Neuroscience and Biobehavioral Reviews 142 (2022) 104918

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

# Pattern analysis of neuroimaging data reveals novel insights on threat learning and extinction in humans

Augustin C. Hennings[a,b,c], Samuel E. Cooper[c], Jarrod A. Lewis-Peacock[b,c,d,e], Joseph E. Dunsmoor[b,c,d,*]

[a] *Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA*
[b] *Institute for Neuroscience, University of Texas at Austin, Austin, TX, USA*
[c] *Department of Psychiatry and Behavioral Sciences, Dell Medical School, University of Texas at Austin, Austin, TX, USA*
[d] *Center for Learning and Memory, Department of Neuroscience, University of Texas at Austin, Austin, TX, USA*
[e] *Department of Psychology, University of Texas at Austin, Austin, TX, USA*

ARTICLE INFO

*Keywords:*
Fear
Extinction
FMRI
Amygdala
Hippocampus
MPFC
MVPA
Univariate

ABSTRACT

Several decades of rodent neurobiology research have identified a network of brain regions that support Pavlovian threat conditioning and extinction, focused predominately on the amygdala, hippocampus, and medial prefrontal cortex (mPFC). Surprisingly, functional magnetic resonance imaging (fMRI) studies have shown inconsistent evidence for these regions while humans undergo threat conditioning and extinction. In this review, we suggest that translational neuroimaging efforts have been hindered by reliance on traditional univariate analysis of fMRI. Whereas univariate analyses average activity across voxels in a given region, multivariate pattern analyses (MVPA) leverage the information present in spatial patterns of activity. MVPA therefore provides a more sensitive analysis tool to translate rodent neurobiology to human neuroimaging. We review human fMRI studies using MVPA that successfully bridge rodent models of amygdala, hippocampus, and mPFC function during Pavlovian learning. We also highlight clinical applications of these information-sensitive multivariate analyses. In sum, we advocate that the field should consider adopting a variety of multivariate approaches to help bridge cutting-edge research on the neuroscience of threat and anxiety.

## 1. Introduction

More precise understanding of the neurobehavioral mechanisms of threat learning and regulation will ultimately improve treatments for psychiatric disorders characterized by excessive fear and anxiety. For over a century, Pavlovian conditioning and extinction have served as widely used models for investigating threat learning processes across species (Haaker et al., 2019; LeDoux, 2000). Neurobiological research in rodent models has revealed the associated microcircuitry in a canonical "threat network", a collection of key regions supporting both the acquisition and extinction of conditioned threat centered predominantly on the amygdala, hippocampus, and subdivisions of the medial prefrontal cortex (mPFC) (Bouton et al., 2021; Tovote et al., 2015). This network has also been identified in humans – both through early lesion studies (LaBar et al., 1995) and early functional neuroimaging (LaBar et al., 1998). However, the reliability of this translation from neurophysiology in rodents to neuroimaging in humans has recently come into question (Fullana et al., 2018; Visser et al., 2021). Specifically, functional magnetic resonance imaging (fMRI) studies have not found consistent engagement of the threat network during threat acquisition and subsequent extinction (Fullana et al., 2016, 2018; Sehlmeyer et al., 2009) (but see Sjouwerman et al., 2020; Wen et al., 2022). This inconsistency likely stems from the primary analytical approach used to analyze fMRI data of human threat learning, namely mass univariate analysis. In this brief review, we detail the limits of the univariate approach and describe how advances in multivariate pattern analysis (MVPA) of fMRI data can overcome these limitations and facilitate a more precise understanding of the neural computations that underlie threat and extinction in humans.

## 2. Limitations of univariate fMRI for the study of human threat learning

Typical human threat learning experiments include a minimum of

[*] Corresponding author at: Institute for Neuroscience, University of Texas at Austin, Austin, TX, USA.
*E-mail address:* joseph.dunsmoor@austin.utexas.edu (J.E. Dunsmoor).

https://doi.org/10.1016/j.neubiorev.2022.104918
Received 16 July 2022; Received in revised form 12 October 2022; Accepted 13 October 2022
Available online 15 October 2022

two types of conditioned stimuli (CSs): a CS paired with an aversive unconditioned stimulus (US), referred to as CS+ , and a CS unpaired with the US that serves as a within-subjects control, referred as the CS- (Lonsdorf et al., 2017). The traditional fMRI experiment of threat learning focuses on univariate differences (i.e., one voxel at a time) in the blood-oxygenation-level-dependent (BOLD) signal between these CS types. These differences are measured by contrasting the average BOLD signal in each voxel for CS+ trials versus CS- trials. The results are then spatially smoothed and statistically threshold to identify the strongest, most reliable responses. At the whole-brain level, univariate approaches consistently reveal stronger responses to the CS+ versus the CS- during threat acquisition in the thalamus, midbrain, anterior insula, and dorsal anterior cingulate cortex (dACC) (Fullana et al., 2016), areas that overlap with the putative salience network of brain regions that mediate responses to salient environmental cues (Seeley, 2019). Notably absent, however, are regions of the canonical threat network constituting the focus of rodent neuroscience (e.g., the amygdala). Likewise, meta-analyses of extinction learning fail to find consistent engagement of the amygdala or ventromedial prefrontal cortex (vmPFC), as would be expected based on the rodent literature (Bouton et al., 2021), but instead find activity in the same salience network regions that are active during threat learning (Fullana et al., 2018).

Considering the absence of consistent threat network activity in human fMRI, one proposal is that the nature of human threat paradigms are not aversive or threatening enough (e.g., a mild electrical shock applied to a finger) to engage the same neurocircuitry as in rodents. Another possibility is that these regions simply do not serve the same functional role in threat acquisition and extinction in humans as they do in rodents. For example, it may be that cognitive processes unique to humans provide top-down signals that alter the role of the threat network during aversive learning (LeDoux and Pine, 2016). But perhaps a more parsimonious explanation is that univariate fMRI analyses are not sensitive enough to detect differential responses between CS+ and CS- stimuli in these regions. This is because the standard univariate approach reduces signal by ignoring voxels with weaker (i.e., non-significant) responses that nevertheless might carry diagnostic information about threat processing.

A critical distinction between univariate analyses and MVPA is that the latter considers the unique informational content present *across* a set of voxels (for review, see Cohen et al., 2017; Haxby et al., 2014; Lewis-Peacock and Norman, 2014). Like conventional methods, the MVPA approach seeks to boost sensitivity by looking at the contributions of multiple voxels. However, to avoid signal loss inherent in the univariate approach, MVPA treats each voxel as a distinct source of information and aggregates this (possibly weak) information across voxels to derive a more precise measurement of the neural response. Ultimately, multivariate approaches allow for both higher sensitivity in detecting neural signals related to threat processing, as well as higher specificity in distinguishing these signals from other cognitive processes (Reddan and Wager, 2018).

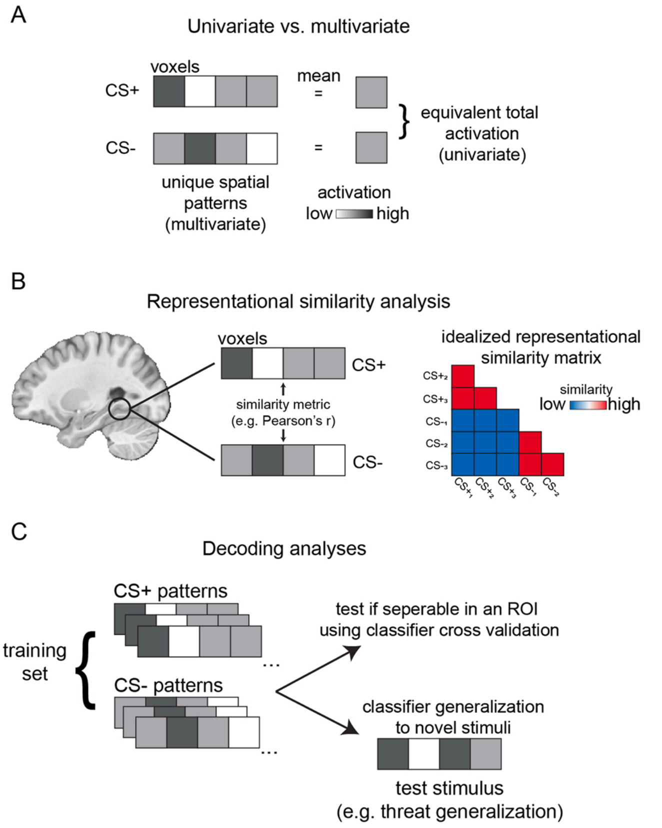
A
Univariate vs. multivariate
voxels
CS+
CS-
mean
=
=
equivalent total activation (univariate)
unique spatial patterns (multivariate)
activation
low
high
B
Representational similarity analysis
voxels
CS+
CS-
similarity metric (e.g. Pearson's r)
idealized representational similarity matrix
similarity
low
high
CS+₂
CS+₃
CS-₁
CS-₂
CS-₃
CS+₁
CS+₂
CS+₃
CS-₁
CS-₂
C
Decoding analyses
training set
CS+ patterns
CS- patterns
...
...
test if seperable in an ROI using classifier cross validation
classifier generalization to novel stimuli
test stimulus (e.g. threat generalization)

**Fig. 1.** Overview and applications of MVPA. A. **Analytic approaches of univariate and multivariate analyses.** A univariate analysis considers the mean activation in a single voxel over time, or in the case of region of interest analyses, the mean activation across voxels. In many regions, CS+ and CS- stimuli do not differ in total mean activation. Multivariate analyses are concerned with the information present in the pattern of activation across voxels in a given region. In this schematic, while the CS+ and CS- stimuli have equivalent mean activation across these 4 voxels, they each have a unique spatial pattern of activity. B. **Representational similarity analysis (RSA).** In RSA, the multivariate patterns of activity for each stimulus are compared to each other using a similarity metric, such as Pearson's r. The results can be visualized in a representational similarity matrix. C. **Decoding analyses**. As with RSA, decoding analyses rely on the unique spatial patterns of activity for different stimuli. In a typical decoding analysis, multiple presentations of stimuli from different categories are submitted to a linear classifier as training data. Using a cross-validation procedure, one approach is to test whether the stimuli from different categories are separable in a given brain region. Another approach is to use the trained classifier to attempt to decode a novel stimulus, such as in threat generalization.

There are two principal applications of MVPA methods to fMRI data (Fig. 1). One approach, referred to as representational similarity analysis (RSA) (Kriegeskorte et al., 2008), quantifies the correlation of activity patterns across collections of voxels. This produces a metric of the similarity structure for pattern-level information in each brain region elicited by different experimental conditions, allowing inferences on how the region represents a stimulus. The other principal application of MVPA involves decoding methods. Decoding analyses are versatile: they can be used to determine if two different neural states, represented by a spatially distributed patterns of activity across voxels, are separable or distinct, and they can be used to attempt to predict the content of an unknown brain state using a decoded state as a reference (Norman et al., 2006). Given the recent proliferation of machine learning techniques and their diverse applications in neuroscience, it is worth noting that the decoders discussed here are relatively simple supervised linear classifiers (e.g., linear support vector machines or logistic regression) as opposed to more complex algorithms (e.g., neural networks). In the specific case of fMRI data, linear classifiers exhibit better generalizability by reducing overfitting to small training sets, and they provide more readily interpretable mappings between neural features and classification outputs (Pereira et al., 2009). The remainder of this review examines how researchers have leveraged these multivariate methods to better identify the neural substrates of threat and extinction processing in humans.

## 3. MVPA reveals the circuits for threat acquisition

### 3.1. Threat network

Strong evidence for the critical role of the amygdala and hippocampus in rodent threat acquisition models have not been reliably translated to humans using univariate fMRI (Fullana et al., 2016). MVPA efforts, however, have yielded results that are more closely aligned with rodent models. Using a decoding analysis, Bach et al. (2011) first reported distinct CS+ and CS- activity patterns in the amygdala during threat acquisition. A series of studies utilizing RSA corroborated and expanded on these results (Visser et al., 2016; Visser et al., 2013, 2015), by comparing the similarity of neural responses to multiple visual categories (e.g., faces and scenes) of CS+ and CS- stimuli. In these studies, the amygdala exhibited increased similarity for CS+ stimuli compared to CS- stimuli, as well as increased similarity between the two categories of CS+ stimuli as compared to the similarity between the two categories of CS- stimuli. Notably, multivariate indicators of amygdala activity have been linked to behavioral expressions of conditioned learning (e.g., skin conductance, Bach et al., 2011; pupillary responses, Visser et al., 2013; and threat expectancy, Braem et al., 2017). It is worth noting, however, that these findings in the amygdala have not consistently replicated (Visser et al., 2016), even across studies that employ similar paradigms (Visser et al., 2011). Nevertheless, the identification of threat-related neural responses in the amygdala using multivariate analysis contradicts numerous null findings that have relied on univariate analysis alone.

The role of the hippocampus in threat acquisition is nuanced (Pape and Pare, 2010). Although hippocampus is necessary for the formation of contextual fear memories in rodents, it may not be crucial for learning simple cued fear associations (Phillips and LeDoux, 1992). Univariate fMRI analyses in humans often report hippocampal *deactivation* (i.e. CS- > CS+) during threat acquisition (Fullana et al., 2016), indicating some degree of task-related hippocampal processing. This deactivation has been interpreted as reflecting the processing of the safety value of the CS-, or alternatively as reflecting a "resting-state" like response given the hippocampus is part of the "default mode network" (Raichle, 2015). However, this interpretation of the hippocampal role in threat processing is not consistent with rodent work, which has successfully parsed the individual roles of subfields across the long axis of the hippocampus (Lacagnina et al., 2019; Meyer et al., 2019; Qin et al., 2021). Echoing this approach, our lab recently found that the posterior hippocampus showed selective reinstatement of threat conditioning stimuli, while the anterior hippocampus showed selective reinstatement for threat extinction stimuli (Hennings et al., 2022). This study used an encoding-to-retrieval RSA approach adapted from neuroimaging of episodic memory (Ritchey et al., 2013), whereby the pattern of activity for each stimulus at encoding is correlated with the pattern of activity to the same stimulus at a memory retrieval test. This analysis can be considered a human analogue for activity-dependent labeling techniques used to observe engram reinstatement in rodents, and as such it provides a powerful approach to measure the integrity of associative memories in humans.

### 3.2. Salience network

In contrast to the inconsistent evidence for the engagement of the threat network (e.g., amygdala and hippocampus), univariate fMRI analyses have provided robust evidence for the involvement of the salience network in threat learning (see Seeley, 2019). In particular, the dACC and anterior insula are consistently activated (CS+ > CS-) during threat acquisition (Fullana et al., 2016; Sehlmeyer et al., 2009). MVPA has corroborated and expanded these findings. For example, neural response patterns in the dACC and insula have been shown to be more similar for CS+ stimuli compared to CS- stimuli during acquisition (Braem et al., 2017; Levine et al., 2021; Visser, de Haan et al., 2016; Visser et al., 2013, 2015). These results have also been extended to threat learning paradigms investigating the role of social interactions, showing increased dACC and insular similarity for intentional vs. unintentional threats (Undeger et al., 2020; but see Undeger et al., 2021). These results have also been linked to behavioral conditioned responses (Visser et al., 2013, 2015). Relatedly, in a threat generalization paradigm which includes benign stimuli that incidentally resemble the CS+ , the anterior insula shows a similar pattern of response to the threat stimuli as to the related stimuli (Onat and Büchel, 2015). Such generalized threat information persists in the dACC and insula 24 h later, suggesting that both regions support the retention and subsequent retrieval of threat associations (Hennings et al., 2022). Overall, univariate and multivariate fMRI analyses converge to provide strong evidence suggesting that the salience network serves a central role in the acquisition and retrieval of threat in humans.

### 3.3. vmPFC

Findings from rodent models suggest that the vmPFC primarily supports the learning and retrieval of extinction memories. Consistent with this view, univariate analyses of human fMRI data have found that the vmPFC is reliably deactivated (i.e., CS- > CS+) during threat learning (Fullana et al., 2016). These univariate findings suggest that vmPFC codes exclusively for safety signals, both during threat learning and extinction. However, multivariate techniques provide evidence that the vmPFC may represent threat associations as well (Hennings et al., 2022). In addition, pattern similarity amongst CS+ stimuli in the vmPFC has been observed to increase over time (Braem et al., 2017; Visser, de Haan et al., 2016; Visser et al., 2013, 2015). These results might suggest a more general function for the vmPFC in learning and updating stimulus value over time (e.g., Battaglia et al., 2020**)**.

### 3.4. Other regions

Whole-brain MVPA analyses have identified shifts or tuning of neural representations during threat acquisition outside of the core threat and salience circuits described above. Li et al. (2008) found aversive learning modulated cortical representations of previously neutral odors in piriform cortex, a specialized sensory region for smell. These learning-dependent representational shifts have been observed across multiple types of stimuli and sensory cortices, including basic visual stimuli in early visual cortex (gabor patches, Yin et al., 2020), and complex visual stimuli in higher-order visual cortices (de Voogd et al., 2016; Dunsmoor et al., 2014; Visser et al., 2011). Similarly, following threat acquisition, neural patterns of CS+ and CS- tones in primary auditory cortex can be reliably separated, both for simple and complex auditory stimuli (Reddan et al., 2018; Staib et al., 2020; Staib and Bach, 2018). In addition to these findings in sensory cortices, multivariate analyses have identified threat-related signals in all major lobes of the brain, including the cerebellum (Faul et al., 2020; Hennings et al., 2022; Levine et al., 2021; Visser et al., 2013, 2015; Visser, de Haan et al., 2016).

### 3.5. MVPA reveals circuitry of extinction learning and retrieval

Rodent models establish that in addition to their role in threat acquisition, the amygdala, hippocampus, and vmPFC also support extinction learning and retrieval (Bouton et al., 2021). MVPA has proved sensitive in detecting the involvement of these regions in extinction processes in humans. For example, representational similarity in these regions for CS+ stimuli decreases as the US is omitted (i.e., as extinction progresses Visser et al., 2013, 2015). These regions also exhibit a significant shift in CS+ representations from the end of acquisition to the end of extinction learning, which is a putative index of extinction memory formation (Graner et al., 2020; Hauner et al., 2013). Extinction learning can also be successfully detected using MVPA at a later testing session. For example, both the vmPFC and anterior hippocampus engage in neural reinstatement of the extinction memory during recall tests (Hennings et al., 2020, 2022). Additionally, multivoxel activation patterns identified in the vmPFC during extinction learning are spontaneously reactivated immediately after extinction during rest (Gerlicher et al., 2018). Moreover, the degree of extinction-related reactivations in the vmPFC at rest predicts subsequent extinction memory retrieval at a later test.

### 3.6. MVPA identifies stable biomarkers of threat learning

The multivariate analyses discussed thus far have focused on determining if a specific region of interest (ROI) contains information about either threat acquisition or extinction processes. This practice follows from the robust rodent and basic human literature mapping the substrates of fear and anxiety. MVPA can also be applied more broadly to classify neural states based on signals across several ROIs. In this approach, a classifier is trained on data from the entire brain simultaneously, which allows for inferences on the relative importance of different regions in the classification process. For example, Reddan et al., (2018) were able to construct a putative threat-predictive neural pattern that was used to test the efficacy of a novel extinction intervention. This predictive pattern was found to have strong weights in core regions of the threat network, including the vmPFC and amygdala. These predictive patterns are robust, as they can generalize to other samples and can be used to predict behavior (e.g., Zhou et al., 2021). Other applications of larger scale pattern analysis include decoding the threat value of ambiguous stimuli in a threat generalization paradigm (Visser, Haver et al., 2016), and identifying separable circuits that support physiological responding vs. subjective fear ratings (Taschereau-Dumouchel, Kawato et al., 2020). Given the difficulty of using univariate signals as predictive markers of threat, generalizable predictive patterns are appealing tools for identifying robust biomarkers of normal threat processing, and potentially of psychopathology.

### 3.7. Clinical applications of MVPA

Continued development of multivariate analyses has the potential to enhance clinical translation efforts. For example, real-time neurofeedback interventions based on univariate signals have sometimes, but not consistently, ameliorated symptoms of posttraumatic stress disorder (Chiba et al., 2019). A possible explanation for this inconsistency is that univariate signals are not reliable markers of threat processes in healthy adults. The amygdala is a common target of these interventions, and MVPA could improve their efficacy by providing more reliable neural signals as neurofeedback targets. For example, Koizumi et al. (2016) used neurofeedback to create an extinction memory without directly exposing participants to the CS+ following threat acquisition. This intervention relied on decoding the visual representation of the CS+ in the ventral visual stream, which is a popular approach in multivariate neurofeedback studies (Taschereau-Dumouchel, Cortese et al., 2020). However, future neurofeedback interventions seeking to modulate threat processes might consider more direct targets related to threat and extinction processing. For example, we have highlighted several studies that have used MVPA to identify threat-related patterns of activity in the amygdala and vmPFC. Future research should test whether directly targeting these patterns with neurofeedback is effective.

## 4. Conclusion

Despite the many strengths that multivariate approaches possess over the traditional univariate approach, it should be noted that they are not a panacea to all neuroimaging woes. Multivariate methods cannot rescue poorly designed studies, nor rescue poor signal quality in spatial regions suffering from BOLD dropout or other signal artefacts. As with any fMRI analysis, great care should be given to study design and preprocessing steps (Mumford et al., 2012, 2014; Turner et al., 2012). Multivariate methods come with their own set of unique pitfalls and interpretative challenges (for in-depth discussions, see Davis and Poldrack, 2013; Dimsdale-Zucker and Ranganath, 2018; Etzel et al., 2013). For example, the directionality of effects can be more difficult to ascertain in representational analyses, as both negative and positive activations can result in strong similarity. Multivariate methods can be relatively complex, and the added experimenter degrees of freedom, if not carefully considered, can lead to unconstrained analysis attempts (i.e., “p-hacking”). As these techniques continue to develop, we encourage would-be practitioners to critically develop a priori hypotheses, and take meaningful steps to reduce the chances of spurious and unreproducible findings (e.g., pre-registration).

Functional MRI has allowed researchers remarkable access to the neural mechanisms of Pavlovian conditioning and extinction in the human brain. However, progress of the neuroscience of fear and anxiety in humans should begin to adopt analytical approaches and computational methods that have been widely implemented in other areas of cognitive neuroscience. We have sought to highlight some of the advances in understanding that are possible when multivariate analyses are applied to fMRI data. Specifically, these analyses are sensitive to the subtle information present in distributed patterns of activity, and they have been used successfully to confirm a role of the amygdala, hippocampus, vmPFC, and dACC in threat and extinction processes. In addition, these techniques have shown that a variety of aversive learning processes are accompanied by representational shifts and separations across the cortex. As a complement, and often an enhancement, to traditional approaches, researchers would benefit from including multivariate analyses in their study of threat learning in humans.

## Data Availability

No data was used for the research described in the article.

## Acknowledgements

A.C.H is funded by the NIH (F31MH124360). S.E.C. is funded by the NIH (F32MH129136). J.A.L.-P. is funded by the NIH (R01EY028746 and R01MH129042). J.E.D. is funded by the NIH (R01 MH122387) and the NSF (CAREER Award 1844792).

## References

Bach, D.R., Weiskopf, N., Dolan, R.J., 2011. A stable sparse fear memory trace in human amygdala. J. Neurosci. 31 (25), 9383–9389. https://doi.org/10.1523/JNEUROSCI.1524-11.2011.

Battaglia, S., Garofalo, S., di Pellegrino, G., Starita, F., 2020. Revaluing the role of vmPFC in the acquisition of pavlovian threat conditioning in humans. J. Neurosci. 40 (44), 8491–8500. https://doi.org/10.1523/JNEUROSCI.0304-20.2020.

Bouton, M.E., Maren, S., McNally, G.P., 2021. Behavioral and neurobiological mechanisms of pavlovian and instrumental extinction learning. Physiol. Rev. 101 (2), 611–681. https://doi.org/10.1152/physrev.00016.2020.

Braem, S., Houwer, J.D., Demanet, J., Yuen, K.S.L., Kalisch, R., Brass, M., 2017. Pattern analyses reveal separate experience-based fear memories in the human right amygdala. J. Neurosci. 37 (34), 8116–8130. https://doi.org/10.1523/JNEUROSCI.0908-17.2017.

Chiba, T., Kanazawa, T., Koizumi, A., Ide, K., Taschereau-Dumouchel, V., Boku, S., Hishimoto, A., Shirakawa, M., Sora, I., Lau, H., Yoneda, H., Kawato, M., 2019. Current status of neurofeedback for post-traumatic stress disorder: a systematic review and the possibility of decoded neurofeedback. Front. Hum. Neurosci. 13 https://www.frontiersin.org/article/10.3389/fnhum.2019.00233.

Cohen, J.D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K.A., Pillow, J., Ramadge, P.J., Turk-Browne, N.B., Willke, T.L., 2017. Computational approaches to fMRI analysis. Nat. Neurosci. 20 (3), 304–313. https://doi.org/10.1038/nn.4499.

Davis, T., Poldrack, R.A., 2013. Measuring neural representations with fMRI: practices and pitfalls. Ann. N. Y. Acad. Sci. 1296 (1), 108–134. https://doi.org/10.1111/nyas.12156.

Dimsdale-Zucker, H.R., Ranganath, C., 2018. Representational similarity analyses. In: Handbook of Behavioral Neuroscience, Vol. 28. Elsevier, pp. 509–525. https://doi.org/10.1016/B978-0-12-812028-6.00027-6.

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

Etzel, J.A., Zacks, J.M., Braver, T.S., 2013. Searchlight analysis: promise, pitfalls, and potential. NeuroImage 78, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041.

Faul, L., Stjepanović, D., Stivers, J.M., Stewart, G.W., Graner, J.L., Morey, R.A., LaBar, K.S., 2020. Proximal threats promote enhanced acquisition and persistence of reactive fear-learning circuits. Proc. Natl. Acad. Sci. U.S.A. 117 (28), 16678–16689. https://doi.org/10.1073/pnas.2004258117.

Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., Radua, J., 2016. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. Mol. Psychiatry 21 (4), 500–508. https://doi.org/10.1038/mp.2015.88.

Fullana, M.A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., Harrison, B.J., 2018. Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants. Neurosci. Biobehav. Rev. 88, 16–25. https://doi.org/10.1016/j.neubiorev.2018.03.002.

Gerlicher, A.M.V., Tüscher, O., Kalisch, R., 2018. Dopamine-dependent prefrontal reactivations explain long-term benefit of fear extinction. Nat. Commun. 9 (1), 4294. https://doi.org/10.1038/s41467-018-06785-y.

Graner, J.L., Stjepanović, D., LaBar, K.S., 2020. Extinction learning alters the neural representation of conditioned fear. Cogn. Affect. Behav. Neurosci. 20 (5), 983–997. https://doi.org/10.3758/s13415-020-00814-4.

Haaker, J., Maren, S., Andreatta, M., Merz, C.J., Richter, J., Richter, S.H., Meir Drexler, S., Lange, M.D., Jüngling, K., Nees, F., Seidenbecher, T., Fullana, M.A., Wotjak, C.T., Lonsdorf, T.B., 2019. Making translation work: Harmonizing cross-species methodology in the behavioural neuroscience of Pavlovian fear conditioning. Neurosci. Biobehav. Rev. 107, 329–345. https://doi.org/10.1016/j.neubiorev.2019.09.020.

Hauner, K.K., Howard, J.D., Zelano, C., Gottfried, J.A., 2013. Stimulus-specific enhancement of fear extinction during slow-wave sleep. Nat. Neurosci. 16 (11), 1553–1555. https://doi.org/10.1038/nn.3527.

Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37 (1), 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325.

Hennings, A.C., McClay, M., Lewis-Peacock, J.A., Dunsmoor, J.E., 2020. Contextual reinstatement promotes extinction generalization in healthy adults but not PTSD. Neuropsychologia 147, 107573. https://doi.org/10.1016/j.neuropsychologia.2020.107573.

Hennings, A.C., McClay, M., Drew, M.R., Lewis-Peacock, J.A., Dunsmoor, J.E., 2022. Neural reinstatement reveals divided organization of fear and extinction memories in the human brain. e5 Curr. Biol. 32 (2), 304–314. https://doi.org/10.1016/j.cub.2021.11.004.

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., Lau, H., 2016. Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. Nat. Hum. Behav. 1 (1), 1–7. https://doi.org/10.1038/s41562-016-0006.

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 (November), 1–28. https://doi.org/10.3389/neuro.06.004.2008.

LaBar, K.S., LeDoux, J.E., Spencer, D.D., Phelps, E.A., 1995. Impaired fear conditioning following unilateral temporal lobectomy in humans. J. Neurosci. 15 (10), 6846–6855. https://doi.org/10.1523/JNEUROSCI.15-10-06846.1995.

LaBar, K.S., Gatenby, J.C., Gore, J.C., LeDoux, J.E., Phelps, E.A., 1998. Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. Neuron 20 (5), 937–945. https://doi.org/10.1016/S0896-6273(00)80475-4.

Lacagnina, A.F., Brockway, E.T., Crovetti, C.R., Shue, F., McCarty, M.J., Sattler, K.P., Lim, S.C., Santos, S.L., Denny, C.A., Drew, M.R., 2019. Distinct hippocampal engrams control extinction and relapse of fear memory. Nat. Neurosci. 22 (5), 753–761. https://doi.org/10.1038/s41593-019-0361-z.

LeDoux, J.E., 2000. Emotion circuits in the brain. Annu. Rev. Neurosci. 23 (1), 155–184.

LeDoux, J.E., Pine, D.S., 2016. Using neuroscience to help understand fear and anxiety: a two-system framework. Am. J. Psychiatry 173 (11), 1083–1093. https://doi.org/10.1176/appi.ajp.2016.16030353.

Levine, S.M., Kumpf, M., Rupprecht, R., Schwarzbach, J.V., 2021. Supracategorical fear information revealed by aversively conditioning multiple categories. Cogn. Neurosci. 12 (1), 28–39. https://doi.org/10.1080/17588928.2020.1839039.

Lewis-Peacock, J.A., Norman, K.A., 2014. Multivoxel pattern analysis of functional MRI data. Cogn. Neurosci. 911–919.

Li, W., Howard, J.D., Parrish, T.B., Gottfried, J.A., 2008. Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. Science 319 (5871), 1842–1845. https://doi.org/10.1126/science.1152837.

Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., Merz, C.J., 2017. Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neurosci. Biobehav. Rev. 77, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026.

Meyer, H.C., Odriozola, P., Cohodes, E.M., Mandell, J.D., Li, A., Yang, R., Hall, B.S., Haberman, J.T., Zacharek, S.J., Liston, C., Lee, F.S., Gee, D.G., 2019. Ventral hippocampus interacts with prelimbic cortex during inhibition of threat response via learned safety in both mice and humans. Proc. Natl. Acad. Sci. U.S.A. 116 (52), 26970–26979. https://doi.org/10.1073/pnas.1910481116.

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. NeuroImage 59 (3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076.

Mumford, J.A., Davis, T., Poldrack, R.A., 2014. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. NeuroImage 103, 130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005.

Onat, S., Büchel, C., 2015. The neuronal basis of fear generalization in humans. Nat. Neurosci. 18 (12), 1811–1818. https://doi.org/10.1038/nn.4166.

Pape, H.C., Pare, D., 2010. Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. Physiol. Rev. 90 (2), 419–463. https://doi.org/10.1152/physrev.00037.2009.

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45, S199–S209. https://doi.org/10.1016/j.neuroimage.2008.11.007.Machine.

Phillips, R.G., LeDoux, J.E., 1992. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. Behav. Neurosci. 106 (2), 274–285. https://doi.org/10.1037/0735-7044.106.2.274.

Qin, C., Bian, X.-L., Wu, H.-Y., Xian, J.-Y., Cai, C.-Y., Lin, Y.-H., Zhou, Y., Kou, X.-L., Chang, L., Luo, C.-X., Zhu, D.-Y., 2021. Dorsal hippocampus to infralimbic cortex circuit is essential for the recall of extinction memory. Cereb. Cortex 31 (3), 1707–1718. https://doi.org/10.1093/cercor/bhaa320.

Raichle, M.E., 2015. The brain's default mode network. Annu. Rev. Neurosci. 38, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030.

Reddan, M.C., Wager, T.D., 2018. Modeling pain using fMRI: from regions to biomarkers. Neurosci. Bull. 34 (1), 208–215. https://doi.org/10.1007/s12264-017-0150-1.

Reddan, M.C., Wager, T.D., Schiller, D., 2018. Attenuating neural threat expression with imagination. e4 Neuron 100 (4), 994–1005. https://doi.org/10.1016/j.neuron.2018.10.047.

Ritchey, M., Wing, E.A., LaBar, K.S., Cabeza, R., 2013. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cereb. Cortex 23 (12), 2818–2828. https://doi.org/10.1093/cercor/bhs258.

Seeley, W.W., 2019. The salience network: a neural system for perceiving and responding to homeostatic demands. J. Neurosci. 39 (50), 9878–9882. https://doi.org/10.1523/JNEUROSCI.1138-17.2019.

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., Konrad, C., 2009. Human fear conditioning and extinction in neuroimaging: a systematic review. PLoS One 4 (6), e5865. https://doi.org/10.1371/journal.pone.0005865.

Sjouwerman, R., Scharfenort, R., Lonsdorf, T.B., 2020. Individual differences in fear acquisition: multivariate analyses of different emotional negativity scales, physiological responding, subjective measures, and neural activation. Sci. Rep. 10 (1), 15283. https://doi.org/10.1038/s41598-020-72007-5.

Staib, M., Bach, D.R., 2018. Stimulus-invariant auditory cortex threat encoding during fear conditioning with simple and complex sounds. NeuroImage 166, 276–284. https://doi.org/10.1016/j.neuroimage.2017.11.009.

Staib, M., Abivardi, A., Bach, D.R., 2020. Primary auditory cortex representation of fear-conditioned musical sounds. Hum. Brain Mapp. 41 (4), 882–891. https://doi.org/10.1002/hbm.24846.

Taschereau-Dumouchel, V., Cortese, A., Lau, H., Kawato, M., 2020. Conducting decoded neurofeedback studies. *nsaa063* Soc. Cogn. Affect. Neurosci.. https://doi.org/10.1093/scan/nsaa063.

Taschereau-Dumouchel, V., Kawato, M., Lau, H., 2020. Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. Mol. Psychiatry 25 (10), 2342–2354. https://doi.org/10.1038/s41380-019-0520-3.

Tovote, P., Fadok, J.P., Lüthi, A., 2015. Neuronal circuits for fear and anxiety. Nat. Rev. Neurosci. 16 (6), 317–331. https://doi.org/10.1038/nrn3945.

Turner, B.O., Mumford, J.A., Poldrack, R.A., Ashby, F.G., 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. NeuroImage 62 (3), 1429–1438. https://doi.org/10.1016/j.neuroimage.2012.05.057.

Undeger, I., Visser, R.M., Olsson, A., 2020. Neural pattern similarity unveils the integration of social information and aversive learning. Cereb. Cortex 30 (10), 5410–5419. https://doi.org/10.1093/cercor/bhaa122.

Undeger, I., Visser, R.M., Becker, N., de Boer, L., Golkar, A., Olsson, A., 2021. Model-based representational similarity analysis of blood-oxygen-level-dependent fMRI captures threat learning in social interactions. R. Soc. Open Sci. 8 (11), 202116 https://doi.org/10.1098/rsos.202116.

Visser, R.M., Scholte, H.S., Kindt, M., 2011. Associative learning increases trial-by-trial similarity of BOLD-MRI patterns. J. Neurosci. 31 (33), 12021–12028. https://doi.org/10.1523/JNEUROSCI.2178-11.2011.

Visser, R.M., Scholte, H.S., Beemsterboer, T., Kindt, M., 2013. Neural pattern similarity predicts long-term fear memory. Nat. Neurosci. 16 (4), 388–390. https://doi.org/10.1038/nn.3345.

Visser, R.M., Kunze, A.E., Westhoff, B., Scholte, H.S., Kindt, M., 2015. Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. Psychoneuroendocrinology 55, 8–20. https://doi.org/10.1016/j.psyneuen.2015.01.021.

Visser, R.M., de Haan, M.I.C., Beemsterboer, T., Haver, P., Kindt, M., Scholte, H.S., 2016. Quantifying learning-dependent changes in the brain: Single-trial multivoxel pattern analysis requires slow event-related fMRI. Psychophysiology 53 (8), 1117–1127. https://doi.org/10.1111/psyp.12665.

Visser, R.M., Haver, P., Zwitser, R.J., Scholte, H.S., Kindt, M., 2016. First steps in using multi-voxel pattern analysis to disentangle neural processes underlying generalization of spider fear. Front. Hum. Neurosci. 10 https://www.frontiersin.org/article/10.3389/fnhum.2016.00222.

Visser, R.M., Bathelt, J., Scholte, H.S., Kindt, M., 2021. Robust BOLD responses to faces but not to conditioned threat: challenging the amygdala's reputation in human fear and extinction learning. J. Neurosci. 41 (50), 10278–10292. https://doi.org/10.1523/JNEUROSCI.0857-21.2021.

de Voogd, L.D., Fernández, G., Hermans, E.J., 2016. Awake reactivation of emotional memory traces through hippocampal–neocortical interactions. NeuroImage 134, 563–572. https://doi.org/10.1016/j.neuroimage.2016.04.026.

Wen, Z., Raio, C.M., Pace-Schott, E.F., Lazar, S.W., LeDoux, J.E., Phelps, E.A., Milad, M. R., 2022. Temporally and anatomically specific contributions of the human amygdala to threat and safety learning. e2204066119 Proc. Natl. Acad. Sci. U.S.A. 119 (26). https://doi.org/10.1073/pnas.2204066119.

Yin, S., Bo, K., Liu, Y., Thigpen, N., Keil, A., Ding, M., 2020. Fear conditioning prompts sparser representations of conditioned threat in primary visual cortex. Soc. Cogn. Affect. Neurosci. 15 (9), 950–964. https://doi.org/10.1093/scan/nsaa122.

Zhou, F., Zhao, W., Qi, Z., Geng, Y., Yao, S., Kendrick, K.M., Wager, T.D., Becker, B., 2021. A distributed fMRI-based signature for the subjective experience of fear. Nat. Commun. 12 (1), 6643. https://doi.org/10.1038/s41467-021-26977-3.