doi:10.1038/nn.2802

9. Li W, Howard JD, Parrish TB, Gottfried JA. Aversive Learning Enhances Perceptual and Cortical Discrimination of Indiscriminable Odor Cues. Science. 2008;319: 1842–1845. doi:10.1126/science.1152837

10. Onat S, Büchel C. The neuronal basis of fear generalization in humans. Nat Neurosci. 2015;advance online publication. doi:10.1038/nn.4166

11. Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci. 2001;2: 194–203.

12. Henderson JM. Human gaze control during real-world scene perception. Trends Cogn Sci. 2003;7: 498–504.

13. Yarbus AL. Eye movements and vision. New York: Plenum Press; 1967.

14. Schumann F, Einhäuser-Treyer W, Vockeroth J, Bartl K, Schneider E, König P. Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. J Vis. 2008;8: 12.1–17. doi:10.1167/8.14.12

15. Peterson MF, Lin J, Zaun I, Kanwisher N. Individual differences in face-looking behavior generalize from the lab to the world. J Vis. 2016;16: 12. doi:10.1167/16.7.12

16. Hayhoe M, Ballard D. Eye movements in natural behavior. Trends Cogn Sci. 2005;9: 188–194. doi:10.1016/j.tics.2005.02.009

17. Dowd EW, Mitroff SR, LaBar KS. Fear generalization gradients in visuospatial attention. Emot Wash DC. 2016;16: 1011–1018. doi:10.1037/emo0000197

18. Cerf M, Frady EP, Koch C. Faces and text attract gaze independent of the task: Experimental data and computer model. J Vis. 2009;9: 10.1-15. doi:10.1167/9.12.10

19. End A, Gamer M. Preferential Processing of Social Features and Their Interplay with Physical Saliency in Complex Naturalistic Scenes. Front Psychol. 2017;8. doi:10.3389/fpsyg.2017.00418

20. Peterson MF, Eckstein MP. Looking just below the eyes is optimal across face recognition tasks. Proc Natl Acad Sci. 2012;109: E3314–E3323. doi:10.1073/pnas.1214269109

21. Jack RE, Garrod OGB, Schyns PG. Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. Curr Biol. 2014;24: 187–192. doi:10.1016/j.cub.2013.11.064

22. Adolphs R. Fear, faces, and the human amygdala. Curr Opin Neurobiol. 2008;18: 166–172. doi:10.1016/j.conb.2008.06.006

23. Dunsmoor JE, Prince SE, Murty VP, Kragel PA, LaBar KS. Neurobehavioral mechanisms of human fear generalization. NeuroImage. 2011;55: 1878–1888. doi:10.1016/j.neuroimage.2011.01.041

24. Kriegeskorte N, Mur M, Bandettini P. Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. Front Syst Neurosci. 2008;2. doi:10.3389/neuro.06.004.2008

25. Benson PJ, Beedie SA, Shephard E, Giegling I, Rujescu D, St. Clair D. Simple Viewing Tests Can Detect Eye Movement Abnormalities That Distinguish Schizophrenia Cases from Controls with Exceptional Accuracy. Biol Psychiatry. 2012;72: 716–724. doi:10.1016/j.biopsych.2012.04.019

26. Arizpe J, Kravitz DJ, Yovel G, Baker CI. Start Position Strongly Influences Fixation Patterns during Face Processing: Difficulties with Eye Movements as a Measure of Information Use. PLoS ONE. 2012;7: e31106. doi:10.1371/journal.pone.0031106

27. Henderson JM, Shinkareva SV, Wang J, Luke SG, Olejarczyk J. Predicting cognitive state from eye movements. 2013; Available: http://dx.plos.org/10.1371/journal.pone.0064937

28. Zelinsky GJ, Peng Y, Samaras D. Eye can read your mind: Decoding gaze fixations to reveal categorical search targets. J Vis. 2013;13: 10.

29. Mehoudar E, Arizpe J, Baker CI, Yovel G. Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. J Vis. 2014;14: 6.

30. Borji A, Itti L. Defending Yarbus: Eye movements reveal observers' task. J Vis. 2014;14: 29–29. doi:10.1167/14.3.29

31. Kanan C, Bseiso DN, Ray NA, Hsiao JH, Cottrell GW. Humans have idiosyncratic and task-specific scanpaths for judging faces. Vision Res. 2015;108: 67–76.

32. Walker-Smith GJ, Gale AG, Findlay JM. Eye movement strategies involved in face perception. Perception. 1977;6: 313–326.

33. Coutrot A, Binetti N, Harrison C, Mareschal I, Johnston A. Face exploration dynamics differentiate men and women. J Vis. 2016;16: 16–16. doi:10.1167/16.14.16

34. Diedrichsen J, Ridgway GR, Friston KJ, Wiestler T. Comparing the similarity and spatial structure of neural representations: A pattern-component model. NeuroImage. 2011;55: 1665–1678. doi:10.1016/j.neuroimage.2011.01.044

35. Masciocchi CM, Mihalas S, Parkhurst D, Niebur E. Everyone knows what is interesting: salient locations which should be fixated. J Vis. 2009;9: 25.1–22. doi:10.1167/9.11.25

36. Reynolds JH, Desimone R. Interacting roles of attention and visual salience in V4. Neuron. 2003;37: 853–863.

37. Ohl FW, Scheich H, Freeman WJ. Change in pattern of ongoing cortical activity with auditory category learning. Nature. 2001;412: 733–736. doi:10.1038/35089076

38. Dunsmoor JE, Murphy GL. Categories, concepts, and conditioning: how humans generalize fear. Trends Cogn Sci. 2015;19: 73–77. doi:10.1016/j.tics.2014.12.003

39. Dunsmoor JE, Kragel PA, Martin A, LaBar KS. Aversive Learning Modulates Cortical Representations of Object Categories. Cereb Cortex. 2014;24: 2859–2872. doi:10.1093/cercor/bht138

40. Vervoort E, Vervliet B, Bennett M, Baeyens F. Generalization of Human Fear Acquisition and Extinction within a Novel Arbitrary Stimulus Category. PLoS ONE. 2014;9. doi:10.1371/journal.pone.0096569

41. Qu LP, Kahnt T, Cole SM, Gottfried JA. De Novo Emergence of Odor Category Representations in the Human Brain. J Neurosci. 2016;36: 468–478. doi:10.1523/JNEUROSCI.3248-15.2016

42. Vervliet B, Geens M. Fear generalization in humans: Impact of feature learning on conditioning and extinction. Neurobiol Learn Mem. 2014;113: 143–148. doi:10.1016/j.nlm.2013.10.002

43. Norrholm SD, Jovanovic T, Briscione MA, Anderson KM, Kwon CK, Warren VT, et al. Generalization of fear-potentiated startle in the presence of auditory cues: a parametric analysis. Front Behav Neurosci. 2014;8. doi:10.3389/fnbeh.2014.00361

44. Yue X, Biederman I, Mangini MC, Malsburg C von der, Amir O. Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. Vision Res. 2012;55: 41–46. doi:10.1016/j.visres.2011.12.012

45. Malcolm GL, Lanyon LJ, Fugard AJB, Barton JJS. Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. J Vis. 2008;8: 2–2. doi:10.1167/8.8.2

46. Schurgin MW, Nelson J, Iida S, Ohira H, Chiao JY, Franconeri SL. Eye movements during emotion recognition in faces. J Vis. 2014;14: 14–14. doi:10.1167/14.13.14

47. Lissek S, Rabin S, Heller RE, Lukenbaugh D, Geraci M, Pine DS, et al. Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. Am J Psychiatry. 2009;167: 47–55. doi:10.1176/appi.ajp.2009.09030410

48. Cha J, Greenberg T, Carlson JM, DeDora DJ, Hajcak G, Mujica-Parodi LR. Circuit-Wide Structural and Functional Measures Predict Ventromedial Prefrontal Cortex Fear Generalization: Implications for Generalized Anxiety Disorder. J Neurosci. 2014;34: 4043–4053. doi:10.1523/JNEUROSCI.3372-13.2014

49. Laufer O, Israeli D, Paz R. Behavioral and Neural Mechanisms of Overgeneralization in Anxiety. Curr Biol. 2016;26: 713–722. doi:10.1016/j.cub.2016.01.023

50. Visser RM, Scholte HS, Beemsterboer T, Kindt M. Neural pattern similarity predicts long-term fear memory. Nat Neurosci. 2013;16: 388–390. doi:10.1038/nn.3345

51. Armann R, Bülthoff I. Gaze behavior in face comparison: The roles of sex, task, and symmetry. Atten Percept Psychophys. 2009;71: 1107–1126. doi:10.3758/APP.71.5.1107

52. Haji-Abolhassani A, Clark JJ. A computational model for task inference in visual search. J Vis. 2013;13: 29.

53. Itti L. New Eye-Tracking Techniques May Revolutionize Mental Health Screening. Neuron. 2015;88: 442–444. doi:10.1016/j.neuron.2015.10.033

54. Kietzmann TC, Gert AL, Tong F, König P. Representational Dynamics of Facial Viewpoint Encoding. J Cogn Neurosci. 2017;29: 637–651. doi:10.1162/jocn_a_01070

55. Arizpe JM, Walsh V, Baker CI. Characteristic visuomotor influences on eye-movement patterns to faces and other high level stimuli. Front Psychol. 2015;6. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4518262/

56. Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J Neurosci. 1998;18: 411–418.

57. Bannerman RL, Milders M, Gelder B de, Sahraie A. Orienting to threat: faster localization of fearful facial expressions and body postures revealed by saccadic eye movements. Proc R Soc Lond B Biol Sci. 2009;276: 1635–1641. doi:10.1098/rspb.2008.1744

58. Lin JY, Murray SO, Boynton GM. Capture of Attention to Threatening Stimuli without Perceptual Awareness. Curr Biol. 2009;19: 1118–1122. doi:10.1016/j.cub.2009.05.021

59. König P, Wilming N, Kietzmann TC, Ossandón JP, Onat S, Ehinger BV, et al. Eye movements as a window to cognitive processes. J Eye Mov Res. 2016;9. doi:10.16910/jemr.9.5.3

60. Lissek S, Biggs AL, Rabin SJ, Cornwell BR, Alvarez RP, Pine DS, et al. Generalization of conditioned fear-potentiated startle in humans: experimental validation and clinical relevance. Behav Res Ther. 2008;46: 678–687. doi:10.1016/j.brat.2008.02.005

61. Greenberg T, Carlson JM, Cha J, Hajcak G, Mujica-Parodi LR. Neural reactivity tracks fear generalization gradients. Biol Psychol. 2013;92: 2–8. doi:10.1016/j.biopsycho.2011.12.007

62. Wilming N, Onat S, Ossandón JP, Açık A, Kietzmann TC, Kaspar K, et al. An extensive dataset of eye movements during viewing of complex images. Sci Data. 2017;4: 160126. doi:10.1038/sdata.2016.126

63. Onat S, Kampermann L. FPSA_FearGen: A Github repository for the analysis and preparation of "Aversive Learning Changes Face-Viewing Strategies—as Revealed by Model-Based Fixation-Pattern Similarity Analysis". [Internet]. 2017. Available: https://github.com/selimonat/FPSA_FearGen.git

64. Blakemore C, Campbell FW. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. J Physiol. 1969;203: 237–260.

65. Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997;10: 433–436.

66. Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis. 1997;10: 437–442.

67. Buracas GT, Boynton GM. Efficient design of event-related fMRI experiments using M-sequences. NeuroImage. 2002;16: 801–813.

68. Liu TT, Frank LR. Efficiency, power, and entropy in event-related FMRI with multiple trial types: Part I: theory. NeuroImage. 2004;21: 387–400. doi:10.1016/j.neuroimage.2003.09.030

69. Watson AB, Pelli DG. QUEST: a Bayesian adaptive psychometric method. Percept Psychophys. 1983;33: 113–120.

70. Benedek M, Kaernbach C. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology. 2010;47: 647–658. doi:10.1111/j.1469-8986.2009.00972.x

71. Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. J Neurosci Methods. 2010;190: 80–91. doi:10.1016/j.jneumeth.2010.04.028

72. Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol TIST. 2011;2: 27.

Author contributions

Conceptualization: SO, CB

Data Curation: SO

Formal Analysis: LK, NW, SO

Funding Acquisition: CB

Investigation: LK, SO

Methodology: LK, NW, SO

Project Administration: SO

Resources: CB

Software: LK, SO

Supervision: SO

Validation: LK, NW, AA, CB, SO

Visualization: LK, SO

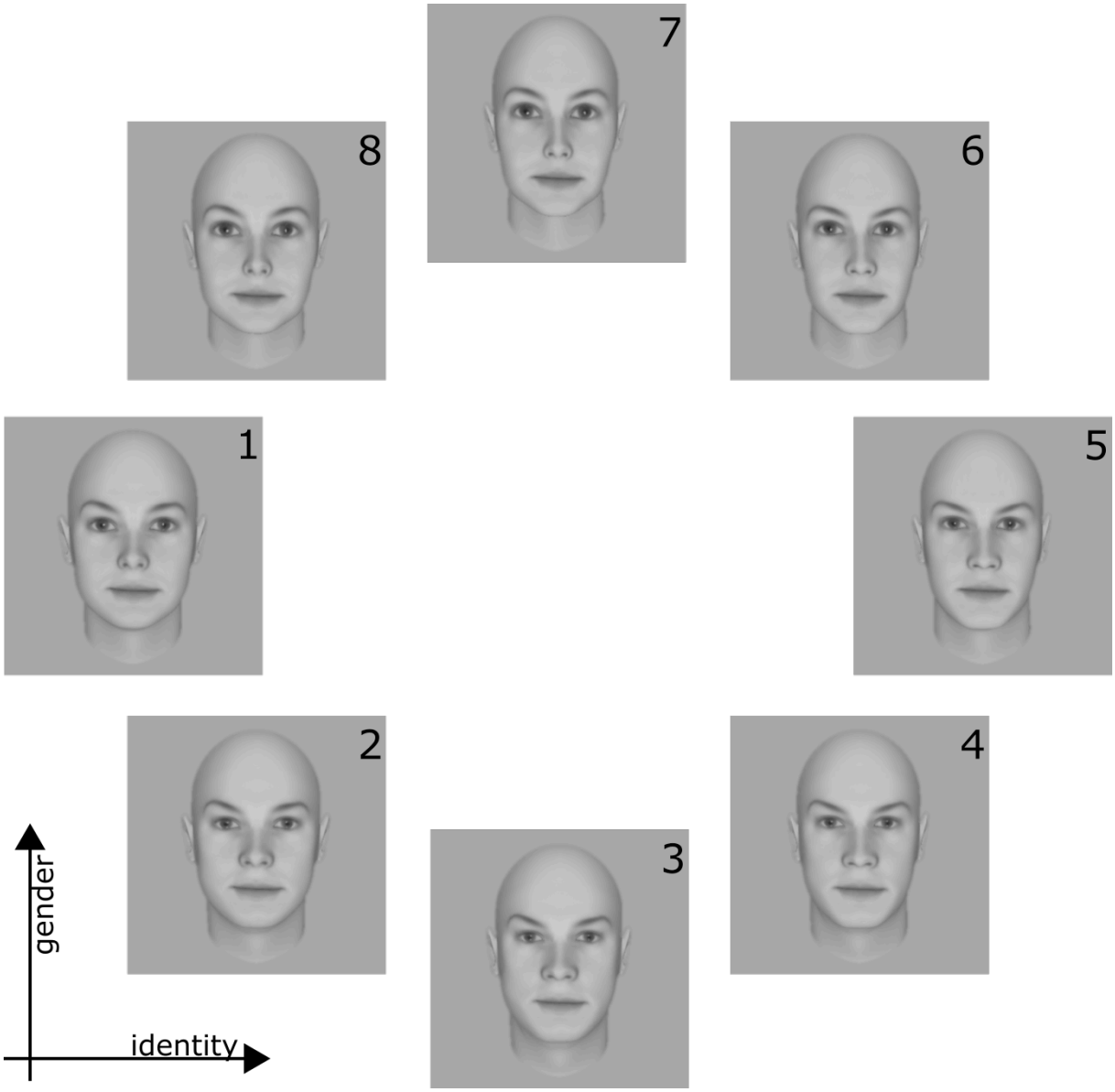
Writing - Original Draft Preparation: LK, SO

Writing – Review, Editing: LK, NW, AA, CB, SO

Competing financial interest

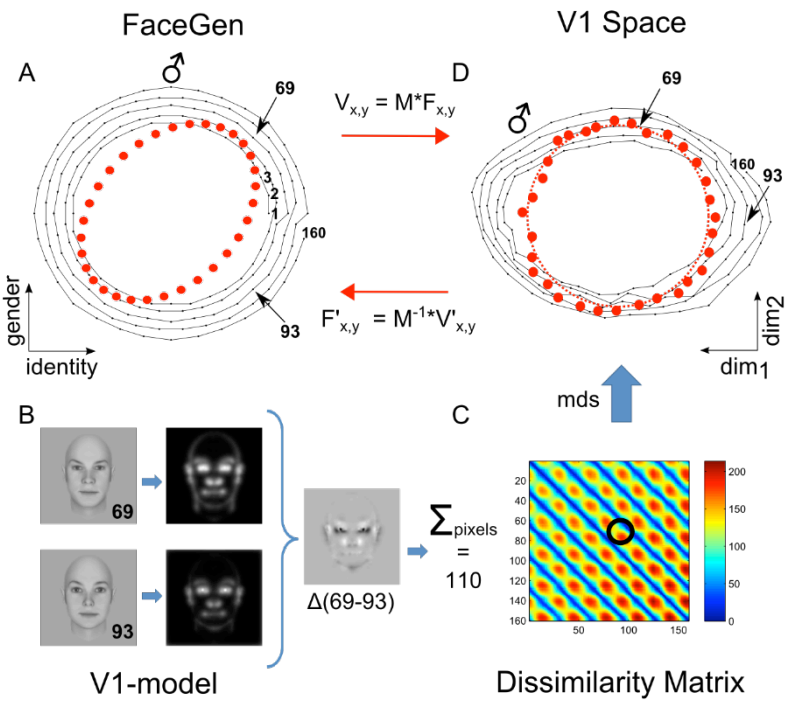
The authors declare no competing financial interests.

Supporting Information



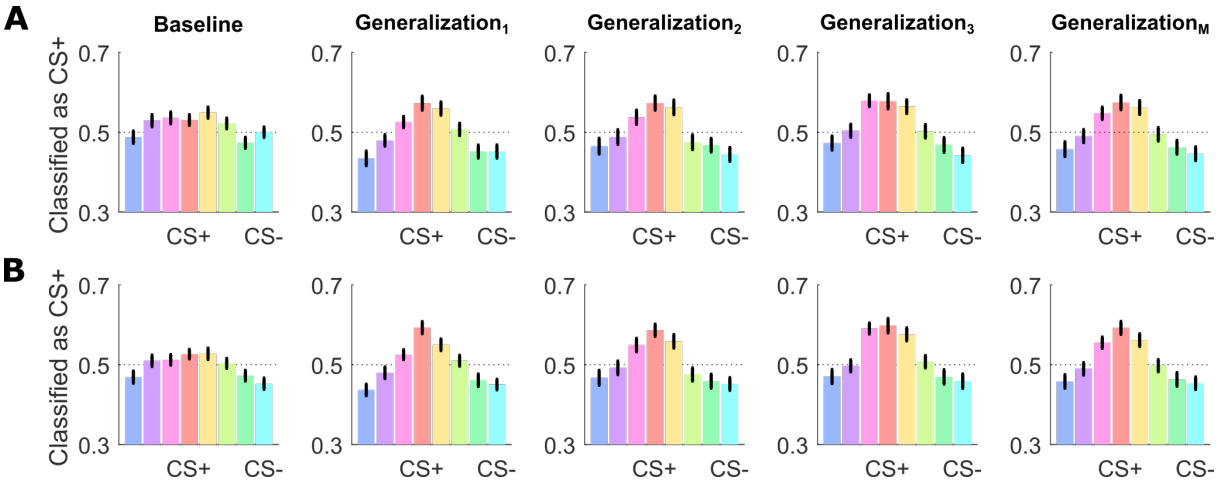
SFig 1

Face Stimuli. Set of 8 faces that were calibrated to form a circular similarity continuum. Faces vary along the two dimensions of gender (vertical axis) and identity (horizontal axis). See SFig 2 for the calibration process.



SFig 2

Calibration of faces using a V1 model tuned to human psychophysics. (A) Using the FaceGen software, 160 faces forming five concentric circles were generated with coordinates varying in gender and identity dimensions (connected black dots in the left panel). Maximally male faces are located at 12 o'clock direction and indicated with the male symbol. (B) V1 representations of faces were modelled according to [44]. This is illustrated for faces 69 and 93. The difference between these two faces resulted in a Euclidean distance of 110. The pair-wise Euclidean distance for all the 160 faces are shown in (C) as a dissimilarity matrix. The resulting dissimilarity matrix exhibits 5 major bands corresponding to 5 concentric circles. By applying MDS, we obtained the representational space of V1 shown in (A, right panel). Note that the most male face is 45° counter-clockwise rotated with respect to the main axes of in V1 representation. The mapping between FaceGen coordinates and V1 representational space thus involved a rotation and scaling which was captured by the matrix *M*. We therefore used the inverse of *M*, to achieve coordinates of perfect circularity based on this V1 model. This ensured that faces along the similarity continuum were characterized by controlled changes for every angular step based on the model used.



SFig 3

Multivariate Classification of FDMs. Classification results for a linear SVM trained to differentiate between CS+ and CS–. Bars (M ± SEM) show the average proportion of trials of all 8 conditions classified as CS+, i.e. correct classification for the actual CS+ (*red*) condition and false alarms for CS– trials (*cyan*). Baseline and three generalization runs are plotted separately. The last column ($Generalization_M$) depicts the average across the three generalization runs. Dotted lines indicate chance level obtained by classification of random label permutation. Results are depicted based on two different dimension reduction methods (see Material and Methods): (A) Classification results based on *N* eigenvectors explaining 90% of total variance. (B) Results based on elbow criterion, i.e. choosing N eigenvectors where the slope of eigenvalues levelled off substantially.

S1 Table

Mixed-effects modeling of the similarity matrices during the baseline phase with the bottom-up model shown in Fig 1B.

Model information:

Number of observations	1708
Fixed effects coefficients	2
Random effects coefficients	122
Covariance parameters	4

Formula:

FPSA_baseline ~ 1 + circle + (1 + circle | subject)

Model fit statistics:

AIC	*BIC*	*LogLikelihood*	*Deviance*
-1682.7	-1650	847.34	-1694.7

Fixed effects coefficients (95% CIs):

Name	*Estimate*	*SE*	*tStat*	*DF*	*pValue*	*Lower*	*Upper*
'(Intercept)'	0.24944	0.0037375	66.741	1706	0	0.24211	0.25677
'circle'	0.063345	0.0079565	7.9614	1706	3.0776e-15	0.047739	0.078951

Random effects covariance parameters (95% CIs):

Group: subject (61 Levels)

Name1	*Name2*	*Type*	*Estimate*	*Lower*	*Upper*
'(Intercept)'	'(Intercept)'	'std'	0.0076928	NaN	NaN
'circle'	'(Intercept)'	'corr'	NaN	NaN	NaN
'circle'	'circle'	'std'	0.044851	NaN	NaN

Group: Error

Name	*Estimate*	*Lower*	*Upper*
'Res Std'	0.14541	NaN	NaN

S2 Table

Mixed-effects modeling of the similarity matrices during the generalization phase with the arousal model shown in Fig 1C.

Model information:

Number of observations	1708
Fixed effects coefficients	2
Random effects coefficients	122
Covariance parameters	4

Formula:

FPSA_generalization ~ 1 + circle + (1 + circle | subject)

Model fit statistics:

AIC	*BIC*	*LogLikelihood*	*Deviance*
-2748.2	-2715.5	1380.1	-2760.2

Fixed effects coefficients (95% CIs):

Name	*Estimate*	*SE*	*tStat*	*DF*	*pValue*	*Lower*	*Upper*
'(Intercept)'	0.25395	0.0029798	85.222	1706	0	0.24811	0.25979
'circle'	0.091957	0.0099814	9.2129	1706	9.0388e-20	0.07238	0.11153

Random effects covariance parameters (95% CIs):

Group: subject (61 Levels)

Name1	*Name2*	*Type*	*Estimate*	*Lower*	*Upper*
'(Intercept)'	'(Intercept)'	'std'	0.011548	0.0071814	0.01857
'circle'	'(Intercept)'	'corr'	1	NaN	NaN
'circle'	'circle'	'std'	0.071586	0.057996	0.088361

Group: Error

Name	*Estimate*	*Lower*	*Upper*
'Res Std'	0.10434	0.10084	0.10797

S3 Table

Mixed-effects modeling of the similarity matrices during the baseline phase with the adversity categorization model shown in Fig 1D.

Model information:

Number of observations	1708
Fixed effects coefficients	3
Random effects coefficients	183
Covariance parameters	7

Formula:

FPSA_baseline ~ 1 + specific + unspecific + (1 + specific + unspecific | subject)

Model fit statistics:

AIC	*BIC*	*LogLikelihood*	*Deviance*
-1764	-1709.6	892.01	-1784

Fixed effects coefficients (95% CIs):

Name	*Estimate*	*SE*	*tStat*	*DF*	*pValue*	*Lower*	*Upper*
'(Intercept)'	0.24944	0.0035811	69.654	1705	0	0.24242	0.25646
'specific'	0.057755	0.013652	4.2304	1705	2.457e-05	0.030977	0.084532
'unspecific'	0.068935	0.010663	6.4647	1705	1.3216e-10	0.048021	0.08985

Random effects covariance parameters (95% CIs):

Group: subject (61 Levels)

Name1	*Name2*	*Type*	*Estimate*	*Lower*	*Upper*
'(Intercept)'	'(Intercept)'	'std'	0.0081486	0.0035081	0.018928
'specific'	'(Intercept)'	'corr'	0.81502	0.80893	0.82093
'unspecific'	'(Intercept)'	'corr'	0.30346	-0.0053757	0.55945
'specific'	'specific'	'std'	0.0913	0.071842	0.11603
'unspecific'	'specific'	'corr'	-0.30479	-0.56025	0.0036179
'unspecific'	'unspecific'	'std'	0.062467	0.045811	0.085177

Group: Error

Name	*Estimate*	*Lower*	*Upper*
'Res Std'	0.13817	0.13344	0.14306

S4 Table

Mixed-effects modeling of the similarity matrices during the generalization phase with the adversity categorization model shown in Fig 1D.

Model information:

Number of observations	1708
Fixed effects coefficients	3
Random effects coefficients	183
Covariance parameters	7

Formula:

FPSA_generalization ~ 1 + specific + unspecific + (1 + specific + unspecific | subject)

Model fit statistics:

AIC	*BIC*	*LogLikelihood*	*Deviance*
-2951.7	-2897.3	1485.9	-2971.7

Fixed effects coefficients (95% CIs):

Name	*Estimate*	*SE*	*tStat*	*DF*	*pValue*	*Lower*	*Upper*
'(Intercept)'	0.25395	0.0028004	90.683	1705	0	0.24846	0.25944
'specific'	0.1205	0.014472	8.3259	1705	1.6951e-16	0.09211	0.14888
'unspecific'	0.063418	0.010001	6.3412	1705	2.9123e-10	0.043803	0.083034

Random effects covariance parameters (95% CIs):

Group: subject (61 Levels)

Name1	*Name2*	*Type*	*Estimate*	*Lower*	*Upper*
'(Intercept)'	'(Intercept)'	'std'	0.011778	0.0077913	0.017805
'specific'	'(Intercept)'	'corr'	0.91907	NaN	NaN
'unspecific'	'(Intercept)'	'corr'	0.6935	NaN	NaN
'specific'	'specific'	'std'	0.10647	0.087673	0.12931
'unspecific'	'specific'	'corr'	0.35345	0.34746	0.35941
'unspecific'	'unspecific'	'std'	0.068277	0.054449	0.085617

Group: Error

Name	*Estimate*	*Lower*	*Upper*
'Res Std'	0.09517	0.091916	0.09854

S5 Table

Mixed-effects modeling of the similarity matrices during the baseline phase with the adversity tuning model shown in Fig 1E.

Model information:

Number of observations	1708
Fixed effects coefficients	4
Random effects coefficients	244
Covariance parameters	11

Formula:

FPSA_baseline ~ 1 + specific + unspecific + Gaussian + (1 + specific + unspecific + Gaussian | subject)

Model fit statistics:

AIC	*BIC*	*LogLikelihood*	*Deviance*
-1755.5	-1673.9	892.76	-1785.5

Fixed effects coefficients (95% CIs):

Name	*Estimate*	*SE*	*tStat*	*DF*	*pValue*	*Lower*	*Upper*
'(Intercept)'	0.24181	0.046873	5.1589	1704	2.7749e-07	0.14988	0.33374
'specific'	0.057483	0.013761	4.1774	1704	3.0984e-05	0.030494	0.084473
'unspecific'	0.069013	0.010773	6.4063	1704	1.9241e-10	0.047884	0.090142
'Gaussian'	0.015884	0.097484	0.16294	1704	0.87058	-0.17532	0.20709

Random effects covariance parameters (95% CIs):

Group: subject (61 Levels)

Name1	*Name2*	*Type*	*Estimate*	*Lower*	*Upper*
'(Intercept)'	'(Intercept)'	'std'	0.054362	0.022712	0.13012
'specific'	'(Intercept)'	'corr'	0.16475	NaN	NaN
'unspecific'	'(Intercept)'	'corr'	-0.98746	-0.9876	-0.98731
'Gaussian'	'(Intercept)'	'corr'	-0.99227	-0.99237	-0.99216
'specific'	'specific'	'std'	0.091417	0.07191	0.11622
'unspecific'	'specific'	'corr'	-0.31842	-0.32004	-0.31681
'Gaussian'	'specific'	'corr'	-0.041041	NaN	NaN
'unspecific'	'unspecific'	'std'	0.063537	0.046652	0.086532
'Gaussian'	'unspecific'	'corr'	0.96022	NaN	NaN
'Gaussian'	'Gaussian'	'std'	0.12196	0.054259	0.27414

Group: Error

Name	*Estimate*	*Lower*	*Upper*
'Res Std'	0.13805	0.13333	0.14293

S6 Table

Mixed-effects modeling of the similarity matrices during the generalization phase with the adversity tuning model shown in Fig 1E.

Model information:

Number of observations	1708
Fixed effects coefficients	4
Random effects coefficients	244
Covariance parameters	11

Formula:

FPSA_generalization ~ 1 + specific + unspecific + Gaussian + (1 + specific + unspecific + Gaussian | subject)

Model fit statistics:

AIC	*BIC*	*LogLikelihood*	*Deviance*
-2945.9	-2864.2	1487.9	-2975.9

Fixed effects coefficients (95% CIs):

Name	*Estimate*	*SE*	*tStat*	*DF*	*pValue*	*Lower*	*Upper*
'(Intercept)'	0.22	0.032419	6.786	1704	1.5856e-11	0.15641	0.28359
'specific'	0.11929	0.014301	8.3414	1704	1.4959e-16	0.091239	0.14734
'unspecific'	0.063765	0.010056	6.3412	1704	2.913e-10	0.044042	0.083488
'Gaussian'	0.070672	0.067933	1.0403	1704	0.29834	-0.062568	0.20391

Random effects covariance parameters (95% CIs):

Group: subject (61 Levels)

Name1	*Name2*	*Type*	*Estimate*	*Lower*	*Upper*
'(Intercept)'	'(Intercept)'	'std'	0.044693	0.010711	0.18649
'specific'	'(Intercept)'	'corr'	-0.89633	-0.89679	-0.89586
'unspecific'	'(Intercept)'	'corr'	-0.73005	-0.73122	-0.72888
'Gaussian'	'(Intercept)'	'corr'	-0.99995	-0.99996	-0.99995
'specific'	'specific'	'std'	0.1047	0.08608	0.12734
'unspecific'	'specific'	'corr'	0.35136	0.35001	0.3527
'Gaussian'	'specific'	'corr'	0.9005	0.90014	0.90086
'unspecific'	'unspecific'	'std'	0.068748	0.054885	0.086113
'Gaussian'	'unspecific'	'corr'	0.72353	0.7229	0.72415
'Gaussian'	'Gaussian'	'std'	0.11757	0.03792	0.36454

Group: Error

Name	*Estimate*	*Lower*	*Upper*
'Res Std'	0.095034	0.091784	0.098399