Contents lists available at ScienceDirect

NeuroImage

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

Cross-paradigm integration shows a common neural basis for aversive and appetitive conditioning

Sanja Klein[a,b,d,*], Onno Kruse[a,b], Isabell Tapia León[b,c], Lukas Van Oudenhove[e,f,g], Sophie R. van 't Hof[h], Tim Klucken[c], Tor D. Wager[g], Rudolf Stark[a,b,d]

[a] Department of Psychotherapy and Systems Neuroscience, Justus Liebig University, Giessen 35394, Germany
[b] Bender Institute for Neuroimaging (BION), Justus Liebig University, Giessen 35394, Germany
[c] Clinical Psychology and Psychotherapy, University Siegen, Siegen 57076, Germany
[d] Center of Mind, Brain and Behavior, Universities of Marburg and Giessen, Marburg 35032, Germany
[e] Department of Chronic Diseases and Metabolism (CHROMETA), Laboratory for Brain-Gut Axis Studies (LaBGAS), Translational Research Centre for Gastrointestinal Disorders TARGID, KU Leuven, Leuven, Belgium
[f] Leuven Brain Institute, KU Leuven, Leuven, Belgium
[g] Department of Psychological and Brain Sciences, Cognitive and Affective Neuroscience Lab, Dartmouth College, Hanover, NH, USA
[h] Department of Psychiatry, Amsterdam University Medical Centers, Amsterdam 1105 AZ, The Netherlands

ARTICLE INFO

Keywords:
fMRI
Conditioning
Learning
Fear
Reward
Striatum

ABSTRACT

Sharing imaging data and comparing them across different psychological tasks is becoming increasingly possible as the open science movement advances. Such cross-paradigm integration has the potential to identify commonalities in findings that neighboring areas of study thought to be paradigm-specific. However, even the integration of research from closely related paradigms, such as aversive and appetitive classical conditioning is rare – even though qualitative comparisons already hint at how similar the 'fear network' and 'reward network' may be. We aimed to validate these theories by taking a multivariate approach to assess commonalities across paradigms empirically. Specifically, we quantified the similarity of an aversive conditioning pattern derived from meta-analysis to appetitive conditioning fMRI data. We tested pattern expression in three independent appetitive conditioning studies with 29, 76 and 38 participants each. During fMRI scanning, participants in each cohorts performed an appetitive conditioning task in which a CS+ was repeatedly rewarded with money and a CS- was never rewarded. The aversive pattern was highly similar to appetitive CS+ > CS- contrast maps across samples and variations of the appetitive conditioning paradigms. Moreover, the pattern distinguished the CS+ from the CS- with above-chance accuracy in every sample. These findings provide robust empirical evidence for an underlying neural system common to appetitive and aversive learning. We believe that this approach provides a way to empirically integrate the steadily growing body of fMRI findings across paradigms.

1. Introduction

Comparing paradigms and results across research areas is necessary to advance knowledge in basic and translational neuroscience. But even very closely related areas of research are often studied in parallel, accumulating data with little cross-fertilization between areas and their respective paradigms. Two such areas are the neural basis of fear learning and reward learning - conceptualized as aversive and appetitive conditioning, respectively. When these intrinsically adaptive learning processes become excessive, they can become the basis for psychological disorders such as anxiety, depression and addiction (Duits et al., 2015; Martin-Soelch et al., 2007). This conceptual distinction is reflected in the Research Domain Criteria (RDoC) framework, with 'fear learning' and 'reward learning' belonging to the separate domains of negative and positive valence systems (Insel, 2014). However, possible common underlying or interacting factors in these disorders (Destoop et al., 2019; Liverant et al., 2014; Xie et al., 2021) can be easily overlooked when we only examine these domains separately. Thus, shedding light on commonalities regarding their basic neural processes is essential going forward. Some efforts have been made to translate neuroimaging evidence from aversive to appetitive conditioning paradigms, but limited to qualitative comparisons and narrative reviews (e.g. Brooks and Berns 2013, Moscarello and LeDoux 2013, Stefanova et al. 2020). Only very recently, a meta-analysis on prediction errors also included a look at appetitive and aversive stimuli at a global level (Corlett et al., 2022). So far, no empirical integration of neuroimaging data from specifically aversive and appetitive conditioning studies has been attempted, although a lot of imaging data – especially on aversive conditioning – already exists and

[*] Corresponding author at: University of Giessen, Otto-Behaghel-Str. 10 H, Giessen 35394, Germany.
E-mail address: sanja.klein.psychol@gmail.com (S. Klein).

https://doi.org/10.1016/j.neuroimage.2022.119594.
Received 3 March 2022; Received in revised form 22 July 2022; Accepted 25 August 2022
Available online 28 August 2022.

Table 1
A detailed overview of regions reported in meta-analyses as well as theoretical models of aversive and appetitive learning. Common regions between aversive and appetitive CS+ are amygdala, NAcc, caudate nucleus, putamen, insula and thalamus.

	Aversive conditioning	Appetitive conditioning
Theoretical models	amygdala, mPFC, hippocampus Tovote et al. (2015) amygdala, mPFC, dmPFC Herry and Johansen (2014)	amygdala, OFC, dACC, vACC, NAcc, caudate nucleus, putamen Martin-Soelch et al. (2007) amygdala, NAcc Averbeck and Costa (2017)
Empirical meta-analytic evidence	dACC, thalamus, anterior insular cortex, amygdala, OFC, putamen, midbrain/substantia nigra Etkin and Wager (2007) amygdala (smaller effects, only in uninstructed studies), anterior insula, putamen, caudate nucleus, dmPFC, dACC, preSMA, thalamus, pallidum Mechias et al. (2010) anterior insular cortex, NAcc, caudate, SMA/preSMA, dlPFC, precuneus, cerebellum Fullana et al. (2016)	amygdala, NAcc, caudate nucleus, putamen, midbrain, thalamus, frontal operculum, insula Chase et al. (2015)

Abbreviations: Nucleus Accumbens (NAcc), prefrontal cortex (PFC), medial PFC (mPFC), dorsomedial PFC (dmPFC), dorsolateral PFC (dlPFC), supplementar motor area (SMA), orbitofrontal cortex (OFC), dorsal/ventral anterior cingulate cortex (dACC/vACC)

qualitatively, activation patterns seem similar. Therefore, our general aims were, first, to attempt the empirical integration of data across the paradigms of aversive and appetitive conditioning. Second, we wanted to demonstrate the feasibility of integrating findings from these two paradigms in order to enable further research across a multitude of other paradigms of varying similarity.

The differential aversive or appetitive conditioning paradigms that are employed in fMRI research in humans are highly alike. An initially neutral stimulus becomes a conditioned stimulus (CS+) after repeated pairing with an aversive or appetitive unconditioned stimulus (UCS, e.g. electric shock or money). A second stimulus (CS-) is never paired with a UCS (Mackintosh, 1975). On the one hand these are striking similarities, on the other hand reward and fear seem diametrically opposed leading to separate investigations into brain regions constituting a fear network or a reward network. The neural correlates of aversive conditioning have been researched extensively in human neuroimaging, which has led to a large body of fMRI results on the topic as well as meta-analyses (for reviews see Etkin and Wager 2007, Fullana et al. 2016, Mechias et al. 2010, Sehlmeyer et al. 2009). In parallel, fMRI studies on appetitive conditioning have begun to accumulate (for reviews see Averbeck and Costa 2017, Chase et al. 2015, Martin-Soelch et al. 2007). It has become increasingly apparent that the findings from aversive and appetitive conditioning are qualitatively similar. The same regions often emerge from separate meta-analyses of responses to a CS+ compared to a CS- in aversive (Etkin and Wager, 2007; Fullana et al., 2016; Mechias et al., 2010) and appetitive (Chase et al., 2015) conditioning, see Table 1 for details. Seminal theoretical models of aversive conditioning focus mainly on the amygdala (Herry and Johansen, 2014; Tovote et al., 2015) while appetitive conditioning models also include striatal regions such as the Nucleus Accumbens (NAcc; Averbeck and Costa 2017, Martin-Soelch et al. 2007). In summary, the amygdala, NAcc, caudate nucleus, putamen, insula and thalamus seem to be involved in both aversive and appetitive learning, based on qualitative comparison of empirical data as well as theoretical models (see Table 1). The cerebellum has been reported in the most recent meta-analysis of aversive learning (Fullana et al., 2016) and since then in another aversive conditioning study in humans (Ernst et al., 2019). This region may be crucial for many different types of outcome prediction (Popa and Ebner, 2018) and has been shown associated with appetitive prediction in animal data (Heffley and Hull, 2019), so cerebellar activity might be another possible commonality between human aversive and appetitive learning. Based on these apparently overlapping regions, it is assumed that the concepts 'fear network' and 'reward network' share mesolimbic dopamine pathways and thus may share a common basis in an anticipatory motivational system related to learning in general (Menon and Uddin, 2010; Moscarello and LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). However, these assumptions are mostly based on qualitative literature reviews. Only few neuroimaging studies have systematically compared aversive and appetitive learning in the same experiment and even then mostly focused on differences instead of similarities (e.g. Breiter et al. 2001, Carter et al. 2009, Lake et al. 2019, Sankar et al. 2019). While elucidating the differences between these mechanisms remains important, quantifying cross-paradigm similarities might provide an even greater opportunity.

In this paper, we adopt a multivariate analysis approach to quantitatively integrate previously published evidence across paradigms and samples in order to better understand the commonalities of aversive and appetitive processes. With the help of machine learning classification algorithms, we can test whether whole-brain patterns of activation are present in a dataset and whether they distinguish between conditions (Weaverdyck et al., 2020; Woo et al., 2017). Multivariate approaches have already been used to great success in finding and validating whole brain response patterns associated with cognitive and affective states, e.g. the experience of pain (Wager et al., 2013), emotions (Kragel and LaBar, 2014; Saarimäki et al., 2016) or perceiving sexual pictures (van 't Hof et al., 2021; for a review on neural signatures see Kragel et al. (2018) based on data from the same kind of paradigm. Here, instead of developing an activation model from similar paradigms, we apply an already existing meta-analytic response pattern from one paradigm (aversive conditioning) to data from a similar paradigm (appetitive conditioning) to empirically identify activation commonalities. Using a meta-analytical pattern instead of training a new aversive conditioning pattern enables us to investigate similarity of our current appetitive conditioning data with the summarized data of numerous past aversive conditioning studies, gathered over many years of research.

In this study, we aim to identify commonalities of a differential activation pattern related to aversive conditioning, based on the meta-analysis by Fullana et al. (2016), with activation patterns in appetitive conditioning paradigms. In order to assess generalizability of the similarities, we carry out the same tests in three independent appetitive conditioning datasets with varying features regarding sample characteristics and procedural details (Kruse et al., 2018, 2020; Tapia León et al., 2019). First, we expect that the brain activation difference between aversive CS+ (avCS+) and aversive CS- (avCS-) will be similar to the activation difference between appetitive CS+ (appCS+) and appetitive CS- (appCS-), measured by a pattern expression score. We expect this for differential activation over the whole brain as well as for a priori anatomical regions of interest (ROIs: NAcc, caudate nucleus, putamen, amygdala, thalamus, insula, cerebellum), which have been implicated in both forms of learning empirically and theoretically but may have traditionally been associated with one paradigm more than the other. Second, we hypothesize that the separate appCS+ and appCS- activation data will differ in their similarity to the avCS+ > avCS- pattern. We expect to accurately discriminate whether a pattern expression score stems from whole brain appCS+ or appCS- data based on the score's size via forced-choice classification. With these analyses, we aim to provide empirical evidence for the neural commonalities of aversive and appetitive conditioning at whole brain and region level.

2. Materials and methods

2.1. Sample descriptions

We used three previously published datasets on appetitive conditioning. All studies were approved by the local ethics committee and were conducted in accordance with the 1964 declaration of Helsinki and its later amendments. Participants gave written informed consent and received 10 € per hour or course credit for their participation plus monetary gains from the tasks.

2.1.1. Active learning/homogeneous sample

The Active Learning/Homogeneous Sample included only male subjects and a between-person acute stress condition (Kruse et al., 2018, see also Kruse et al. (2017). For our analysis, we included only the no-stress control group (n = 29, control group from Kruse et al. (2018) for this analysis. The mean age was M = 23.83 (SD = 2.80). Because this sample was the control group in a strictly timed stress experiment, the overall procedure was more rigorously controlled and standardized than in the other two samples.

2.1.2. Active learning/heterogeneous sample

The Active Learning/Heterogeneous Sample was larger (n = 76, Kruse et al. 2020) and included 36 men and 40 women with a mean age of M = 23.76 (SD = 3.73).

2.1.3. Passive learning/heterogeneous sample

The Passive Learning/Heterogeneous Sample (n = 38, Tapia León et al. 2019) also included men as well as women (22 men, 16 women) with a mean age of M = 23.50 (SD = 3.54).

2.2. Conditioning paradigms

2.2.1. Active learning/homogeneous sample and active learning/heterogeneous sample

The same uninstructed differential conditioning paradigm was used in both the Active Learning/Homogeneous Sample (Kruse et al., 2018) and Active Learning/Heterogeneous Sample (Kruse et al., 2020). In each trial, the subject was presented with a CS+ or CS- (blue or yellow rectangle) and then with a target (white square), upon which they were instructed to press a button as quickly as possible. Reactions within target presentation time were rewarded with 50 cents (UCS) only if a CS+ was presented before the target (timing of the target was predetermined, so that approx. 62% of all CS+ trials were rewarded). Fast reactions after a CS- were never rewarded. Participants were instructed to pay attention to possible contingencies before the task and received the money they won after scanning. The paradigm included 21 CS+ and 21 CS- trials. The first two trials (always one CS+ and one CS-) were excluded from further analyses, since learning could not have taken place yet, leaving 20 CS+ and CS- trials each per subject. For more detailed information about the paradigm please see the original publications for the Active Learning/Homogeneous Sample (Kruse et al., 2018) and the Active Learning/Heterogeneous Sample (Kruse et al., 2020). See also Fig. 1 for graphical representation of the task.

2.2.2. Passive learning/heterogeneous sample

In the Passive Learning/Heterogeneous Sample (Tapia León et al., 2019), an instructed differential conditioning paradigm without any behavioral reaction component was used. Participants were presented with a CS+ or CS- (blue or yellow rectangles) followed by feedback about reward/no reward. Half of the CS+ trials were rewarded with 50 cents (UCS) while the CS- was never rewarded. Participants were instructed about the relationships between CS and UCS before the task and received the money from the experiment after leaving the scanner. The paradigm included 20 CS+ trials and 20 CS- trials. For more detailed information about the paradigm see the original publication for the Passive Learning/Heterogeneous Sample (Tapia León et al., 2019). See also the right half of Fig. 1 for graphical representation of the task.

2.3. Appetitive sample data

MRI images for all samples were acquired using the same 3 T whole-body tomograph (Siemens Prisma). Preprocessing and first level analyses were performed using Matlab and Statistical Parametric Mapping (SPM 12) implemented in Matlab R2012a (The MathWorks Inc.). Event-related general linear models in each sample included appCS+ and appCS- in addition to other task and nuisance regressors. All following analyses use appCS+, appCS- as well as appCS+ > appCS- first level contrast images from these models. For detailed information on data acquisition, preprocessing and first level analysis, please see the supplementary information (S1 and S2) or the original sample publications (Active Learning/Homogeneous Sample: Kruse et al. 2018; Active Learning/Heterogeneous Sample: Kruse et al. 2020; Passive Learning/Heterogeneous Sample: Tapia León et al. 2019).

For this study, we additionally created group level contrast maps using paired t-tests on CSF-scaled and winsorized appCS+ and appCS- maps with custom code available from the authors' website (https://canlab.github.io; CANlab, code used for this publication available from https://github.com/s-kline/aversive-appetitive-conditioning). These were only used for visualization purposes (see Fig. 2) and not part of any subsequent analysis.

Finally, to judge how well activation data can be distinguished between appCS+ and appCS- condition without the aversive pattern, we performed multivariate predictive modeling analyses on the appetitive data only using custom code (https://canlab.github.io; CANlab, 2020, https://github.com/s-kline/aversive-appetitive-conditioning). In each conditioning sample, a classifier was trained and tested to distinguish between appCS+ and appCS- using whole-brain Support Vector Machines (Burges, 1998; Gramfort et al., 2013). We used 5-fold cross-validation blocked by subject (i.e., leaving out all images from a particular participant together), which allows every subject to serve as both training and test data at one point. The classifiers were trained on whole-brain appCS+ > appCS- first level contrast images masked with a gray matter mask. Each SVM model resulted in a pattern of weights of each voxel predicting the appCS+ or appCS- stimulus presentation (appCS+ > appCS- predictive weight map) and an intercept (offset) value. Bootstrap resampling (with 5,000 bootstrap samples; see also Wager et al. 2013) was used to estimate voxel-wise p-values for each predictive weight map. We tested for significant clusters in the predictive weight maps thresholded at P = .05, FDR (false discovery rate)-corrected.

2.4. Aversive conditioning pattern

For the aversive conditioning pattern, we used a whole brain pattern which discriminates within aversive conditioning paradigms between CS+ (avCS+) and CS- (avCS-). This avCS+ > avCS- pattern was the result of a meta-analysis of 27 independent fear conditioning data sets (total subjects N = 677, 54% male; Fullana et al. 2016). Specifically, Fullana et al. computed functional activation differences between avCS+ and avCS- for each study, either from original contrast maps or

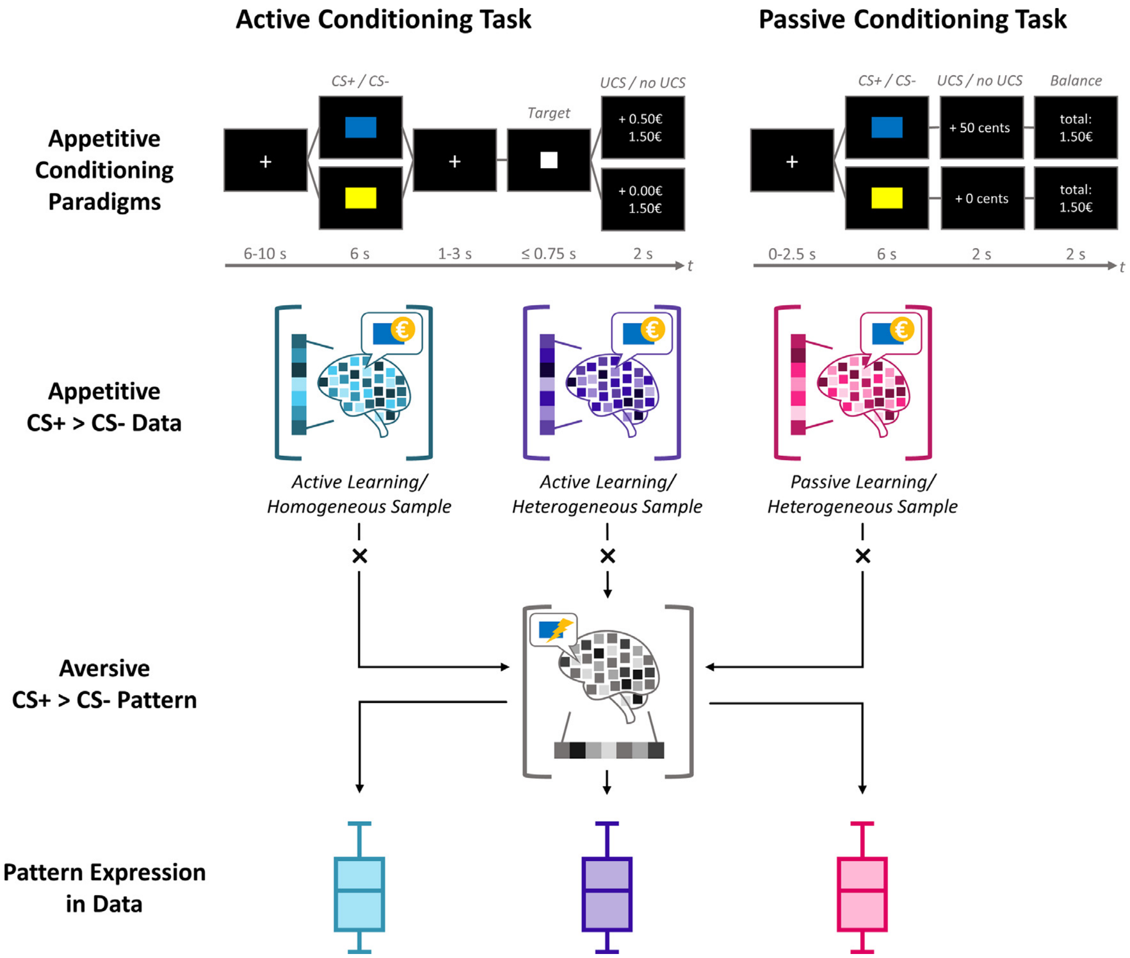
Active Conditioning Task
Passive Conditioning Task
Appetitive Conditioning Paradigms
CS+ / CS-
Target
UCS / no UCS
+ 0.50€ 1.50€
+ 0.00€ 1.50€
+
6-10 s
6 s
1-3 s
≤ 0.75 s
2 s
t
CS+ / CS-
UCS / no UCS
Balance
+ 50 cents
+ 0 cents
total: 1.50€
total: 1.50€
0-2.5 s
6 s
2 s
2 s
t
Appetitive CS+ > CS- Data
Active Learning/ Homogeneous Sample
Active Learning/ Heterogeneous Sample
Passive Learning/ Heterogeneous Sample
Aversive CS+ > CS- Pattern
Pattern Expression in Data

Fig. 1. Methods Summary

Note: Both active samples underwent the same active appetitive conditioning task. In each trial, subjects could win 50 cents with a fast reaction to the target only if a CS+ was shown before. The Passive Sample underwent a passive appetitive conditioning task. Subjects were shown CS+ and CS- and subsequent wins of 50 cents or nothing. Activation maps related to appetitive CS+ and CS- presentation averaged over the whole task were computed for each subject in each sample. The aversive conditioning pattern was applied to these subject-specific maps using cosine similarity metric. The pattern expression values reflect the magnitude of similarity between two normalized image vectors.

the peak coordinates reported in the studies. They then created a brain map of the effect size of the difference between the two conditions for each study using AES-SDM software (www.sdmproject.com/) and with these maps conducted a voxel-wise random-effects meta-analysis with weighting for sample size and variance. Fullana et al. (2016) found several large bilateral clusters demonstrating consistently significant functional activations during aversive conditioning (avCS+ > avCS-) including anterior insular cortex, NAcc, caudate nucleus, dACC and lateral cerebellum. Most of the included studies used electric shocks as UCS and simple geometric shapes as CS. The whole brain map of z-values associated with the difference between avCS+ and avCS- is available on Neurovault (https://identifiers.org/neurovault.collection:2472). We obtained this map of z-values from Neurovault and used it as the pattern associated with avCS+ > avCS- in our similarity analysis.

2.5. Similarity analysis

We followed the same analysis steps in each sample, using custom code available from the authors' website (https://canlab.github.io; CANlab; code used for this publication available from https://github.com/s-kline/aversive-appetitive-conditioning): (i) First, we computed pattern expression scores in the whole brain appCS+ > appCS- contrast images. (ii) Second, we computed the pattern expression in each of the ROIs NAcc, caudate nucleus, putamen amygdala, thalamus, insula and cerebellum. (iii) Finally, we computed pattern expression scores in the separate appCS+ and appCS- activation maps, which were then used in a classification analysis to test if we can distinguish appCS+ from appCS- condition based on these scores. The significance threshold for all tests was $P < .05$.

(i) To apply the pattern to our data, we initially resampled the pattern map to the space of the functional data using trilinear interpolation. Then, we used cosine similarity metric to assess the degree of similarity between the avCS+ > avCS- pattern and the individual unthresholded appCS+ > appCS- contrast image of each subject: For every subject of each of the three appetitive conditioning samples, we calculated a pattern expression score, which measures the similarity of the contrast image to the aversive conditioning pattern. As pattern expression score, we used the cosine similarity metric, which indicates to what extent the pattern image vector and the data image vector from one participant point in the same direction (Bisandu et al., 2019; Bobadilla-Suarez et al., 2020; van Oudenhove et al., 2020). For each appetitive sample participant, we calculated the dot product between the avCS+ > avCS- pattern image and their appCS+ > appCS- contrast image and divided it by the product of the two image vectors length, normalizing the result. Cosine similarity can range from 1 (indicating exact similarity, i.e. exactly the same direction of the vectors) over 0 (indicating no relation, orthogonal

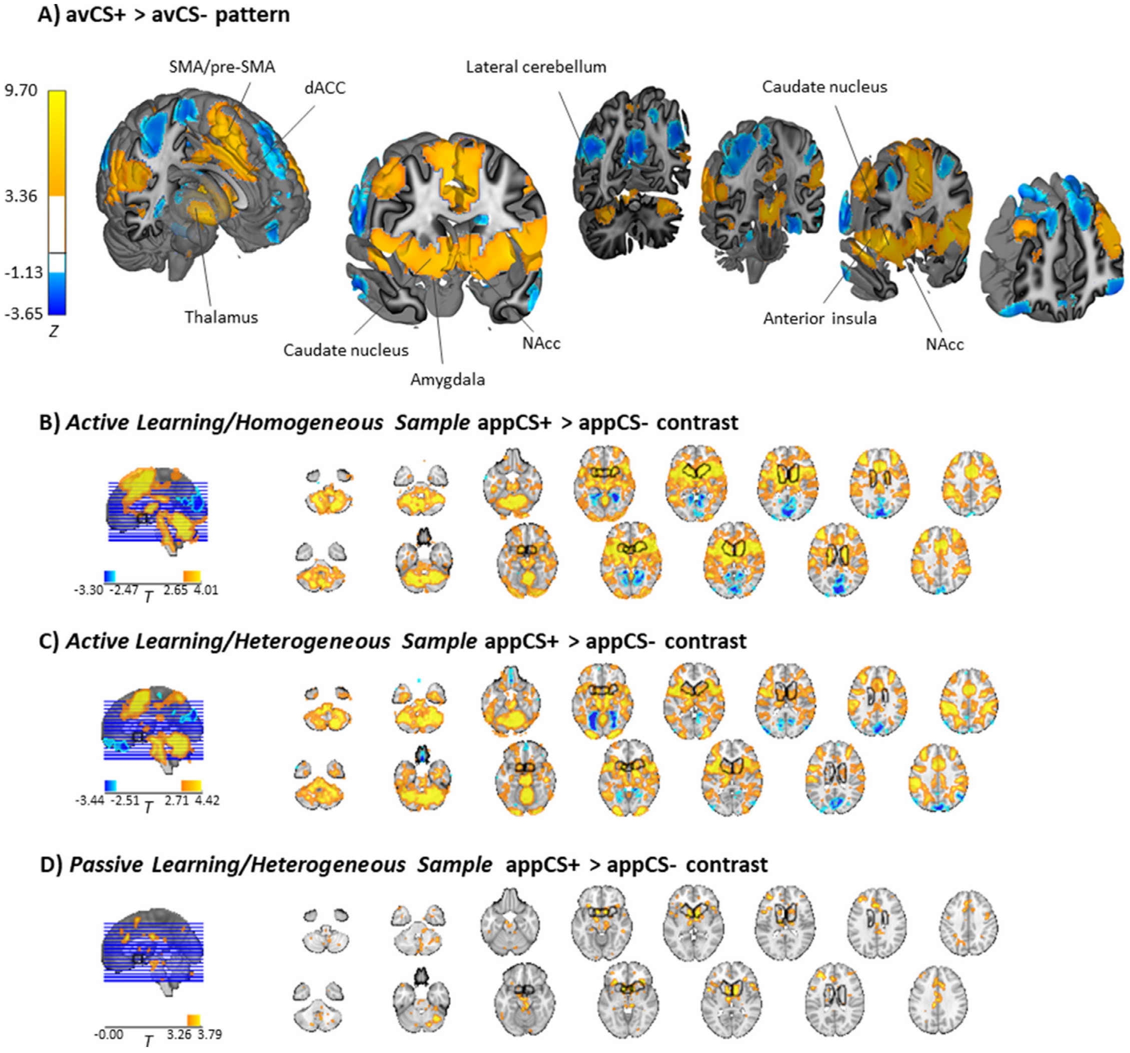

Fig. 2. Brain maps of aversive conditioning pattern and appetitive conditioning data
Note: Pattern related to aversive conditioning from meta-analysis (A). Weight map consisting of z-values is displayed on 4 coronal slices and two central cutaways showing the basal ganglia with region labels (SMA: supplementary motor area). The pattern is thresholded at $P < .005$, cluster size > 10, see Fullana et al. (2016) for details. Main effects of appetitive CS+ versus appetitive CS- in Samples (B, C, D). Contrast maps are the result of a paired t-test between CSF-scaled and winsorized activation maps of appCS+ and appCS- conditions, thresholded at $P < .05$ FDR-corrected. Midline sagittal and two rows of axial slices are shown for each sample, black outlines indicating NAcc and caudate nucleus. Anatomical images were adapted from the 7T high-resolution atlas of Keuken et al. (2014).

vectors) to -1 (indicating complete inversion, exactly opposite vector direction). Thus, in our analysis, positive cosine similarity (between 0 and 1) results when positive contrast values (appCS+ > appCS-) are found in voxels that are also positive in the aversive conditioning pattern. In accordance with that, positive cosine similarity also results when negative contrast values (appCS+ < appCS-) are found in voxels that are also negative in the aversive conditioning pattern. Equivalently, negative cosine similarity (between 0 and -1) results when positive contrast values are found in voxels that are negative in the aversive conditioning pattern and vice versa. Using this approach resulted in one pattern expression score per participant, which indicated the similarity between individual appetitive conditioning contrast images and the aversive conditioning pattern. Finally, we tested whether the appetitive conditioning contrast images were significantly similar to the aversive conditioning pattern using standard binomial tests with t-statistics, i.e. if cosine similarity was significantly different from 0.

(ii) For the ROI analysis, we masked the appCS+ > appCS- contrast images with anatomical masks for the NAcc (from the SPM anatomy toolbox), caudate nucleus, putamen (both from striatum parcellation by Pauli et al. 2016), amygdala (from the SPM anatomy toolbox), thalamus, insula (both from Harvard Oxford Atlas) and cerebellum (from Diedrichsen et al. 2009). This resulted in seven new images that only contained data in the voxels encompassed by the respective ROI. We then calculated the pattern expression scores in these images, which restricts the analysis to only the voxels within the ROI for both contrast image and pattern. Otherwise, we employed the same steps, cosine similarity metric and significance test as for the whole brain analysis described under (i).

(iii) We also computed cosine similarity of the avCS+ > avCS- pattern to the separate appCS+ and appCS- activation maps of each subject to use for classification analysis. This resulted in two pattern expression scores per participant, one indicating similarity of the pattern with appCS+ and the other one indicating similarity of the pattern with appCS-. To assess how the pattern expression scores for appCS+ and appCS- images differed from each other, we tested whether they could accurately predict the condition label appCS+ or appCS-. For this purpose, we computed forced-choice classification performance where the image with the higher avCS+ > avCS- pattern expression scores is labeled as appCS+ and the image with the smaller pattern expression score is labeled as appCS- using receiver operating characteristics (ROC; for

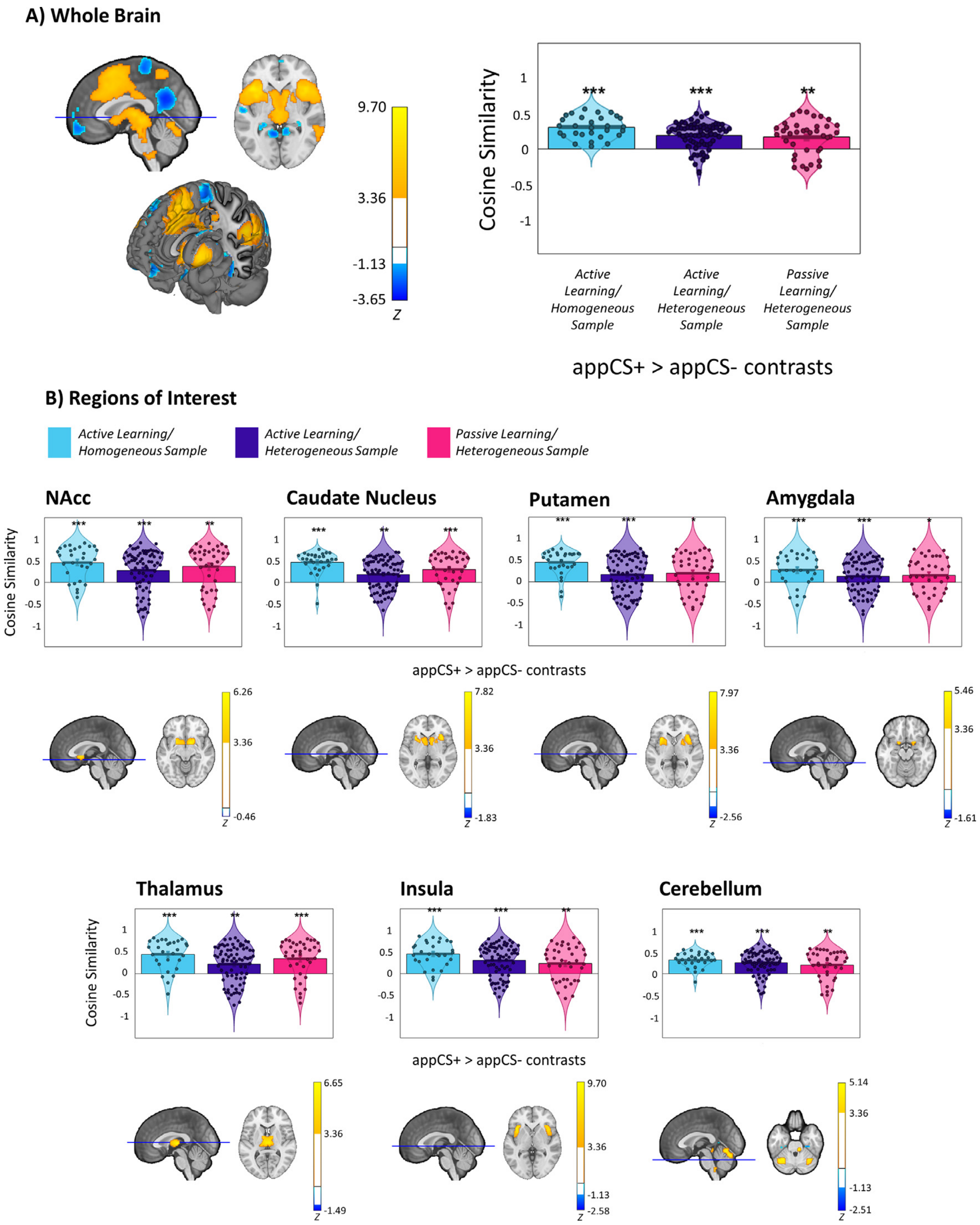

Fig. 3. Similarity between Aversive Conditioning Pattern and Appetitive Conditioning Data
Note: Results of similarity analysis for appetitive conditioning data in A) whole brain, and B) Regions of Interest NAcc, caudate nucleus, putamen, amygdala, thalamus, insula and cerebellum. For each region, the aversive conditioning pattern is shown mapped onto canonical anatomical sections (axial slice indicated by the line on mid-sagittal slice) and for the whole brain also onto respective brain cutaways adapted from the 7T high-resolution atlas of Keuken et al. (2014). The pattern is thresholded at $P < .005$, cluster size > 10, see Fullana et al. (2016) for details. Bar plots show the cosine similarity between aversive conditioning pattern and appetitive conditioning contrasts with each subject as a dot, violin plots illustrating the data distribution and error bars indicating standard error of means. *** indicates $P < .001$, ** indicates $P < .01$.

an introduction see Tharwat, 2021). We report accuracy measures and statistics of this classification based on the pattern expression values.

2.6. Control analyses

To support our assumption that similarity between aversive and appetitive conditioning is not solely driven by a common level of cognitive demand or emotional arousal features of both tasks, we performed control analyses. Specifically, pattern expression of other published whole brain multivariate patterns related to these concepts in the appetitive conditioning data were assessed. These were multivariate signatures related to cognitive control (Kragel et al., 2018), cognitive demand in a stroop task (Silvestrini et al., 2020), negative affect induced by pictures (Chang et al., 2015), as well as fearfulness and surprise induced

Table 2
Mean cosine similarity of avCS+ > avCS- pattern to appCS+ > appCS- contrast for whole brain and ROIs with standard error, statistics and effect size.

Region	Dataset	Cosine similarity	SE	T	p	Cohens d
Whole	Active Learning/Homogen.	0.304	0.028	10.85	<.001	2.02
Brain	Active Learning/Heterogen.	0.184	0.021	8.76	<.001	1.01
	Passive Learning/Heterogen.	0.160	0.039	4.13	<.001	0.67
NAcc	Active Learning/Homogen.	0.452	0.071	6.40	<.001	1.19
	Active Learning/Heterogen.	0.274	0.054	5.08	<.001	0.58
	Passive Learning/Heterogen.	0.360	0.077	4.68	<.001	0.76
Caudate	Active Learning/Homogen.	0.459	0.048	9.53	<.001	1.77
Nucleus	Active Learning/Heterogen.	0.179	0.041	4.39	<.001	0.50
	Passive Learning/Heterogen.	0.294	0.060	4.90	<.001	0.79
Putamen	Active Learning/Homogen.	0.444	0.052	8.52	<.001	1.58
	Active Learning/Heterogen.	0.159	0.049	3.24	.002	0.37
	Passive Learning/Heterogen.	0.191	0.072	2.66	.012	0.43
Amygdala	Active Learning/Homogen.	0.278	0.068	4.08	<.001	0.76
	Active Learning/Heterogen.	0.131	0.045	2.93	.005	0.34
	Passive Learning/Heterogen.	0.158	0.065	2.43	.020	0.40
Thalamus	Active Learning/Homogen.	0.443	0.064	6.93	<.001	1.29
	Active Learning/Heterogen.	0.212	0.048	4.42	<.001	0.51
	Passive Learning/Heterogen.	0.339	0.069	4.94	<.001	0.80
Insula	Active Learning/Homogen.	0.450	0.053	8.56	<.001	1.59
	Active Learning/Heterogen.	0.297	0.042	7.03	<.001	0.81
	Passive Learning/Heterogen.	0.229	0.064	3.55	.001	0.58
Cerebellum	Active Learning/Homogen.	0.309	0.031	9.82	<.001	1.82
	Active Learning/Heterogen.	0.247	0.030	8.28	<.001	0.95
	Passive Learning/Heterogen.	0.197	0.053	3.74	<.001	0.61

by music and films (Kragel and LaBar, 2015) available from the authors' website (https://canlab.github.io; CANlab). We computed expression of these patterns in each sample and tested for significance same as for the aversive pattern (see Section 2.5). If the similarity between aversive conditioning pattern and appetitive conditioning data is at least somewhat specific to conditioning, the similarity to these control patterns should be smaller in comparison. To test this, we performed paired t-tests to compare control pattern similarity and aversive conditioning pattern similarity with the appetitive conditioning data.

3. Results

3.1. Aversive pattern expression in appetitive contrast data

In line with our expectations, the aversive pattern was expressed significantly in the contrast images of every sample (all p < .001). Pattern expression was largest in the Active Learning/Homogeneous Sample with a mean cosine similarity of 0.304 (SE = 0.028, t = 10.85) and a very large effect size (Cohens d = 2.02). In the Active Learning/Heterogeneous Sample, pattern expression was moderate (cosine similarity = 0.184, SE = 0.021, t = 8.76, d = 1.01), but statistics and effect size of the similarity were still high; higher than in Passive Learning/Heterogeneous Sample (cosine similarity = 0.160, SE = 0.039, t = 4.13, d = 0.67).

As expected, pattern expression scores were also significantly large in all a priori ROIs. We found the highest scores in the striatal regions, thalamus and insula, moderately high scores in the cerebellum and moderate scores in the amygdala (for detailed statistics, see Table 2). Cosine Similarities between avCS+ > avCS+ pattern and the appCS+ > appCS- contrasts in the independent datasets are presented in Fig. 3 for whole brain and ROI data. For visual comparison, the aversive pattern as well as group contrast maps are shown in Fig. 2.

3.2. Classification of appCS+ versus appCS- by pattern expression

We computed pattern expression scores for the avCS+ > avCS- pattern in the separate appCS+ and appCS- conditions (see Fig. 4A and supplemental Table 1) to use for classification analysis. Classification results indicated that the aversive conditioning pattern could distinguish appCS+ from appCS- images accurately in every sample (classification performance in all three samples is presented in Fig. 4B). Forced choice classification effect size was largest in the Active Learning/Homogeneous Sample (100% accuracy, d = 2.08). The effect was also large in the Active Learning/Heterogeneous Sample (84% accuracy, d = 1.05) and moderate in the Passive Learning/Heterogeneous Sample (74% accuracy, d = 0.80). Importantly, the classification accuracies of the pattern for appCS+ versus appCS- were significantly above chance in all samples (all p < .05, see Table 3). These results are in line with the previous appCS+ > appCS- pattern expression results. Mean pattern expression scores were high in the appCS+ condition, supporting the notion that appCS+ activation data and avCS+ > avCS- pattern are highly similar. The pattern was also significantly expressed in the appCS- condition in every sample, likely due to basic similarities of CS+ and CS- conditions in both aversive and appetitive conditioning.

3.3. Control pattern expression in appetitive contrast data

As expected, all control patterns showed lower pattern expression in the appetitive sample data than the aversive conditioning pattern with all mean cosine similarity values < 0.08 (see Table 4 for detailed results). Only the pattern related to cognitive demand in a stroop task (Silvestrini et al., 2020) was significantly expressed in the appetitive sample data. This is probably because the stroop task has basic visual features and reaction demands in common with the appetitive conditioning paradigms. The pattern related to fearfulness (Kragel and LaBar, 2015) was significantly negatively expressed in the Active Learning/Homogenous Sample. In line with our expectations, the aversive conditioning pattern was expressed more strongly in appetitive conditioning data than any pattern related to cognitive demands and emotion processing (all p<.05 in paired t-tests, see Table 4).

3.4. Appetitive conditioning SVM classification

For all conditioning samples, we obtained predictive weight maps through support vector machine (SVM) classification (shown in supplemental Fig. 1A–C). The classifier trained on the Active Learning/Homogeneous Sample performed with 100% accuracy and a large effect size (d = 2.62), indicating that the cross-validated SVM scores

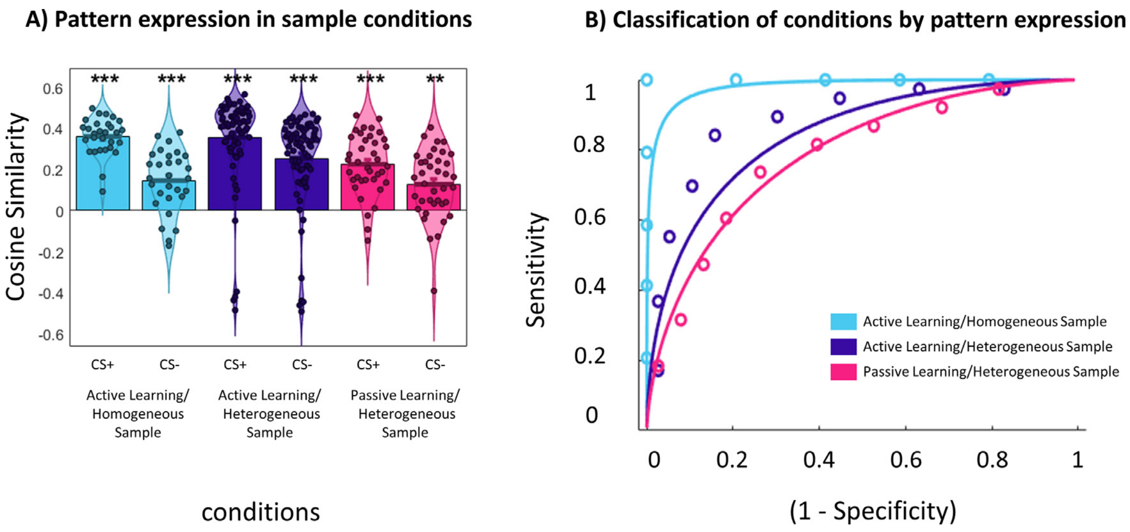

Fig. 4. Classification of Appetitive Data by Aversive Pattern

Note: (A) Bar plot showing cosine similarity between aversive pattern and appetitive sample conditions with each subject as a dot, violin plots illustrating the data distribution and error bars indicating standard error of means. (B) ROC plot showing aversive pattern performance on appCS+ vs. appCS- classification of data from all three samples. The threshold for classification, calculated with optimal balanced error rate was 0.0 for all samples. *** indicates $P < .001$, ** indicates $P < .01$.

Table 3

Performance of avCS+ > avCS- pattern classifying appCS+ versus appCS- conditions in three datasets. Accuracy with standard error (SE), specificity and sensitivity with confidence interval (CI) are presented to demonstrate the performance of the patterns using forced choice classification. Effect size indicates Cohen's d. *** indicates $p < .001$, ** indicates $p < .01$

Dataset	Accuracy (%)		Specificity (%)		Sensitivity (%)		Effect Size
		SE		CI		CI	
Active Learning/Homogen.	100***	0.0	100	100-100	100	100-100	2.08
Active Learning/Heterogen.	84***	4.0	84	76-91	84	76-91	1.05
Passive Learning/Heterogen.	74**	7.1	74	58-86	74	59-88	0.80

were higher for appCS+ than appCS- in every subject. The classifier trained on the Active Learning/Heterogeneous Sample performed moderately accurate (accuracy = 91%, d = 1.93) as did the classifier trained on the Passive Learning/Heterogeneous Sample (accuracy = 89%, *d* = 1.58). Accuracy was significantly above chance level (50%) as assessed with a binomial test for all classifiers ($P < .001$). Specificity, sensitivity, effect size, and accuracy for all three samples are presented in supplemental Table 1 (see also Supplemental Fig. 1D).

In the Active Learning/Homogeneous Sample predictive weight map, clusters significantly predicting the appCS+ versus appCS- condition were found. Clusters with positive effects (i.e. associated with the appCS+ compared to appCS-) were located in the NAcc, caudate nucleus, putamen, brainstem, cerebellum and somatomotor cortex. The weight maps of the Active Learning/Heterogeneous Sample and the Passive Learning/Heterogeneous Sample were predictive over the whole brain. There were no clusters limited to specific brain regions, which reached significance (all $P > .05$, FDR-corrected). All significant clusters from Active Learning/Homogeneous Sample are shown in supplemental Table 2.

4. Discussion

The goal of this study was to integrate neuroimaging findings from aversive with appetitive conditioning paradigms to empirically identify commonalities, and to show the feasibility of cross-paradigm integration with this example. Similarity of these processes in the brain has already been hypothesized but based mainly on qualitative literature reviews (Menon and Uddin, 2010; Moscarello and LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). We wanted not only to quantitatively assess the aversive pattern expression in an appetitive sample but also to determine if results would generalize across multiple appetitive conditioning datasets with differences in task, procedure, instruction, and sample makeup. To address this question, we analyzed three independent previously published appetitive conditioning datasets: The Active Learning/Homogeneous Sample (Kruse et al., 2018), the Active Learning/Heterogeneous Sample (Kruse et al., 2020) and the Passive Learning/Heterogeneous Sample (Tapia León et al., 2019). The aversive conditioning pattern was expressed significantly in the activation maps of all three appetitive conditioning datasets. Furthermore, we were able to accurately classify appCS+ from appCS- in all samples using the aversive pattern. These results provide robust empirical evidence for aversive and appetitive learning processes sharing common neural mechanisms.

The results are in line with previous research (Carter et al., 2009; Lake et al., 2019; Sankar et al., 2019) and we are now able to quantify the long-assumed similarity of aversive and appetitive learning processes at a neural level (Menon and Uddin, 2010; Moscarello and LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). Our results suggests that the activation difference between avCS+ and avCS- contains neural activation which is independent of UCS valence. This common activation might represent the acquired salience of both CS+ (Ogawa et al., 2013; Treviño, 2015). Furthermore, our results are consistent with a regional overlap in activation related to both negative and positive affective processing (Satpute et al., 2015) and appetitive and aversive prediction errors (Corlett et al., 2022). Animal studies recording activity from single cells and neuron populations have also shown that while appetitive and aversive CS+ may evoke distinct neural responses, they are often co-localized in the same anatomical areas (O'Neill et al., 2018; Shabel and Janak, 2009; Tye et al., 2010; Xiu et al., 2014). Thus, the commonalities we found may also reflect valence-specific activation in the same voxels. Finally, no pattern re-

Table 4
Mean cosine similarity of control patterns to appCS+ > appCS- contrast for whole brain with standard error, statistics and effect size (columns 1-7). Comparison between similarity of the control pattern and similarity of the avCS+>avCS- pattern with the respective appCS+ > appCS- contrast is shown with statistics (columns 8-9).

							Comparison with mean cosine similarity of avCS+>avCS- pattern	
Pattern	Dataset	Cosine similarity	SE	T	p	Cohens d	T	p
Cognitive Control, Kragel, Kano et al. (2018)	Active Learning/Homogen.	-0.009	0.010	-0.91	0.371	-0.17	9.18	<.001
	Active Learning/Heterogen.	-0.002	0.005	-0.31	0.758	-0.04	8.38	<.001
	Passive Learning/Heterogen.	-0.007	0.009	-0.76	0.452	-0.12	3.99	<.001
Cognitive Demand (Stroop; Silvestrini et al. 2020)	Active Learning/Homogen.	0.076	0.011	7.03	<.001	1.31	8.87	<.001
	Active Learning/Heterogen.	0.050	0.007	7.07	<.001	0.81	7.53	<.001
	Passive Learning/Heterogen.	0.054	0.012	4.60	<.001	0.75	3.56	.001
Fearful, Kragel and LaBar (2015)	Active Learning/Homogen.	-0.029	0.008	-3.70	<.001	-0.69	12.26	<.001
	Active Learning/Heterogen.	-0.008	0.005	-1.63	.108	-0.19	8.81	<.001
	Passive Learning/Heterogen.	0.008	0.007	1.07	.291	0.17	3.87	<.001
Surprise, Kragel and LaBar (2015)	Active Learning/Homogen.	<0.000	0.010	-0.02	.988	-0.00	11.32	<.001
	Active Learning/Heterogen.	<0.000	0.007	-0.04	.968	-0.01	8.49	<.001
	Passive Learning/Heterogen.	0.010	0.009	1.13	.264	-0.18	3.77	<.001
Picture Induced Negative Affect, Chang et al. (2015)	Active Learning/Homogen.	0.002	0.004	0.56	.583	0.10	10.67	<.001
	Active Learning/Heterogen.	-0.001	0.003	-0.25	.805	-0.03	8.68	<.001
	Passive Learning/Heterogen.	0.005	0.004	1.33	.192	0.22	3.99	<.001

lated to cognitive task demands or affective processing was expressed as highly in the appetitive conditioning data as the aversive conditioning pattern. This indicates that their similarity may be in part specific to the underlying motivational learning processes and not exclusively due to common task demands or basic sensory features.

In addition to similarity over the whole brain, we also found high similarity in the NAcc, caudate nucleus, putamen amygdala, thalamus, insula and cerebellum ROIs. This fits well with the roles of NAcc in reward and loss anticipation (Oldham et al., 2018), caudate nucleus in processing motivational values of actions (Balleine and O'Doherty, 2010), putamen in stimulus-response associations (Everitt and Robbins, 2013), amygdala in representing the CS-UCS relationship (Moscarello and LeDoux, 2013) and the cerebellum in predictive coding and motor responses (Lange et al., 2015) found in past conditioning studies. The thalamus is likely important as a sensory input region for the amygdala in both appetitive and aversive conditioning (Gründemann, 2021; Tye et al., 2008) while the insula may be involved in learning under uncertainty (Gorka et al., 2016; Morriss et al., 2019). Co-localization of aversive and appetitive learning responses in amygdala (O'Neill et al., 2018; Shabel and Janak, 2009; Tye et al., 2010) and striatal regions (Xiu et al., 2014) have already been found in animal studies and more recently, in a human fMRI meta-analysis (Corlett et al., 2022). Here, similarity was most notably high in NAcc and caudate nucleus, indicating that these striatal regions especially may be crucial for motivational salience learning. Further underpinning this interpretation, the SVM classifier trained on appetitive data only (Active Learning/Homogeneous Sample) also revealed clusters predicting appCS+ vs. appCS- in the NAcc, caudate nucleus and cerebellum (see supplemental Table 1). Importantly, while co-localized fMRI activation in these regions points to them being involved in appetitive as well as aversive learning, it may not necessarily indicate them performing the same functions during aversive and appetitive conditioning. For example, animal evidence suggests that activation in the NAcc shell indicates the motivational valence of both an appCS+ and an avCS+ arranged along a rostrocaudal gradient with more anterior activation indicating positive valence (approach signal) and more posterior activation indicating negative valence (avoidance signal; Berridge and Kringelbach 2015). Activation in the NAcc core most likely indicates an unsigned motivational salience signal based on the input it receives from the ventral tegmental areal (Bromberg-Martin et al., 2010). Thus, combined signals from the NAcc may be important for approach behavior in appetitive conditioning and avoidance behavior in aversive conditioning (Gentry et al., 2019) but signal motivational salience of the CS+ in both.

Our findings are particularly relevant since altered aversive and appetitive conditioning are considered the basis for psychological disorders characterized by excessive avoidance and approach behavior, respectively (Duits et al., 2015; Martin-Soelch et al., 2007). As of yet, very little is known about commonalities and overlaps between these disorder categories. Here, we have provided proof of concept for an approach which facilitates finding commonalities in such separate concepts. Further integration of data across more different affective learning paradigms and RDoC domains – and across patient samples - may help fill these knowledge gaps and pave the way towards transdiagnostic biomarkers (Insel, 2014; Woo et al., 2017).

Our results support the practicability of quantitative cross-paradigm integration. We found high whole brain similarity between aversive and appetitive CS+ > CS- contrasts in all samples (see Fig. 3). Effects were larger in some samples than others but present and significantly strong in all of them. These results demonstrate how empirical knowledge can be gained from disparate paradigms by quantifying their similarity. Using an existing software toolbox (https://canlab.github.io; CANlab), and an openly available meta-analysis (Fullana et al., 2016), we could efficiently integrate our current appetitive data with a multitude of past aversive conditioning studies. Empirical cross-paradigm integration has rarely been done up until now – in this study, we could illustrate the feasibility of our approach. Considering the exponential increase in fMRI publications over the last two decades and the difficulties to collect large datasets at single institutions, data integration across studies is becoming an increasingly essential analysis tool. Tools such as these are much needed if we want to better understand the connections between the diverse published evidence and our own data. Here, using this method, we found remarkably high neural similarity with the aversive activation pattern in every appetitive sample included. This enables us to make conclusions not only about the neural similarity of aversive and appetitive learning itself but also about the generalizability of this similarity.

To verify and examine generalizability of the results, we included different appetitive conditioning samples. The otherwise often troublesome fact that many experiments on appetitive conditioning vary in details can be used to our advantage here. By including diverse studies, we can quantify the variance between them and thus try to evaluate how much those details actually affect results while at the same time assessing the generalizability of cross-paradigm similarity. In our analysis, we included three different samples, to examine how generalizable the integration results are. Results were significant across all three samples despite some small differences in effect sizes and classification accuracies. This variance in results may have been due to several reasons: (1) Smaller sample size and increased homogeneity may improve the estimation of experimental variance because of decreased noise. Some studies suggest that increased neural activation variance in conditioning paradigms can be due to hormone fluctuation differences in subjects assigned female at birth, depending on whether they use hormonal birth control (Merz et al., 2018). Thus, samples including mostly cis men may show especially low inter-subject variance. (2) A similar point can be made concerning a more standardized and strictly controlled study protocol – this likely reduces error variance. (3) Less instruction and increased action demands in an appetitive task may lead to it being more arousing overall and thus closer to the (presumably higher) arousal level in an aversive task. Both points (1) and (2) were given in Active Learning/Homogeneous Sample and (3) was a notable difference between Passive Learning and Active Learning samples. The influence of active versus passive task design on the similarity remains to be examined more closely but recent findings suggest that the common neurocircuitry between these types of tasks mirrors the commonalities we found here (mainly ventral and dorsal striatum; Corlett et al. 2022). SVM classification based on the appetitive data only also worked best in Active Learning/Homogeneous Sample, further illustrating how reduced inter-subject variance may improve modeling results generally. Thus, our results highlight the brain activation differences between appetitive conditioning experiments which vary only slightly in task and sample characteristics. At the same time, by integrating over a diversity of methods and samples, we could show that the similarities between patterns of activation associated with aversive and appetitive CS+ can be generalized across this diversity.

In all three samples, we also found the aversive pattern positively expressed in the appCS- condition to varying degrees (see Fig. 4A). Possible explanations for this include: First, basic similarity of the conditions – aversive pattern as well as both appCS+ and appCS- data likely contain activation related to general visual processing, attention etc. leading to a small baseline of similarity. Second, the appCS- may have acquired aversive properties since it signaled absence of a reward (Matsumoto and Hikosaka, 2009; Mollick et al., 2021). This is supported by a post-conditioning drop in appCS- valence ratings in the two Active Learning samples (Kruse et al., 2018, 2020). Part of the appCS- activation data may then reflect these aversive properties. However, the appCS+ condition was still more similar to the avCS+>avCS- pattern, indicating that the pattern primarily codes acquired salience rather than valence.

4.1. Limitations and future directions

Human fMRI data has limited spatial resolution compared to animal studies utilizing methods like single-unit recording or optical imaging. Thus, while we found the BOLD responses to aversive and appetitive CS quite similar at a voxel level, neuronal responses could still be dissociable at a much more precise spatial scale than possible to measure here (e.g. neuronal populations). Another limitation was that the appetitive samples differed in key details but were all collected at the same site. This may have made the overall procedures similar; the scanner itself, other facilities and some of the data collection staff were identical for all samples. Furthermore, while our appetitive learning paradigms are intentionally held similar to classical fear conditioning, the majority of appetitive learning paradigms used in human fMRI are more diverse than this (e.g. reinforcement learning with varying probabilities, risk-taking; Averbeck and Costa 2017, Sherman et al. 2018). The diversity of paradigms out there is a considerable resource that presents countless possibilities for further study with our integration approach. Aversive conditioning data could be integrated with a broader range of appetitive learning datasets that have more procedural and task variance between them. Doing this will enable us to more closely narrow down the factors involved in their similarity. To better understand dissimilarities, integration could also be done with increasingly different paradigms, for example starting with affectively neutral associative conditioning. This could also answer the open question, whether the similarities found here are due to both paradigms involving learning contexts or if the emotional context that they share is more important. Another open question is how the similarity between appetitive and aversive conditioning is mediated by using primary versus secondary UCS. Future studies could disentangle this effect from affective valence by repeating the analysis with more primary appetitive UCS such as food or water instead of money as a secondary reward. Further integration with more different paradigms may reveal how much of the similarity found here can be attributed to common basic features of most fMRI tasks, such as sensory processing, attention and motor action. We included control patterns related to basic cognitive and emotional processing as a first step in this validation process. An alternative to using an aversive meta-analysis pattern as we did here would be to train an aversive conditioning classifier and testing it in appetitive data. Having a sample where each participant performs an appetitive as well as aversive conditioning paradigm would also enable to train aversive conditioning patterns individually for each participant and testing similarity with appetitive conditioning data in the same individual. Such finer-grained aversive patterns may be more precise predictors and also illuminate possible individual differences concerning the similarity between aversive and appetitive learning. Finally, since there are different avenues of integrating data (e.g. principal component analysis), future work could also expand the methods some more to see if more information can be gained from other similarity metrics. Altogether, continuing rigorous cross-paradigm-integration may provide important clinical insights, as it allows to build on existing transdiagnostic approaches to mental health: For instance, the RDoC initiative seeks to characterize mental disorders by impaired functioning in various domains (such as fear and reward learning) rather than existing disorder categories (Insel, 2014). In this framework, fear and reward learning are separate domains, but cross-paradigm-integration findings could demonstrate the benefits of working not only within those domains but across them as well.

5. Conclusion

In conclusion, this study demonstrates the similarity of aversive and appetitive conditioning at fMRI pattern level across multiple independent appetitive datasets. These commonalities may have important implications for etiological models of fear- and reward-related disorders. Enabled by the open science movement and multivariate analysis methods, we could quantitatively integrate past evidence from one paradigm with current data from another. Using the example of aversive and appetitive conditioning, we have demonstrated that this approach is not only viable but extremely valuable when trying to connect data from different paradigms. It presents an opportunity to integrate rather than compare past findings with current studies and thus make better use of the ever-growing body of fMRI studies.

Code and data availability

Data were analyzed using CANlab neuroimaging analysis tools available at https://github.com/canlab/, analysis code specific for this publication available from https://github.com/s-kline/aversive-appetitive-conditioning.

The data from the appetitive conditioning samples are available upon request from the corresponding author. The data are not publicly available due to ethical or privacy restrictions. The meta-analysis pattern is openly available at Neurovault (https://identifiers.org/neurovault.collection:2472).

Declaration of Competing Interest

None.

Credit authorship contribution statement

Sanja Klein: Writing – original draft, Conceptualization, Methodology, Formal analysis, Investigation. **Onno Kruse:** Writing – review & editing, Investigation. **Isabell Tapia León:** Writing – review & editing, Investigation. **Lukas Van Oudenhove:** Writing – review & editing, Methodology. **Sophie R. van 't Hof:** Writing – review & editing. **Tim Klucken:** Writing – review & editing, Resources, Funding acquisition. **Tor D. Wager:** Writing – review & editing, Resources, Software, Methodology. **Rudolf Stark:** Writing – review & editing, Resources, Funding acquisition.

Data Availability

Data will be made available on request.

Funding

The work of SK, TK, and RS on this article was carried out in the context of the Research Unit ACSID, FOR2974, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 411232260. The data collection for the Active Learning/Heterogeneous Sample (DFG KL2500/5-1) and for the Passive Learning/Heterogeneous Sample (DFG KL2500/4-1) was also funded by the German research association.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119594.

References

Averbeck, B.B., Costa, V.D., 2017. Motivational neural circuits underlying reinforcement learning. Nat. Neurosci. 20 (4), 505–512. doi:10.1038/nn.4506.

Balleine, B.W., O'Doherty, J.P, 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35 (1), 48–69. doi:10.1038/npp.2009.131, Official Publication of the American College of Neuropsychopharmacology.

Berridge, K.C., Kringelbach, M.L., 2015. Pleasure systems in the brain. Neuron 86 (3), 646–664. doi:10.1016/j.neuron.2015.02.018.

Bisandu, D.B., Prasad, R., Liman, M.M., 2019. Data clustering using efficient similarity measures. J. Stat. Manag. Syst. 22 (5), 901–922. doi:10.1080/09720510.2019.1565443.

Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., Love, B., 2020. Measures of neural similarity. Comput. Brain Behav. 3 (4), 369–383. doi:10.1007/s42113-019-00068-5.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P., 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 30 (2), 619–639. doi:10.1016/S0896-6273(01)00303-8.

Bromberg-Martin, E.S., Matsumoto, M., Hikosaka, O., 2010. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68 (5), 815–834. doi:10.1016/j.neuron.2010.11.022.

Brooks, A.M., Berns, G.S., 2013. Aversive stimuli and loss in the mesocorticolimbic dopamine system. Trends Cogn. Sci. 17 (6), 281–286. doi:10.1016/j.tics.2013.04.001.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167. https://canlab.github.io.

CANlab. CANlab neuroimaging analysis tools.

Carter, R.M., MacInnes, J.J., Huettel, S.A., Adcock, R.A., 2009. Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. Front. Behav. Neurosci. 3. doi:10.3389/neuro.08.021.2009.

Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A., Wager, T.D., 2015. A sensitive and specific neural signature for picture-induced negative affect. PLoS Biol. 13 (6), e1002180. doi:10.1371/journal.pbio.1002180.

Chase, H.W., Kumar, P., Eickhoff, S., Dombrovski, A.Y., 2015. Reinforcement learning models and their neural correlates: an activation likelihood estimation meta-analysis. Cogn. Affect. Behav. Neurosci. 15 (2), 435–459. doi:10.3758/s13415-015-0338-7.

Corlett, P.R., Mollick, J.A., Kober, H., 2022. Meta-analysis of human prediction error for incentives, perception, cognition, and action. Neuropsychopharmacology doi:10.1038/s41386-021-01264-3, Official Publication of the American College of Neuropsychopharmacology. Advance online publication.

Destoop, M., Morrens, M., Coppens, V., Dom, G., 2019. Addiction, anhedonia, and comorbid mood disorder. A narrative review. Front. Psychiatry 10 (311). doi:10.3389/fpsyt.2019.00311.

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., Ramnani, N., 2009. A probabilistic MR atlas of the human cerebellum. Neuroimage 46 (1), 39–46. doi:10.1016/j.neuroimage.2009.01.045.

Duits, P., Cath, D.C., Lissek, S., Hox, J.J., Hamm, A.O., Engelhard, I.M., van den Hout, M.A., Baas, J.M.P., 2015. Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depress. Anxiety 32 (4), 239–253. doi:10.1002/DA.22353.

Ernst, T.M., Brol, A.E., Gratz, M., Ritter, C., Bingel, U., Schlamann, M., Maderwald, S., Quick, H.H., Merz, C.J., Timmann, D., 2019. The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. Elife 8. doi:10.7554/eLife.46831.

Etkin, A., Wager, T.D., 2007. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. Am. J. Psychiatry 164 (10), 1476–1488. doi:10.1176/appi.ajp.2007.07030504.

Everitt, B.J., Robbins, T.W., 2013. From the ventral to the dorsal striatum: devolving views of their roles in drug addiction. Neurosci. Biobehav. Rev. 37 (9), 1946–1954. doi:10.1016/j.neubiorev.2013.02.010, Pt A.

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. doi:10.1038/mp.2015.88.

Gentry, R.N., Schuweiler, D.R., Roesch, M.R., 2019. Dopamine signals related to appetitive and aversive events in paradigms that manipulate reward and avoidability. Brain Res. 1713, 80–90. doi:10.1016/j.brainres.2018.10.008.

Gorka, S.M., Nelson, B.D., Phan, K.L., Shankman, S.A., 2016. Intolerance of uncertainty and insula activation during uncertain reward. Cogn. Affect. Behav. Neurosci. 16 (5), 929–939. doi:10.3758/s13415-016-0443-2.

Gramfort, A., Thirion, B., Varoquaux, G., 2013. Identifying predictive regions from fMRI with TV-L1 prior. In: Davatzikos, C. (Ed.), Proceedings of the International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, Philadelphia, PA, USA, pp. 17–20. doi:10.1109/PRNI.2013.14 22-24 June 2013.

Gründemann, J., 2021. Distributed coding in auditory thalamus and basolateral amygdala upon associative fear learning. Curr. Opin. Neurobiol. 67, 183–189. doi:10.1016/j.conb.2020.11.014.

Heffley, W., Hull, C., 2019. Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. Elife (8) doi:10.7554/eLife.46764.

Herry, C., Johansen, J.P., 2014. Encoding of fear learning and memory in distributed neuronal circuits. Nat. Neurosci. 17 (12), 1644–1654. doi:10.1038/nn.3869.

Insel, T.R., 2014. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. Am. J. Psychiatry 171 (4), 395–397. doi:10.1176/appi.ajp.2014.14020138.

Kragel, P.A., Kano, M., van Oudenhove, L., Ly, H.G., Dupont, P., Rubio, A., Delon-Martin, C., Bonaz, B.L., Manuck, S.B., Gianaros, P.J., Ceko, M., Reynolds Losin, E.A., Woo, C.W., Nichols, T.E., Wager, T.D., 2018. Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. Nat. Neurosci. 21 (2), 283–289. doi:10.1038/s41593-017-0051-7.

Kragel, P.A., Koban, L., Barrett, L.F., Wager, T.D., 2018. Representation, pattern information, and brain signatures: From neurons to neuroimaging. Neuron 99 (2), 257–273. doi:10.1016/j.neuron.2018.06.009.

Kragel, P.A., LaBar, K.S., 2014. Advancing emotion theory with multivariate pattern classification. Emot. Rev. 6 (2), 160–174. doi:10.1177/1754073913512519.

Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cogn. Affect. Neurosci. 10 (11), 1437–1448. doi:10.1093/scan/nsv032.

Kruse, O., Klein, S., Tapia León, I., Stark, R., Klucken, T., 2020. Amygdala and nucleus accumbens involvement in appetitive extinction. Hum. Brain Mapp. 41 (7), 1833–1841. doi:10.1002/hbm.24915.

Kruse, O., Tapia León, I., Stalder, T., Stark, R., Klucken, T., 2018. Altered reward learning and hippocampal connectivity following psychosocial stress. Neuroimage 171, 15–25. doi:10.1016/j.neuroimage.2017.12.076.

Kruse, O., Tapia León, I., Stark, R., Klucken, T., 2017. Neural correlates of appetitive extinction in humans. Soc. Cogn. Affect. Neurosci. 12 (1), 106–115. doi:10.1093/scan/nsw157.

Lake, J.I., Spielberg, J.M., Infantolino, Z.P., Crocker, L.D., Yee, C.M., Heller, W., Miller, G.A., 2019. Reward anticipation and punishment anticipation are instantiated in the brain via opponent mechanisms. Psychophysiology 56 (8), e13381. doi:10.1111/psyp.13381.

Lange, I., Kasanova, Z., Goossens, L., Leibold, N., Zeeuw, C.I., de van Amelsvoort, T., Schruers, K., 2015. The anatomy of fear learning in the cerebellum: A systematic meta-analysis. Neurosci. Biobehav. Rev. 59, 83–91. doi:10.1016/j.neubiorev.2015.09.019.

Liverant, G.I., Sloan, D.M., Pizzagalli, D.A., Harte, C.B., Kamholz, B.W., Rosebrock, L.E., Cohen, A.L., Fava, M., Kaplan, G.B., 2014. Associations among smoking, anhedonia, and reward learning in depression. Behav. Ther. 45 (5), 651–663. doi:10.1016/j.beth.2014.02.004.

Mackintosh, N.J., 1975. A theory of attention: variations in the associability of stimuli with reinforcement. Psychol. Rev. 82 (4), 276–298. doi:10.1037/H0076778.

Martin-Soelch, C., Linthicum, J., Ernst, M., 2007. Appetitive conditioning: neural bases and implications for psychopathology. Neurosci. Biobehav. Rev. (3) 31. doi:10.1016/j.neubiorev.2006.11.002.

MATLAB R2012a, 2012. The MathWorks Inc.

Matsumoto, M., Hikosaka, O., 2009. Representation of negative motivational value in the primate lateral habenula. Nat. Neurosci. 12 (1), 77–84. doi:10.1038/nn.2233.

Mechias, M.L., Etkin, A., Kalisch, R., 2010. A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. Neuroimage 49 (2), 1760–1768. doi:10.1016/j.neuroimage.2009.09.040.

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214 (5-6), 655–667. doi:10.1007/s00429-010-0262-0.

Merz, C.J., Kinner, V.L., Wolf, O.T., 2018. Let's talk about sex … differences in human fear conditioning. Curr. Opin. Behav. Sci. 23, 7–12. doi:10.1016/j.cobeha.2018.01.021.

Mollick, J.A., Chang, L.J., Krishnan, A., Hazy, T.E., Krueger, K.A., Frank, G.K.W., Wager, T.D., O'Reilly, R.C, 2021. The neural correlates of cued reward omission. Front. Hum. Neurosci. 15, 615313. doi:10.3389/fnhum.2021.615313.

Morriss, J., Gell, M., van Reekum, C.M., 2019. The uncertain brain: a co-ordinate based meta-analysis of the neural signatures supporting uncertainty during different contexts. Neurosci. Biobehav. Rev. 96, 241–249. doi:10.1016/j.neubiorev.2018.12.013.

Moscarello, J.M., LeDoux, J.E., 2013. The contribution of the amygdala to aversive and appetitive pavlovian processes. Emot. Rev. 5 (3), 248–253. doi:10.1177/1754073913477508.

Ogawa, M., van der Meer, M.A.A., Esber, G.R., Cerri, D.H., Stalnaker, T.A., Schoenbaum, G., 2013. Risk-responsive orbitofrontal neurons track acquired salience. Neuron 77 (2), 251–258. doi:10.1016/j.neuron.2012.11.006.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., Lorenzetti, V., 2018. The anticipation and outcome phases of reward and loss processing: a neuroimaging meta-analysis of the monetary incentive delay task. Hum. Brain Mapp. 39 (8), 3398–3418. doi:10.1002/hbm.24184.

O'Neill, P.K., Gore, F., Salzman, C.D., 2018. Basolateral amygdala circuitry in positive and negative valence. Curr. Opin. Neurobiol. 49, 175–183. doi:10.1016/j.conb.2018.02.012.

Pauli, W.M., O'Reilly, R.C., Yarkoni, T., Wager, T.D, 2016. Regional specialization within the human striatum for diverse psychological functions. Advance online publication. Proc. Natl. Acad. Sci. 113 (7), 1907–1912. doi:10.1073/PNAS.1507610113.

Popa, L.S., Ebner, T.J., 2018. Cerebellum, predictions and errors. Front. Cell. Neurosci. 12 (524). doi:10.3389/fncel.2018.00524.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2016. Discrete neural signatures of basic emotions. Cereb. Cortex 26 (6), 2563–2573. doi:10.1093/cercor/bhv086.

Sankar, A., Yttredahl, A.A., Fourcade, E.W., Mickey, B.J., Love, T.M., Langenecker, S.A., Hsu, D.T., 2019. Dissociable neural responses to monetary and social gain and loss in women with major depressive disorder. Front. Behav. Neurosci. 13 (149). doi:10.3389/fnbeh.2019.00149.

Satpute, A.B., Kang, J., Bickart, K.C., Yardley, H., Wager, T.D., Barrett, L.F., 2015. Involvement of sensory regions in affective experience: a meta-analysis. Front. Psychol. 6. doi:10.3389/fpsyg.2015.01860.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27 (9), 2349–2356. doi:10.1523/JNEUROSCI.5587-06.2007.

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. doi:10.1371/journal.pone.0005865.

Shabel, S.J., Janak, P.H., 2009. Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. Proc. Natl. Acad. Sci. U.S.A. 106 (35), 15031–15036. doi:10.1073/pnas.0905580106.

Sherman, L., Steinberg, L., Chein, J., 2018. Connecting brain responsivity and real-world risk taking: Strengths and limitations of current methodological approaches. Dev. Cogn. Neurosci. 33, 27–41. doi:10.1016/j.dcn.2017.05.007.

Silvestrini, N., Chen, J.I., Piché, M., Roy, M., Vachon-Presseau, E., Woo, C.W., Wager, T.D., Rainville, P., 2020. Distinct fMRI patterns colocalized in the cingulate cortex underlie the after-effects of cognitive control on pain. Neuroimage 217, 116898. doi:10.1016/j.neuroimage.2020.116898.

Statistical Parametric Mapping (SPM 12), 2014. Wellcome Department of Cognitive Neurology, London, UK.

Stefanova, E., Dubljević, O., Herbert, C., Fairfield, B., Schroeter, M.L., Stern, E.R., Urben, S., Derntl, B., Wiebking, C., Brown, C., Drach-Zahavy, A., Loeffler, Kathrin, L, A., Albrecht, F., Palumbo, R., Boutros, S.W., Raber, J., Lowe, L, 2020. Anticipatory feelings: neural correlates and linguistic markers. Neurosci. Biobehav. Rev. 113, 308–324. doi:10.1016/j.neubiorev.2020.02.015.

Tapia León, I., Kruse, O., Stark, R., Klucken, T., 2019. Relationship of sensation seeking with the neural correlates of appetitive conditioning. Soc. Cogn. Affect. Neurosci. 14 (7), 769–775. doi:10.1093/scan/nsz046.

Tharwat, A., 2021. Classification assessment methods. Appl. Comput. Inform. 17 (1), 168–192. doi:10.1016/j.aci.2018.08.003.

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

Treviño, M., 2015. Associative learning through acquired salience. Front. Behav. Neurosci. 9 (353). doi:10.3389/fnbeh.2015.00353.

Tye, K.M., Cone, J.J., Schairer, W.W., Janak, P.H., 2010. Amygdala neural encoding of the absence of reward during extinction. J. Neurosci. 30 (1), 116–125. doi:10.1523/JNEUROSCI.4240-09.2010.

Tye, K.M., Stuber, G.D., Ridder, B.de, Bonci, A., Janak, P.H., 2008. Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. Nature 453 (7199), 1253–1257. doi:10.1038/nature06963.

van Oudenhove, L., Kragel, P.A., Dupont, P., Ly, H.G., Pazmany, E., Enzlin, P., Rubio, A., Delon-Martin, C., Bonaz, B., Aziz, Q., Tack, J., Fukudo, S., Kano, M., Wager, T.D., 2020. Common and distinct neural representations of aversive somatic and visceral stimulation in healthy individuals. Nat. Commun. 11 (1), 5939. doi:10.1038/s41467-020-19688-8.

van 't Hof, S.R., van Oudenhove, L., Janssen, E., Klein, S., Reddan, M.C., Kragel, P.A., Stark, R., Wager, T.D., 2021. The brain activation-based sexual image classifier (BASIC): a sensitive and specific fMRI activity pattern for sexual image processing. Cereb. Cortex doi:10.1093/cercor/bhab397, Advance online publication.

Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.W., Kross, E., 2013. An fMRI-based neurologic signature of physical pain. N. Engl. J. Med. 368 (15), 1388–1397. doi:10.1056/NEJMoa1204471.

Weaverdyck, M.E., Lieberman, M.D., Parkinson, C., 2020. Tools of the trade. Multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. Soc. Cogn. Affect. Neurosci. 15 (4), 487–509. doi:10.1093/scan/nsaa057.

Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 20 (3), 365–377. doi:10.1038/nn.4478.

Xie, J., Fang, P., Zhang, Z., Luo, R., Dai, B., 2021. Behavioral inhibition/activation systems and depression among females with substance use disorder: the mediating role of intolerance of uncertainty and anhedonia. Front. Psychiatry 12, 644882. doi:10.3389/fpsyt.2021.644882.

Xiu, J., Zhang, Q., Zhou, T., Zhou, T.T., Chen, Y., Hu, H., 2014. Visualizing an emotional valence map in the limbic forebrain by TAI-FISH. Nat. Neurosci. 17 (11), 1552–1559. doi:10.1038/nn.3813.