RESEARCH ARTICLE

WILEY

Pattern classification based on the amygdala does not predict an individual's response to emotional stimuli

Tim Varkevisser[1,2,3] | Elbert Geuze[1,2] | Max A. van den Boom[4,5] |
Karlijn Kouwer[6] | Jack van Honk[3,7] | Remko van Lutterveld[1,2]

[1]University Medical Center, Utrecht, The Netherlands

[2]Brain Research and Innovation Center, Ministry of Defence, Utrecht, The Netherlands

[3]Utrecht University, Utrecht, The Netherlands

[4]Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, Minnesota, USA

[5]Department of Neurology and Neurosurgery, University Medical Center, Utrecht, The Netherlands

[6]Department of Biological and Medical Psychology, Faculty of Psychology, University of Bergen, Bergen, Norway

[7]University of Cape Town, Cape Town, South Africa

Correspondence
Tim Varkevisser, Brain Research and Innovation Center, Dutch Ministry of Defence, Lundlaan 1, 3584 EZ Utrecht, The Netherlands.
Email: t.varkevisser-2@umcutrecht.nl

Abstract

Functional magnetic resonance imaging (fMRI) studies have often recorded robust univariate group effects in the amygdala of subjects exposed to emotional stimuli. Yet it is unclear to what extent this effect also holds true when multi-voxel pattern analysis (MVPA) is applied at the level of the individual participant. Here we sought to answer this question. To this end, we combined fMRI data from two prior studies ($N = 112$). For each participant, a linear support vector machine was trained to decode the valence of emotional pictures (negative, neutral, positive) based on brain activity patterns in either the amygdala (primary region-of-interest analysis) or the whole-brain (secondary exploratory analysis). The accuracy score of the amygdala-based pattern classifications was statistically significant for only a handful of participants (4.5%) with a mean and standard deviation of 37% ± 5% across all subjects (range: 28–58%; chance-level: 33%). In contrast, the accuracy score of the whole-brain pattern classifications was statistically significant in roughly half of the participants (50.9%), and had an across-subjects mean and standard deviation of 49% ± 6% (range: 33–62%). The current results suggest that the information conveyed by the emotional pictures was encoded by spatially distributed parts of the brain, rather than by the amygdala alone, and may be of particular relevance to studies that seek to target the amygdala in the treatment of emotion regulation problems, for example via real-time fMRI neurofeedback training.

KEYWORDS
amygdala, emotion, fMRI, machine learning, multi-voxel pattern analysis (MVPA), pattern classification, task reactivity

1 | INTRODUCTION

Traditional views on the neural basis of emotion often ascribe a central role to the amygdala (Phelps & LeDoux, 2005). In many such views, the amygdala is considered the key component of the neural apparatus responsible for the generation and expression of affect (LeDoux, 2003; Murray, 2007). The core premise behind this framework is straightforward: When an emotional stimulus is encountered by an individual, an amygdala response ensues—one that can be measured under experimental conditions with functional magnetic resonance imaging (fMRI). If that response is disproportionate—as is the case in individuals with affect regulation problems—emotions like fear

and anger have the opportunity to run rampant. These ideas have permeated the neuroscientific landscape to such an extent that they have become textbook knowledge (e.g., see Lane & Nadel, 2002).

Over the last few decades, a multitude of fMRI tasks have been developed to help researchers induce an emotional state in participants (see Phan et al., 2002 for a review). Many of these tasks assume that a neural response can be evoked in the amygdala when participants are exposed to emotional pictures (see Costafreda et al., 2008 for an overview). For example, in the 'Hariri hammer task'—a task named after its renowned ability to induce amygdala activation at a group-level—pictures of human faces that bear an emotional expression are shown to participants (Hariri et al., 2000). In our own work, we have often employed the emotional processing task by Van Buuren et al. (2011), which instead utilizes photographs of objects or scenes to evoke an emotional state in study participants; this task too has been shown to elicit robust activation of the amygdala when measured across subjects (Heesink et al., 2018; Van Buuren et al., 2011; Van Rooij et al., 2014).

While many of the emotion provocation tasks that are used today can indeed induce a response inside the amygdala that is robust across participants, recent evidence suggests that the reproducibility of such task effects may actually be quite low at the single-subject level (Elliott et al., 2020; Nord et al., 2017). This calls into question the utility and suitability of task-evoked amygdala activation as an interindividual biomarker of emotion-related processing—a notion that may have far-reaching consequences for the validity of targeted intervention strategies, such as real-time fMRI neurofeedback training (e.g., see Nicholson et al., 2017; Young et al., 2014 and; see also Linhartová et al., 2019 for a review). In fact, many of the fMRI tasks commonly used to target the amygdala can trace their origins to a branch of experimental research that focuses on the discovery of average group effects, rather than differences between individuals. The statistical practices put in place to analyse such data tend to be somewhat ill-suited to uncover task effects that occur at the level of the individual participant (Infantolino et al., 2018). This lack of intra-subject sensitivity can be explained—at least in part—by the fact that many of these methods are unable to model the covariance that may exist between (adjacent) voxels; that is, they can only measure task effects circumscribed to isolated voxels (Mahmoudi et al., 2012).

Multi-voxel pattern analysis (MVPA) offers a viable alternative to these standard mass-univariate statistics—one that more readily allows for inferences to be made at the single-subject level, due to its ability to capture spatially distributed patterns of brain activity (e.g., see Arbabshirani et al., 2017 for a review). Here, we sought to apply MVPA in order to determine whether the patterns of BOLD activity inside a participant's amygdala can be used to predict the valence categories (negative, neutral, or positive) of affective pictures during an emotional processing task. To achieve this goal, we combined the fMRI measurements from two prior studies (total $N = 112$) (Heesink et al., 2018; Van Rooij et al., 2014), and—for each participant—trained a linear support vector machine (SVM) to classify the valence of emotional pictures based on patterns of brain activation both inside the amygdala (primary region-of-interest [ROI]

analysis) and within the whole-brain (secondary exploratory analysis). We hypothesized that classification based on the amygdala alone would already be conducive to predict the valence category of emotional pictures, but that performance would improve when regions other than the amygdala were added. Indeed, such results would align with previous works that have explored the use of MVPA for emotion classification (Baucom et al., 2012; Bush et al., 2018; Habes et al., 2013; Saarimäki et al., 2016; Yuen et al., 2012). Although previous studies have reported significant increases in task-induced amygdala activation in veterans with emotion dysregulation problems (i.e., relative to healthy controls), such as posttraumatic stress disorder (PTSD) (Bryant et al., 2008; Shin et al., 2006) or intermittent explosive disorder (IED) (Coccaro et al., 2007; McCloskey et al., 2016), no such group effects could be recorded by Heesink et al. (2018) and Van Rooij et al. (2014)—the two studies from which we drew our study sample and data. We therefore did not expect to find any significant effects of psychiatric diagnosis here.

2 | MATERIALS AND METHODS

2.1 | Participants

Neuroimaging data from two prior studies at our department were combined in the present work: (1) the Biological Effects of Traumatic Experiences, Treatment, and Recovery study (BETTER; Van Rooij et al., 2014), and (2) the Military Aggression Regulation Study (MARS; Heesink et al., 2018). Both studies were approved by the Medical Ethical Review Board of the University Medical Centre Utrecht in the Netherlands. Both research projects were in accordance with the Declaration of Helsinki. The original aim of the BETTER study was to identify possible biomarkers of treatment response in military veterans suffering from PTSD. In the BETTER study, functional MR images were acquired both before and 6–8 months after PTSD treatment, however, only the pre-treatment data were used here; for an in-depth description of the treatment response findings, see Van Rooij et al. (2015). The main goal of the MARS study was to identify potential biomarkers of impulsive aggression problems in a veteran sample. A total of 57 veterans with a primary diagnosis of PTSD were recruited in the BETTER project, along with 29 veterans that had no history of mental health issues. Another 29 veterans with impulsive aggression problems were included in the MARS study, as well as 30 non-aggressive veteran controls. One participant was enrolled in both studies and was therefore excluded from the BETTER dataset. Of the remaining 144 participants eligible for inclusion in the current work, one individual was excluded on account of her being the only woman in the combined dataset, three participants had to be excluded because of missing scan data, nine participants were excluded due to large artefacts in the raw MR images, eight participants were excluded because of bad placement of the field-of-view (FOV), six individuals were excluded due to unreliable registration of the behavioural emotional processing task scan data, and two participants were excluded because they gave too many incongruent responses during the

emotional processing task, which we defined as >75% incongruent responses for any stimulus category. Finally, three participants were excluded after quality control of the pre-processed scan data (see Section 2.4 for further details). Thus, a total of 112 participants were included in this study.

2.2 | Data acquisition

All MR imaging data of both the BETTER and MARS protocols were collected on the same 3 Tesla Philips Achieva system (Phillips Medical Systems, Best, the Netherlands) using the same acquisition parameters. Functional image runs consisted of 322 T2*-weighted echo planar images that were acquired interleaved with the following settings: TR = 1600 ms; TE = 23 ms; FA = 72.5°; FOV = 256 × 208 × 120 mm; 30 transverse slices; matrix = 64 × 64; voxel size = 4 × 4 × 3.60 mm, 0.4 mm gap. Due to the tilted angle of the FOV placement of these images, some parts of the occipital cortex and most of the cerebellum could not be imaged (see Supplementary Figure S1). A high-resolution 3D sensitivity encoded (SENSE) T1-weighted anatomical image was collected to facilitate spatial normalization and localization: TR = 10 ms; TE = 4.6 ms; FA = 8°; FOV = 240 × 240 × 160; 200 sagittal slices; matrix = 304 × 299.

2.3 | Experimental paradigm

A total of 96 pictures were taken from the International Affective Picture System (IAPS) and incorporated into the emotional processing task (Lang et al., 1997). Thirty-two pictures of each of the valence categories neutral, negative, and positive were extracted from the IAPS database. Each trial consisted of a single stimulus presented for a duration of 2 s, followed by an evaluation screen asking the participant to rate the picture as neutral, negative, or positive by pressing one of three buttons with the thumb of their right hand. The response period had a maximum duration of 2 s; if a response was given prior to the full elapse of this interval, a fixation cross was shown for the remainder of that interval. The task consisted of four blocks of 24 pictures presented in pseudorandomized order, with each block containing 8 pictures per valence category. Each block of pictures was followed by a 32 s rest period during which a fixation cross was shown to the participant. Figure 1a displays a schematic overview of the emotional processing task, which is identical to the one detailed previously by Heesink et al. (2018), Van Buuren et al. (2011), Van Rooij et al. (2014, 2015), and Vink et al. (2014). These studies all observed robust group-level amygdala activation in both civilian and military samples.

2.4 | Image pre-processing

The functional images were slice-time corrected and re-aligned. The anatomical image was segmented and co-registered to the mean

functional image in order to transform the grey matter, white matter, and cerebrospinal fluid (CSF) maps to native space. All pre-processing steps were performed in SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK) with default settings unless otherwise specified. Jenkinson's algorithm was used to calculate framewise displacement (FD) of the head (Jenkinson et al., 2002), and in-house software written in MATLAB 9.5 (The MathWorks Inc., Natick, MA, 2018) was used to first binarize the white-matter and CSF segmentation masks at an (inclusive) probability threshold of >80%, and subsequently erode each of these two masks by 1 voxel in each of the cardinal dimensions x, y, and z in order to avoid partial volume sharing with grey matter (Chai et al., 2012; as in Varkevisser et al., 2017). Quality control of the pre-processed data led to the exclusion of one participant due to an unresolved artefact that arose during segmentation. In addition, two participants had to be excluded due to excessive head motion, which we defined as more than 25% of scan volumes with an FD above 0.5 mm (Siegel et al., 2014).

2.5 | Trial pattern estimation (feature extraction)

Brain activation patterns for each of the 96 picture trials in a participant's task run were estimated in SPM12 by using the least-squares separate (LS-S) method of Mumford et al. (2012). This technique enables one to disentangle the task activation patterns of single adjacent trials in a fast-event related design. The LS-S approach fits one general linear model (GLM) for each trial in a task run, all the while controlling for the combined effect of all other trials in the time-series for any given model (Mumford et al., 2012; see also Figure 1b). Translated to the current situation, the LS-S approach was implemented by creating a single GLM for each trial in a participant's task run. The stimulus onset and duration were modelled as predictor-of-interest, alongside a second regressor that simultaneously modelled the onsets and durations of all other remaining trials in the task run. The white matter and CSF signals—calculated by taking the average of all voxels in the corresponding segmentation mask for each of the 322 scan volumes—and the six motion realignment parameters (RPs) were added as additional covariates. To further limit possible confounding that may be introduced by in-scanner head motion, we also added the temporal derivatives of the six RPs (calculated using the backward-shifted method) as well as their quadratic terms, and the quadratic terms of the derivatives to the individual models (Friston et al., 1996). An intercept term was also included in the GLMs. A high-pass filter with a cut-off frequency of 1/128 Hz was applied to the data in order to adjust for low-frequency scanner drift. The trial-wise subject-level beta-weight maps yielded by the LS-S method were converted into t-statistic maps by dividing each element by its own standard error (Misaki et al., 2010). Importantly, only those trials for which the participant's response was congruent with the original neutral, positive, or negative IAPS rating were included in the LS-S extraction process; all incongruent trials were excluded from the MVPAs. Note that this led to a variable number of trials per valence category for each participant. All participants with >75% incongruent responses for any
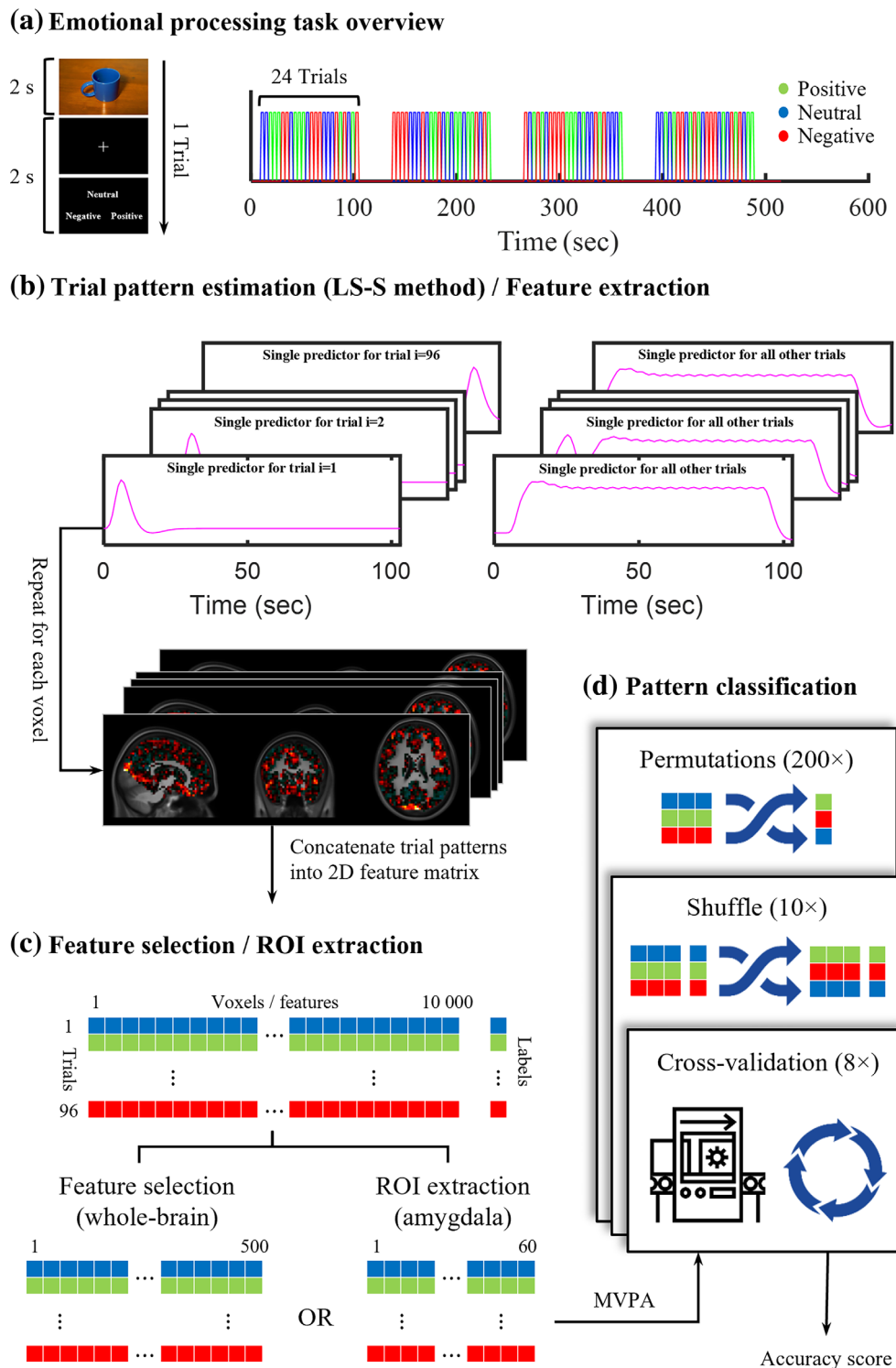
FIGURE 1 Schematic overview of the participant-level analysis pipeline. (a) Overview of the emotional processing task. (b) Trial pattern estimation strategy. The LS-S method was used to fit one GLM for each trial in a participant's task run (Mumford et al., 2012). Per GLM, the stimulus onset and duration of the trial in question were modelled as predictor-of-interest, alongside a second regressor that simultaneously modelled the onsets and durations of all other remaining trials in the task run. The white matter and CSF signals, the six RPs, as well as their expansion terms (i.e., temporal derivatives, quadratic terms, and quadratic terms of derivatives) were added as additional covariates-of-no-interest. An intercept term was also added to each trial's GLM. (c) Feature selection or ROI extraction. As part of the whole-brain analyses, data reduction was applied via a combination of univariate feature selection and recursive feature elimination (see main text for further details). For the ROI-analyses, no additional data reduction steps were applied besides selecting only those voxels within the amygdala ROI mask. (d) MVPA. The pattern classifier was embedded in a repeated stratified k-fold cross-validation scheme that randomly shuffled the temporal arrangement of all trials within a participant's task run a total of ten times and conducted an eight-fold cross-validation for each repeat (or shuffle). Permutations testing was conducted to assess the statistical significance of the classifier's performance (200 permutations). GLM, general linear model; LS-S, least-squares separate; MVPA, multi-voxel pattern analysis; ROI, region-of-interest

stimulus category were excluded from analysis. Crucially, this threshold translates to a minimum of eight congruent trials per picture category, which is the minimum number of examples required to have at least one unique trial pattern per category, in each data-split of our stratified eight-fold cross-validation scheme (see Section 2.8 for further details).

2.6 | Amygdala ROI definition

The primary goal of the current study was to determine the extent to which patterns of BOLD activation in the amygdala can be used to predict the valence categories of a single participant's picture trials (neutral, negative, positive) during an emotional processing task. Our ROI—the amygdala—was defined based on probabilistic maps of the basolateral, centromedial, and superficial sub-nuclei obtained from the Anatomy toolbox of SPM12 (Amunts et al., 2005; Eickhoff et al., 2005). The sum of these probabilistic maps was binarized at a threshold of ≥0.5 to generate a mask for the whole amygdala. The resulting binary mask was then resliced to the same reference space and resolution as the pre-processed functional images of each participant, by using the deformation fields generated during segmentation. All classification analyses were carried out in native space. The amygdala mask had a mean and standard deviation of 68 ± 6 voxels across all participants (range: 55–85 voxels). No additional feature selection was conducted for the ROI-based MVPAs (see also Figure 1c).

2.7 | Whole brain feature selection

The secondary aim of the current study was to determine the extent to which patterns of BOLD activation inside the whole-brain can be used to predict the valence categories of a single participant's picture trials (neutral, negative, positive) during the emotional processing task. To accomplish this goal, feature selection was conducted to prevent overfitting our pattern classifier due to the large number of voxels included in the participant-level grey matter masks (see Figure 1c) (Pereira et al., 2009). We applied a two-step feature reduction strategy that combined an initial crude selection based on univariate t-testing (Habes et al., 2013; see also Yuen et al., 2012), followed by a more fine-grained selection of the remaining voxels via recursive feature elimination (RFE; De Martino et al., 2008). In the first step, (one sample) t-values were computed in order to quantify the level of activation of each voxel across all trials belonging to a given valence category (e.g., neutral). Note that three of these t-values were thus calculated for each voxel: one for each of the three valence categories. For each valence category, we then selected the top 50% of the voxels with the highest t-values (Habes et al., 2013; Yuen et al., 2012). The union of these three voxel selections constituted the initial set of selected features and represented the collection of voxels that were activated most by the task across all stimulus categories (Van den Boom et al., 2019). This initial set of features was subjected to a second round of more fine-grained data reduction that selected

the top 5% of the remaining voxels via RFE (De Martino et al., 2008). At each iteration of the RFE, 20% of the feature set were removed by the algorithm. Thus, while many features were rejected at the first iteration, in a relative sense, progressively fewer and fewer features were removed at all subsequent iterations as the algorithm honed down on its pre-set target of 5% of all voxels in the grey matter mask. By setting the step size of the RFE to 20%, we were able to limit computational resources—which were quite substantial when combined with our permutation testing strategy (see Section 2.9 for further details). The value of 20% was chosen based on prior work by Wottschel et al. (2019), who found this step size to be an agreeable compromise between computation time and classification performance. Across all participants, a mean of 623 grey matter voxels remained after feature selection (standard deviation [SD]: 56 voxels; range: 503–814 voxels).

2.8 | Multi-voxel pattern analysis

MVPA was performed by training an SVM with a linear kernel to classify the valence categories (negative, neutral, or positive) of individual trials within a participant's task run, based on the associated patterns of BOLD activation inside the amygdala or whole-brain. To this end, we implemented the '*SVC*' module from scikit-learn using in-house code written in Python 3.7 (Pedregosa et al., 2011; Van Rossum & Drake Jr, 1995). The regularization parameter of the support vector classifier was fixed to C = 1 and was set to adjust for possible class imbalance. This adjustment was necessary as only congruent trials were included in the classification process, meaning that the number of instances could, in theory (and did in practice), differ between classes for any given participant. The decision function of the algorithm was set to one-versus-rest ('*ovr*') in order to accommodate for the multi-class nature of our situation. The pattern classifier was embedded in a repeated stratified k-fold cross-validation scheme that (1) randomly shuffled the temporal arrangement of all trials within a participant's task run a total of ten times, all the while maintaining the correspondence between class labels and trial patterns, and (2) conducted an eight-fold cross-validation strategy for each of these ten repeats (compare middle and bottom panels of Figure 1d) (Varoquaux et al., 2017). Data-splits were stratified to ensure the overall percentage of instances per class was maintained within each cross-validation fold. Performance metrics were computed at each iteration of the repeated stratified cross-validation scheme (10 repeats × 8 cross-validations) and consisted of a balanced accuracy score—a measure of overall performance, accounting for possible class imbalances—and a confusion matrix to provide an indication of prediction accuracy (sensitivity) per valence category. The performance metrics were averaged over iterations to yield a single accuracy score (or confusion matrix) per participant. The mean performance (balanced accuracy or confusion matrix) of valence classification across individuals was also calculated by averaging over all participants. Feature weights were extracted at each iteration and were averaged per class to produce three classifier weight maps for each participant (see also Section 2.12).

2.9 | Statistical testing

Permutation tests were conducted to assess the statistical significance of the subject-level accuracy scores. To this end, class labels were randomly shuffled across the trial patterns at each permutation (see outer/top panel of Figure 1d). Accuracy scores were computed at each permutation and pooled to provide a null-distribution against which the significance of each participant's true accuracy score could be determined; the α-level against which this true accuracy score was evaluated was adjusted for the number of participants—in our case equal to the number of conducted tests—via Bonferroni correction (i.e., $\alpha_{corrected} = .05 \ /112$). A total of 200 permutations were conducted per participant (Habes et al., 2013; see also Yuen et al., 2012). A grand null-distribution in which all 200 permutations of all 112 participants were pooled was also computed in order to assess the statistical significance of the mean accuracy score across all participants (again at an α-level of 5%).

2.10 | Sensitivity analyses

In order to assess the possible influence of class imbalance and response congruency on classification performance, the main analyses were repeated twice; once by setting the number of instances in each class equal to that of the smallest category, randomly sampling instances from the larger available subset for the non-smallest classes, and once by entering the patterns of *all* trial responses—both the congruent and incongruent ones—into the MVPA of each participant.

2.11 | Motion correction benchmarks

Spearman correlation coefficients were computed between the accuracy scores (amygdala ROI-based or whole-brain) and average FD across all individuals to quantify the extent to which in-scanner head motion was able to confound classification performance in spite of our motion correction and nuisance regression pipeline.

2.12 | Group-level feature weight maps

Exploratory group maps were computed by first normalizing the subject-level feature weight maps and then calculating the average weight across participants at each coordinate of standard MNI space. Normalization of the subject-level feature weight maps was performed by applying the deformation fields yielded earlier during the segmentation step of the pre-processing pipeline. No further statistical testing was performed on the group-level feature weight maps. These images were included as a rough indication of localisation and to provide a comparison to the overall patterns of activation obtained via standard univariate group-analysis of task reactivity (see Section 2.13).

2.13 | Univariate analysis

Univariate analyses of the task fMRI data were performed in order to produce group activation maps that could be compared to—and help to facilitate the interpretation of—the group-level feature weight maps yielded by the whole-brain MVPA. All univariate (group) analyses of the task fMRI data were conducted in SPM12 with default settings unless otherwise specified. In brief, the pre-processed functional scans were first normalized to standard MNI space by using the deformation fields created during segmentation. First-level GLM regression analyses were then conducted in order to estimate, for each participant separately, task reactivity on a voxel-by-voxel basis. For each participant, a GLM was specified that modelled the stimulus onsets and durations of the three valence categories (negative, neutral, positive) as separate predictors-of-interest (2 s boxcar) alongside the pre-processed BOLD signal as outcome variable. Only those trials for which the participant's response was congruent with the original IAPS rating were entered into these models. The white matter and CSF signals—obtained by calculating the average over all voxels in the corresponding segmentation masks, separately for each scan volume—as well as the six RPs, were additionally added to the GLMs as covariates-of-no-interest. A high-pass filter with a cut-off frequency of 1/128 Hz was applied to the data in the model in order to adjust for low-frequency scanner drift. As we used a one-versus-rest decision function for our main MVPAs (e.g., negative versus neutral and positive; see Section 2.8), it was important that we defined our univariate contrasts in such a way that the generated group activation (*t*-)maps could readily be compared to the group-level feature weight maps—the main output of the classification analyses. For each participant, contrast maps were therefore generated for the following first-level contrasts: (1) negative > positive + neutral, (2) positive > negative + neutral. Second-level analyses were then conducted for each of these contrasts (separately) via one-sample *t* tests. The threshold of significance for the group activation (*t*-) maps was adjusted for multiple comparisons by applying a familywise error rate (FWER) correcting threshold of $p < .05$ at the voxel-level.

3 | RESULTS

3.1 | Demographics

Demographic information of the 112 participants included in the final sample is presented in Table 1. About half of the participants ($n = 64$) had a diagnosis for one or more Axis I psychiatric disorders, the most frequently recorded of which was PTSD ($n = 47$), followed by mood disorders ($n = 32$), IED ($n = 20$), and anxiety disorders ($n = 18$). Twenty-two participants were being treated with one or more psychotropic drugs at the time of inclusion; the most frequently recorded drug classes were selective serotonin reuptake inhibitors (SSRI) ($n = 13$) and benzodiazepines ($n = 11$).

TABLE 1 Demographic and clinical information of the study sample.

Demographic	Participants (N = 112)
Age (median, IQR [years])	36 (13.5)
Gender (no. [m/f])	112/0
Handedness (no. [left/right/ambidextrous])[a]	7/90/14
Education level (no. [ISCED: 0/1/2/3/5])[a]	0/1/14/71/25
Psychiatric diagnosis (no.)[b]	
None	48
PTSD	47
IED[c]	20
Mood disorders	32
Schizophrenia and other psychotic disorders	0
Substance-related disorders	4
Anxiety disorders	18
Somatoform disorders	3
ADHD	0
Medication use (no.)[d]	
None	90
SSRI	13
Benzodiazepines	11
SARI	2
Antipsychotics	3
β-blockers	1
Ritalin	1
Melatonin	1

Abbreviations: ADHD, attention deficit hyperactivity disorder; IED, intermittent explosive disorder; IQR, interquartile range; ISCED, international scale for education; no, number of; PTSD, posttraumatic stress disorder; SARI, serotonin antagonist and reuptake inhibitors; SSRI, selective serotonin reuptake inhibitors.
[a]One participant had missing data for these demographics.
[b]The structured clinical interview for DSM IV disorders (SCID) was administered in the BETTER study, while the mini-international neuropsychiatric interview (MINI) was administered in the MARS study. Psychiatric co-morbidity is not taken into account by this table.
[c]IED was ascertained via research diagnostic criteria as published by Coccaro (2011). This diagnosis was only registered in the MARS study.
[d]Polypharmacy is not taken into account by this table.

3.2 | Behavioural data

The number of responses that were congruent with the original IAPS rating of the initial 96 picture trials had a mean and standard deviation of 79.81 ± 8.03 across all 112 participants (range: 55–94). The number of congruent responses had a median of 30 responses (interquartile range [IQR] = 3) for the negative stimulus class, 26.5 responses (IQR = 6) for the neutral class, and 26 responses (IQR = 7) for the positive valence category (range: 10–32 for all three classes). The number of congruent responses was found to differ significantly across the three stimulus classes (Friedman's analysis of variance [ANOVA]: $\chi^2(2) = 47.07$, $p < .001$), where negative responses were

given slightly but significantly more often than neutral (Wilcoxon signed-rank test: $z = -5.182$, $p < .001$) or positive ($z = -6.293$, $p < .001$) responses; no significant difference was found between the number of positive and neutral responses ($z = -0.612$, $p = .541$). This slight but significant difference in the number of responses per stimulus category could bias the performance of pattern classification towards the prediction of the largest, in our case, negative picture class—a point we will address further in the sections devoted to the main results of the MVPAs (see Section 3.4).

3.3 | Motion correction benchmarks

No significant correlation was recorded between the balanced accuracy scores and average FD data across all 112 participants for the amygdala ROI-based ($r = -.045$, $p = .641$) or whole-brain ($r = -.154$, $p = .105$) MVPAs. These results suggest that the combination of motion correction and nuisance regression was relatively successful in mitigating the possible confounding of in-scanner head motion.
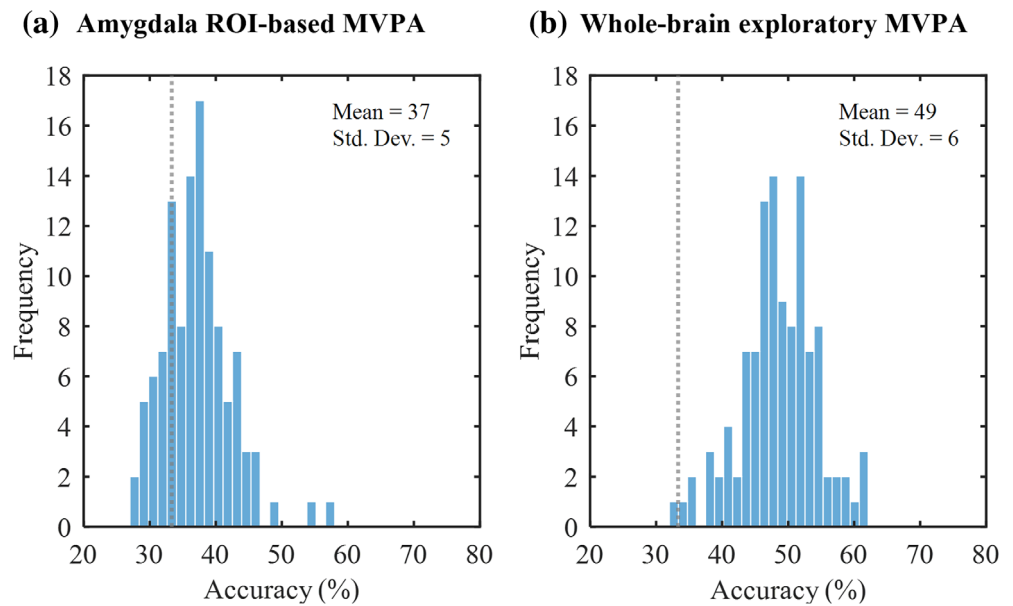
3.4 | Multi-voxel pattern analysis

3.4.1 | Amygdala ROI-based classification

Only 5 out of 112 participants (4.5%) had a classification accuracy that was statistically significant (Bonferonni-corrected) when only those voxels that reside within the amygdala were used to predict the valence categories (negative, neutral, positive) of emotional pictures from an individual's trial patterns. Spearman's rank correlation indicated no significant correlation between the number of voxels and amygdala-based classification accuracy across all 112 participants ($r = .007$, $p = .937$). The mean accuracy of valence classification across all participants was 37%—a score that lies only slightly above chance level (i.e., 33%)—and had a standard deviation of 5% (range: 28–58%; see Figure 2a). When tested against the permutation distribution across all participants, this mean accuracy of valence classification was found to be non-significant ($p = .211$). Independent samples t tests[1] showed that there were no significant group differences when the classification accuracies of participants with a diagnosis of PTSD ($t(93) = -0.205$, $p = .838$), IED ($t(25.08) = 0.343$, $p = .735$), a mood disorder ($t(78) = -0.367$, $p = .714$), or an anxiety disorder ($t(24.75) = -0.347$, $p = .732$) (see Supplementary Figure S2)—the four most frequent forms of psychopathology listed in Table 1—were compared to the accuracy scores of the subsample of participants who were medication- and diagnosis-free. Furthermore, there were no significant group differences when the classification accuracies of participants who were taking SSRI's ($t(14.09) = 0.781$, $p = .448$) or benzodiazepines ($t(57) = -0.682$, $p = .498$) were compared to the

---

[1]A regular Student's t test was performed when a non-significant Levene's test indicated that the assumption of equal variances likely held true, whereas Welch's t test was conducted in all other cases (i.e., in case Levene's test was shown to be significant).

FIGURE 2 Frequency distributions of the balanced accuracy scores of the ROI-based (a) and whole-brain (b) participant-level classifications. The mean accuracy and standard deviation of valence classification across all participants ($N = 112$) are denoted in the upper right corner of the histograms in panels (a) and (b). The dotted vertical grey line in each panel represents chance level accuracy (33%). MVPA, multi-voxel pattern analysis; ROI, region-of-interest; Std. Dev, standard deviation.

(a) Amygdala ROI-based MVPA
Mean = 37
Std. Dev. = 5

(b) Whole-brain exploratory MVPA
Mean = 49
Std. Dev. = 6

accuracy scores of medication- and diagnosis-free participants (see Supplementary Figure S2).

The average accuracy of valence classification for each of the three stimulus categories (negative, neutral, positive) across all participants was examined via a confusion matrix (see Figure 3a). In terms of sensitivity, only the negative valence category (46%, $p = .008$) was found to be significantly decodable after Bonferroni correction ($\alpha_{Bonferroni} = .017$), whereas the neutral (34%, $p = .417$) and positive (31%, $p = .651$) picture classes were not. In order to evaluate whether the classification accuracies were affected by the slight but significant class imbalance (see Section 3.2; see also Section 2.10), we repeated the above analysis; but—for each participant—set the number of instances in each class equal to that of the smallest category, randomly sampling instances from the larger available subset for the non-smallest classes (median number of responses = 23, IQR = 6.75, range: 10–30). As a result, the sensitivity value of the negative picture category was slightly lower (i.e., 43%) and rendered non-significant after adjusting for multiple comparisons ($p = .031$); otherwise, the main results remained largely unchanged (see Supplementary Figure S3). Also, in order to evaluate the impact of response congruency on classification performance (see Section 2.10), we again repeated the above analysis, but now included the patterns of *all* trial responses—both the congruent and incongruent ones—to the MVPA of each participant; the sensitivity value of the negative picture class was again slightly lower (i.e., 42%) and rendered non-significant after adjusting for multiple comparisons ($p = .046$), but otherwise, the main results remained largely unchanged (see Supplementary Figure S4).

3.4.2 | Whole-brain classification

Slightly more than half of the participants (57 out of 112, or 50.9%) had a classification accuracy that was statistically significant (Bonferroni-corrected) when those voxels within a person's grey matter mask that survived feature selection were used to predict the valence categories (negative, neutral, positive) of emotional pictures from an individual's trial patterns. The mean accuracy of valence classification across all participants was 49%—a score that is considerably higher than what would be expected based on chance (i.e., 33%)—and had a standard deviation of 6% (range: 33–62%; see Figure 2b). When tested against the permutation distribution across all participants, the mean accuracy of valence classification was found to be statistically significant ($p = .004$). Independent samples t tests again showed that there were no significant group differences when the classification accuracies of participants with a diagnosis of PTSD ($t(93) = -1.259$, $p = .211$), IED ($t(60.74) = 1.147$, $p = .256$), a mood disorder ($t(78) = -1.402$, $p = .165$), or an anxiety disorder ($t(54.96) = -0.791$, $p = .432$) (see Supplementary Figure S2) were compared to the accuracy scores of the participants who were medication- and diagnosis-free. Again, there were also no significant group differences when the classification accuracies of participants who were taking SSRI's ($t(52.91) = -1.148$, $p = .256$) or benzodiazepines ($t(57) = -1.376$, $p = .174$) were compared to the accuracy scores of medication- and diagnosis-free participants (see Supplementary Figure S2).

The average accuracy of valence classification for each of the three stimulus categories (negative, neutral, positive) across all participants was again examined via a confusion matrix (see Figure 3b). In terms of sensitivity, both the negative (57%, $p < .001$) and neutral (48%, $p = .005$) picture classes were found to be significantly decodable after Bonferroni correction ($\alpha_{Bonferroni} = .017$), whereas the positive (41%, $p = .081$) valence category was not. In order to evaluate whether the classification accuracies were affected by the slight but significant class imbalance (see Section 3.2; see also Section 2.10), we repeated the above analysis; but—for each participant—set the number of instances in each class equal to that of the smallest category, randomly sampling instances from the larger available subset for the non-smallest classes (median number of responses = 23, IQR = 6.75,
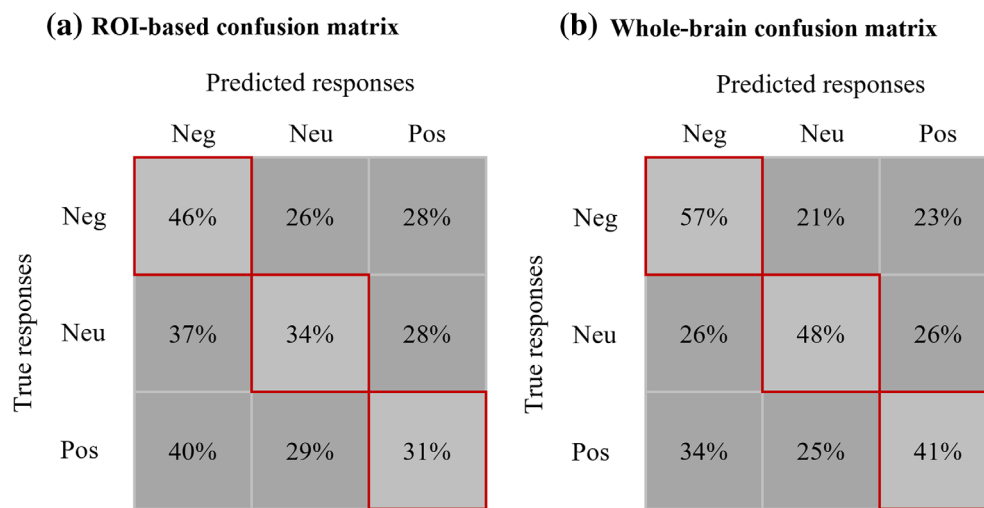
(a) ROI-based confusion matrix

Predicted responses

(b) Whole-brain confusion matrix

Predicted responses

FIGURE 3 Group-level confusion matrices displaying the true (rows) versus predicted (columns) stimulus categorisations of the ROI-based (a, left) and whole-brain (b, right) MVPAs. All cells are normalized by rows, such that each element shows the percentage relative to the total number of true responses in that valence category (negative, neutral, or positive). The diagonal elements are framed in red in order to highlight the true positive (i.e., sensitivity) rates for each of the three stimulus categories. MVPA, multi-voxel pattern analysis; Neg, negative; Neu, neutral; Pos, positive.; ROI, region-of-interest

range: 10–30). The sensitivity value of the positive picture category was slightly raised (i.e., 44%) but still did not reach significance after adjusting for multiple comparisons ($p = .038$); the rest of the main results also remained largely unchanged (see Supplementary Figure S3). Furthermore, in order to evaluate the impact of response congruency on classification performance (see Section 2.10), we again repeated the above analysis, but now included the patterns of *all* trial responses—both the congruent and incongruent ones—to the MVPA of each participant; again, the main results remained largely unchanged (see Supplementary Figure S4).

### 3.5 | Group-level feature weight maps

Group-level feature weight maps of each of the three valence categories are presented in Figure 4. For the negative picture class, comparatively high positive weights were observed in the middle temporal gyrus (bilateral), precuneus, medial prefrontal cortex, left precentral gyrus, and—perhaps most notably—within the amygdala (bilateral), whereas relatively high negative weights were observed in bilateral regions of the lateral prefrontal cortex and lateral parietal cortex (bilateral) (see Figure 4a). For the positive picture class, a similar pattern of feature weights was observed but—in many cases—with the plus and minus signs reversed, such that comparatively high *negative* weights were observed in the middle temporal gyrus (bilateral), precuneus, and medial prefrontal cortex, whereas relatively high *positive* weights were observed in bilateral regions of the lateral prefrontal cortex and lateral parietal cortex (bilateral) (see Figure 4b). A notable exception to this reversal was the amygdala—inside which the magnitude of feature weights for the positive (vs. the negative) picture class seemed to be markedly lower regardless of (plus or minus) sign. The group-level feature weight map of the neutral picture class appeared

to be a combination of the patterns observed for the negative and positive picture classes—albeit with markedly lower feature weight values in an absolute sense (see Figure 4c). Setting the number of instances per class equal to that of the smallest category did not overtly change the patterns observed in the group-level feature weight maps for any of the three picture categories (see Supplementary Figure S5). Also, repeating the analyses with the trial patterns of *all* responses—both the congruent and incongruent ones—as input for the MVPAs did not overtly change the patterns observed in the group-level feature weight maps for any of the three picture categories (see Supplementary Figure S6).

3.6 | Univariate analyses

The group activation maps yielded by the standard univariate (one sample) analyses of the task fMRI data are presented in Figure 5 (see also Supplementary Figure S7). The negative > positive + neutral contrast indicated strong and statistically significant activation in the middle temporal gyrus (bilateral), as well as—notably—within the (bilateral) amygdala, along with weaker, but nevertheless significant activation in midline brain structures such as the medial prefrontal cortex and posterior cingulate cortex/precuneus; regions of significant deactivation included the lateral parietal cortex and regions of the lateral prefrontal cortex, in both cases, mostly lateralized to the right hemisphere. Note that these are essentially the same brain regions for which high positive and negative feature weight values were observed earlier (see Figure 4a). For the positive > negative + neutral contrast, strong and statistically significant activation was recorded in the medial prefrontal cortex and posterior cingulate cortex/precuneus, the latter cluster extending towards the (bilateral) calcarine fissure and lingual gyrus, mirroring

FIGURE 4 Raw (unthresholded) group-level feature weight maps for the negative (a), positive (b), and neutral (c) picture classes. In each panel, sagittal, coronal, and horizontal slices are presented in the top, middle, and bottom rows, respectively, with the corresponding slice-coordinates indicated below each subplot. The column on the far right of each panel provides a visual representation of the locations of each of the slices presented from right to left in case of the upper rows (sagittal plane), from posterior to anterior in case of the middle rows (coronal plane), or from inferior to superior in case of the bottom rows (horizontal plane). Warm colours represent positive weights and cool colours represent negative weights. Note that as a one-versus-rest (ovr) decision function was employed for the participant-level MVPAs, each stimulus class was contrasted against the other two remaining classes (e.g., negative versus neutral and positive). The results are overlaid on an average brain created from high-resolution normalizations of the T1-weighted anatomical images of all 112 included participants. MVPA, multi-voxel pattern analysis.
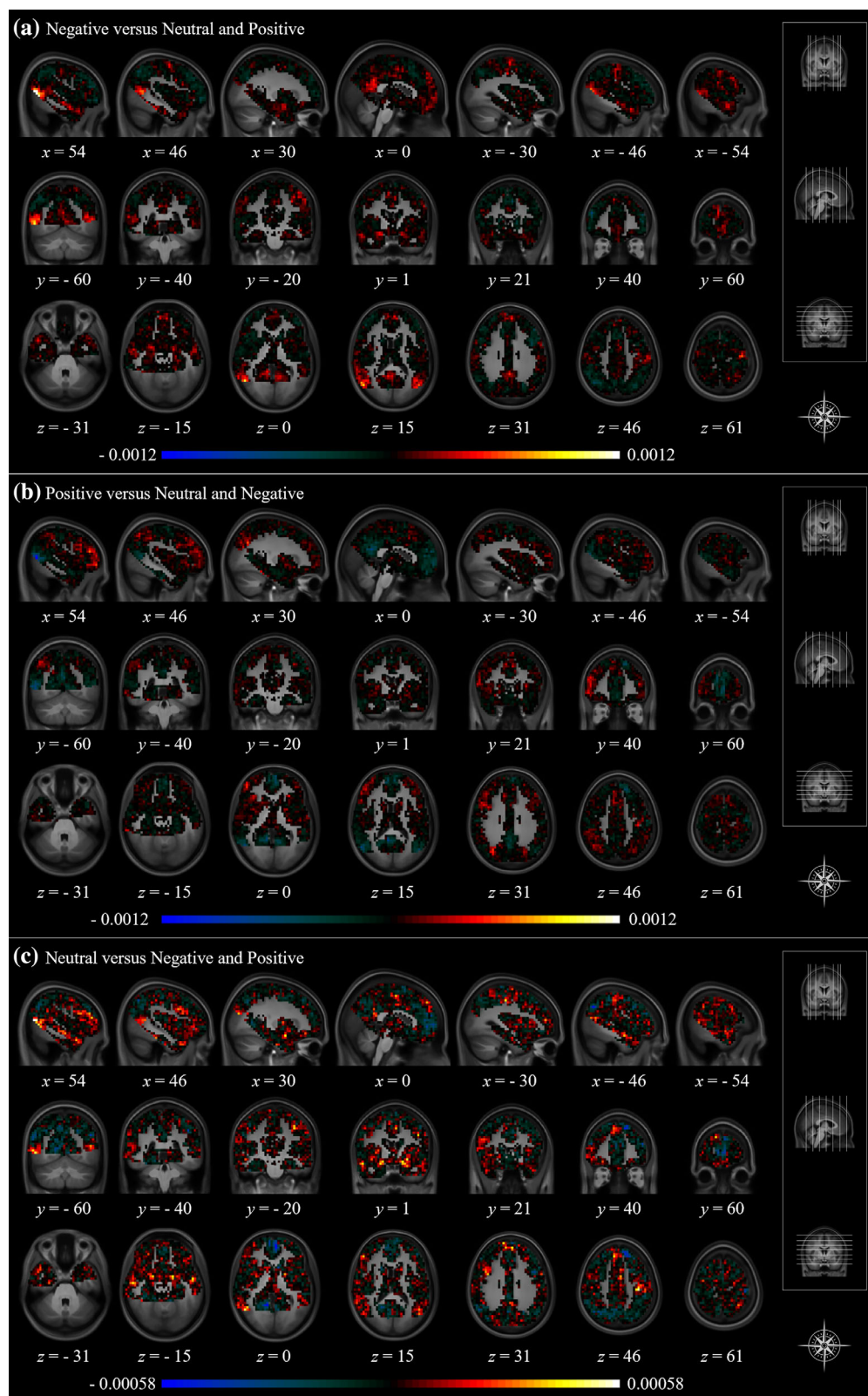
the pattern of high negative feature weight values visible at this location in Figure 4b. Another small region of significant activation was recorded in the dorsolateral prefrontal cortex. No regions of significant deactivation were recorded for the positive > negative + neutral contrast.

No significant group differences were recorded for either the negative > positive + neutral or positive > negative + neutral contrasts when participants with a diagnosis of PTSD, IED, a mood disorder, or an anxiety disorder were (separately) compared to the subset of medication- and diagnosis-free participants. Similarly, there were
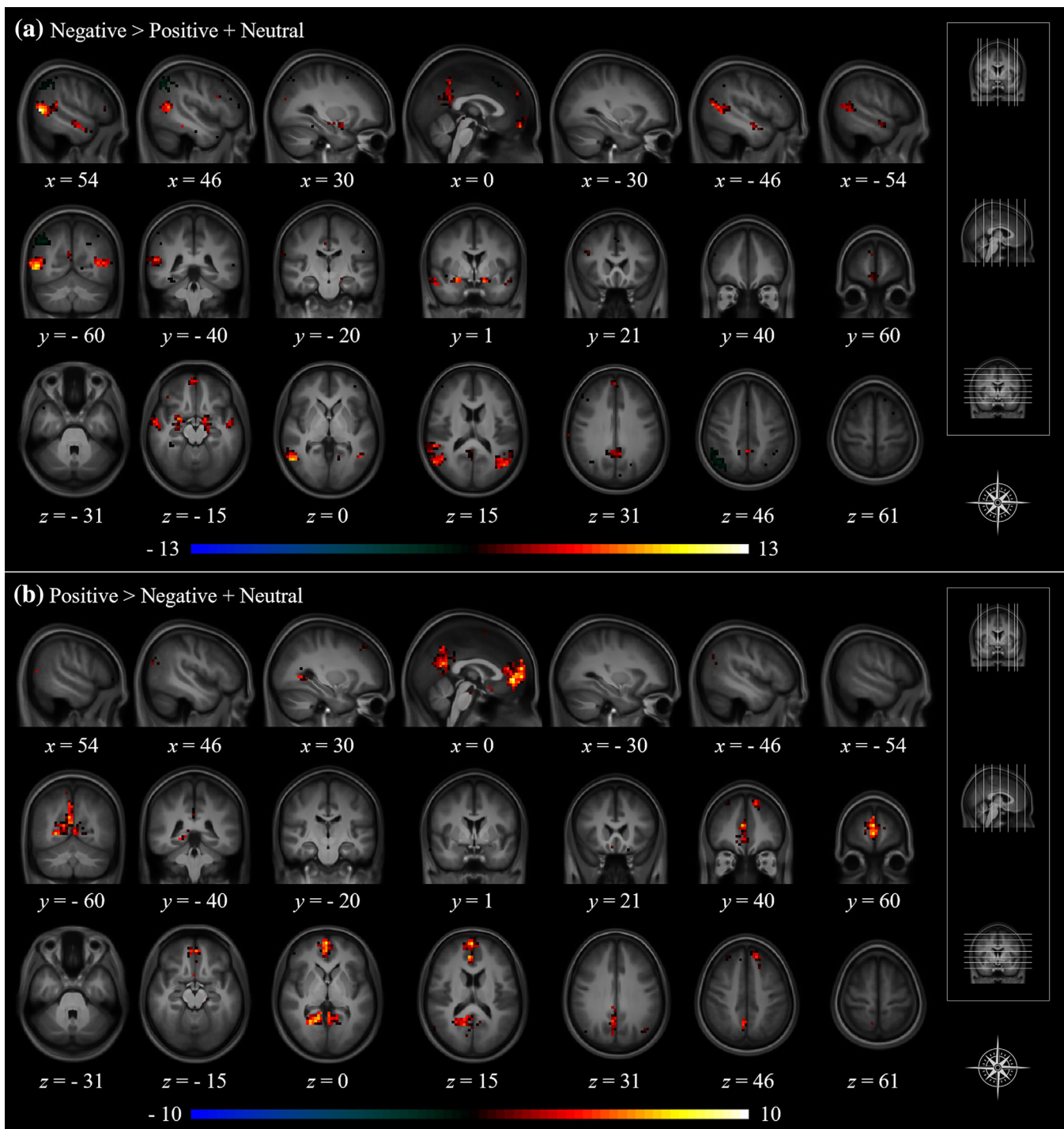
FIGURE 5 Group activation (*t*-)maps for the (a) negative > positive + neutral and (b) positive > negative + neutral contrasts. In each panel, sagittal, coronal, and horizontal slices are presented in the top, middle, and bottom rows, respectively, with the corresponding slice-coordinates indicated below each subplot. The column on the far right of each panel provides a visual representation of the locations of each of the slices presented from right to left in case of the upper rows (sagittal plane), from posterior to anterior in case of the middle rows (coronal plane), or from inferior to superior in case of the bottom rows (horizontal plane). Warm colours represent (relative) activation and cool colours deactivation. The results are overlaid on an average brain created from high-resolution normalizations of the T1-weighted anatomical images of all 112 included participants. A voxel-wise FWER-corrected threshold of *t* > 4.76 was applied to each group activation map (corresponding to *p* < .05). FWER = familywise error-rate

no significant group differences for the two contrasts when participants who were taking SSRI's or benzodiazepines were compared to medication- and diagnosis-free participants.[2]

---

[2]Each of these six comparisons was conducted via two-sample *t* tests at the second-level. A FWER-correction threshold of *p* < .05 was applied at the voxel-level, as with the one-sample univariate analyses.

4 | DISCUSSION

The main goal of this study was to determine whether the patterns of BOLD reactivity inside the amygdala can be used to predict the valence categories (neutral, negative, positive) of emotional stimuli during a picture presentation task. To this end, we combined the data from two prior studies (Heesink et al., 2018; Van Rooij et al., 2014), and—for each participant ($N = 112$)—tried to decode the valence classes using the BOLD responses within the amygdala (hypothesis-driven analyses) and the whole-brain (exploratory analyses). Contrary to our expectations, classification based on the amygdala alone did not support an accurate prediction of emotional valence, as we were able to decode valence classes significantly only in a handful of participants (less than 5%), with the distribution of accuracy scores centring (mean = 37%) only a few percent above chance-level (33%). On average, pattern classification based on whole-brain task reactivity resulted in roughly half of the participants showing a significant accuracy score (50.9%), with an increase in mean prediction accuracy of about 10% (49%), relative to the aforementioned across-subjects mean of the amygdala-based MVPAs. Results indicated no significant group effects of psychiatric diagnosis (PTSD, IED, mood disorder, anxiety disorder) and/or medication-status (SSRI or benzodiazepines).

The low accuracy of amygdala-based valence classification is consistent with an earlier report by Saarimäki et al. (2016). In that study, a linear neural network classifier was trained to decode five emotional states (disgust, fear, happiness, sadness, or neutral) in 21 healthy volunteers by using the task reactivity patterns inside the amygdala. Similar to what we observed here, the mean accuracy across participants was found to be relatively low although statistically significant (i.e., approximately 30%, where 20% represented chance-level and 23% the threshold of significance)—a result we could not replicate here. One factor that may have contributed to this discrepancy in findings includes the difference in sample size and population between the study of Saarimäki et al. (2016) ($N = 21$ volunteers) and that of ours ($N = 112$ veterans). Other contributing factors include the nature and number of the emotional categories being classified; that is, the five basic emotions, disgust, fear, happiness, sadness, and neutral in the task of Saarimäki et al. (2016), versus the broad strokes of the three-class (negative, neutral, positive) emotional valence model employed here. It is possible that the performance of our classifier would have been higher if a similar, more nuanced categorisation of affect had been used to train the algorithm. Still another factor that may have contributed to the discrepancy in findings is the difference in algorithms used for the classification itself; that is, an SVM in the current study versus a linear neural network classifier in the study by Saarimäki et al. (2016). Finally, we draw attention to the task stimuli and design of Saarimäki et al. (2016)—namely a series of short and naturalistic (although unstandardized) video fragments—as these were markedly different from the set of well-validated and standardized (but possibly less ecologically valid) IAPS photographs employed here.

The higher accuracy of whole-brain versus amygdala-based valence classification seems to be consistent with a number of earlier reports. In the above-cited work by Saarimäki et al. (2016), for example, the mean accuracy of whole-brain valence classification across all 21 participants was 47% (chance-level: 20%; significance threshold: 23%). This score exceeded the mean accuracy of amygdala-based valence classification by more than 17%—a somewhat larger increment than the 11% increase in classification accuracy we reported here. Similar to our findings (see Figure 4), the group maps by Saarimäki et al. (2016) also highlighted the influence of midline brain structures, the middle temporal gyrus, and—most notably—the amygdala. In the only other work that also utilized a fast event-related design, Bush et al. (2018) trained a linear SVM to decode the valence categories (negative, positive) of emotional pictures using whole-brain BOLD reactivity patterns, and recorded a mean accuracy of 85% across all participants, with all 19 participants demonstrating a significant accuracy score at single-subject level. The group maps by Bush et al. (2018) again highlighted the importance of midline brain structures along with the amygdala. In another study, Habes et al. (2013) used a linear SVM to classify the valence categories of emotional pictures based on whole-brain patterns of BOLD activation in nine patients suffering from major depression. Statistically significant mean accuracy scores of 86, 89, and 92% were observed when negative and neutral, positive and neutral, and negative and positive trials were contrasted to one another, respectively; all contrasts yielded group maps that again highlighted the influence of the amygdala. Other relevant works confirm the relatively high intra-subject accuracy of whole-brain pattern classification of emotional valence, with predication accuracies ranging between 60 and 92% (Baucom et al., 2012; Yuen et al., 2012). Taken together, these past and present findings seem to converge to indicate that task effects of emotion provocation fMRI can indeed be captured at the level of the individual participant, but that it requires the information contained by spatially distributed patterns of brain activation, rather than the reactivity within the amygdala alone. This notion resonates closely with the idea that complex brain functions can best be understood as emergent properties of the large-scale interconnected nature of the (human) brain—the core premise of the field of network neuroscience (e.g., see Bassett & Sporns, 2017 for an overview). Indeed, the two constellations of brain regions we recorded here—both of the middle temporal gyri, the precuneus, and the medial prefrontal cortex on the one hand, versus the lateral prefrontal and parietal cortices on the other hand—bare a remarkable degree of similarity to perhaps the two most well-known large-scale brain networks, namely the default mode network (DMN: medial prefrontal cortex, posterior cingulate cortex/precuneus, and lateral temporal cortex/inferior parietal cortex) and the central executive network (CEN: lateral parietal cortex and dorsolateral prefrontal cortex), respectively (Fox et al., 2005; Raichle, 2015).

The strength of our work lies in its relatively large sample size ($N = 112$) and the clinical diversity of the study sample (see Table 1). Still, our work was subject to a number of limitations: First, the results of our work may not be fully generalizable to a non-military and/or female population, as only male veterans were included in our sample. Second, the low resolution of our 3 T fMRI data might have limited decoding performance to a certain degree; it is possible that higher accuracy scores would have been obtained at a higher resolution

and/or magnetic field-strength—particularly when considering small brain regions such as the amygdala (Sladky et al., 2013). Similarly, our decoding performance might have been undermined by the relatively low number of available trials per picture category. Third, although our confound regression strategy was relatively successful in mitigating the confounding effect of in-scanner head motion on prediction accuracy, it is possible that the stringency of the model may have led to some inadvertent filtering out of emotional information as well. Fourth, our in-scanner task utilized a hybrid block/fast event-related design. It should be noted, however, that the LS-S technique by Mumford et al. (2012) was specifically designed for feature extraction in simple fast event-related designs, not hybrid block/fast event-related designs. Nonetheless, given that prior work by Valente et al. (2019) has indicated that the LS-S method performs equally well when extracting individual trial patterns within a single block (i.e., block-wise LS-S), versus considering all trials across the entire scan run (i.e., run-wise LS-S)—which is what we did here—we do not expect our divergent task design to have had much of an impact on the main findings. Fifth, the fast-event related nature of our emotional processing task may have limited our ability to fully capture the information unique to each and every trial in a participant's task run. Although the LS-S technique of Mumford et al. (2012) was indeed specifically designed to help deal with this issue, the temporally correlated nature of our in-scanner task—that is to say, its short inter-trial interval—may nonetheless have prevented us to fully isolate the signals of adjacent trials in a participant's time-series—in spite of our use of the method.

4.1 | Conclusions

In conclusion, our findings indicate that pattern classification based on the amygdala alone is insufficient to provide an accurate prediction of the valence categories (negative, neutral, positive) of emotional stimuli at the single-subject level. This outcome may be of particular relevance to studies that seek to target the amygdala in the treatment of emotion regulation problems, for example via real-time fMRI neurofeedback training. Classification of whole-brain BOLD reactivity led to accuracy scores that were considerably higher—and more often statistically significant—than the amygdala-based MVPAs, suggesting that the encoded emotional information was contained by spatially distributed patterns of brain reactivity, rather than being confined to the amygdala volume. In line with this latter notion, exploratory group maps pointed towards a set of brain regions commonly associated with either the DMN or central executive network as playing likely important—yet seemingly opposing—parts in decoding the valence of the emotional pictures.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID
Tim Varkevisser https://orcid.org/0000-0003-3441-9486
Elbert Geuze https://orcid.org/0000-0003-3479-2379
Remko van Lutterveld https://orcid.org/0000-0003-1844-0584

REFERENCES

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., Habel, U., Schneider, F., & Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anatomy and Embryology*, *210*(5), 343–352. https://doi.org/10.1007/s00429-005-0025-5

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, *145*, 137–145. https://doi.org/10.1016/j.neuroimage.2016.02.079

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Baucom, L. B., Wedell, D. H., Wang, J., Blitzer, D. N., & Shinkareva, S. V. (2012). Decoding the neural representation of affective states. *NeuroImage*, *59*(1), 718–727. https://doi.org/10.1016/j.neuroimage.2011.07.037

Bryant, R. A., Kemp, A. H., Felmingham, K. L., Liddell, B., Olivieri, G., Peduto, A., Gordon, E., & Williams, L. M. (2008). Enhanced amygdala and medial prefrontal activation during nonconscious processing of fear in posttraumatic stress disorder: An fMRI study. *Human Brain Mapping*, *29*(5), 517–523. https://doi.org/10.1002/hbm.20415

Bush, K. A., Gardner, J., Privratsky, A., Chung, M. H., James, G. A., & Kilts, C. D. (2018). Brain states that encode perceived emotion are reproducible but their classification accuracy is stimulus-dependent. *Frontiers in Human Neuroscience*, *12*, 1–15. https://doi.org/10.3389/fnhum.2018.00262

Chai, X. J., Castañán, A. N., Öngür, D., & Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *NeuroImage*, *59*(2), 1420–1428. https://doi.org/10.1016/j.neuroimage.2011.08.048

Coccaro, E. F. (2011). Intermittent explosive disorder: Development of integrated research criteria for diagnostic and statistical manual of mental disorders, fifth edition. *Comprehensive Psychiatry*, *52*(2), 119–125. https://doi.org/10.1016/j.comppsych.2010.05.006

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biological Psychiatry*, *62*, 168–178. https://doi.org/10.1016/j.biopsych.2006.08.024

Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews*, *58*(1), 57–70. https://doi.org/10.1016/j.brainresrev.2007.10.012

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58. https://doi.org/10.1016/j.neuroimage.2008.06.037

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data.

NeuroImage, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, *31*(7), 792–806. https://doi.org/10.1177/0956797620916786

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355. https://doi.org/10.1002/mrm.1910350312

Habes, I., Krall, S. C., Johnston, S. J., Yuen, K. S. L., Healy, D., Goebel, R., Sorger, B., & Linden, D. E. J. (2013). Pattern classification of valence in depression. *NeuroImage: Clinical*, *2*, 675–683. https://doi.org/10.1016/j.nicl.2013.05.001

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *Neuroreport*, *11*(1), 43–48. https://doi.org/10.1097/00001756-200001170-00009

Heesink, L., Gladwin, T. E., Vink, M., van Honk, J., Kleber, R., & Geuze, E. (2018). Neural activity during the viewing of emotional pictures in veterans with pathological anger and aggression. *European Psychiatry*, *47*, 1–8. https://doi.org/10.1016/j.eurpsy.2017.09.002

Infantolino, Z. P., Luking, K. R., Sauder, C. L., Curtin, J. J., & Hajcak, G. (2018). Robust is not necessarily reliable: From within-subjects fMRI contrasts to between-subjects comparisons. *NeuroImage*, *173*, 146–152. https://doi.org/10.1016/j.neuroimage.2018.02.024.Robust

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Lane, R. D., & Nadel, L. (2002). *Cognitive neuroscience of emotion*. Oxford University Press.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39–58.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, *23*, 727–738. https://doi.org/10.1023/A:1025048802629

Linhartová, P., Látalová, A., Kóša, B., Kašpárek, T., Schmahl, C., & Paret, C. (2019). fMRI neurofeedback in emotion regulation: A literature review. *NeuroImage*, *193*, 75–92. https://doi.org/10.1016/j.neuroimage.2019.03.011

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel pattern analysis for fMRI data: A review. *Computational and Mathematical Methods in Medicine*, *2012*, 1–14. https://doi.org/10.1155/2012/961257

McCloskey, M. S., Phan, K. L., Angstadt, M., Fettich, K. C., Keedy, S., & Coccaro, E. F. (2016). Amygdala hyperactivation to angry faces in intermittent explosive disorder. *Journal of Psychiatric Research*, *79*, 34–41. https://doi.org/10.1016/j.jpsychires.2016.04.006

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, *53*(1), 103–118. https://doi.org/10.1016/j.neuroimage.2010.05.051

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

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences*, *11*(11), 489–497. https://doi.org/10.1016/j.tics.2007.08.013

Nicholson, A. A., Rabellino, D., Densmore, M., Frewen, P. A., Paret, C., Kluetsch, R., Schmahl, C., Théberge, J., Neufeld, R. W. J., McKinnon, M. C., Reiss, J., Jetly, R., & Lanius, R. A. (2017). The neurobiology of emotion regulation in posttraumatic stress disorder: Amygdala downregulation via real-time fMRI neurofeedback. *Human Brain Mapping*, *38*(1), 541–560. https://doi.org/10.1002/hbm.23402

Nord, C. L., Gray, A., Charpentier, C. J., Robinson, O. J., & Roiser, J. P. (2017). Unreliability of putative fMRI biomarkers during emotional face processing. *NeuroImage*, *156*, 119–127. https://doi.org/10.1016/j.neuroimage.2017.05.024

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

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

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*(2), 331–348. https://doi.org/10.1006/nimg.2002.1087

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/j.neuron.2005.09.025

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, *38*, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

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

Shin, L. M., Rauch, S. L., & Pitman, R. K. (2006). Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Annals of the New York Academy of Sciences*, *1071*, 67–79. https://doi.org/10.1196/annals.1364.007

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., & Petersen, S. E. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human Brain Mapping*, *35*(5), 1981–1996. https://doi.org/10.1002/hbm.22307

Sladky, R., Baldinger, P., Kranz, G. S., Tröstl, J., Höflich, A., Lanzenberger, R., Moser, E., & Windischberger, C. (2013). High-resolution functional MRI of the human amygdala at 7 T. *European Journal of Radiology*, *82*(5), 728–733. https://doi.org/10.1016/j.ejrad.2011.09.025

Valente, G., Kaas, A. L., Formisano, E., & Goebel, R. (2019). Optimizing fMRI experimental design for MVPA-based BCI control: Combining the strengths of block and event-related designs. *NeuroImage*, *186*, 369–381. https://doi.org/10.1016/j.neuroimage.2018.10.080

Van Buuren, M., Vink, M., Rapcencu, A. E., & Kahn, R. S. (2011). Exaggerated brain activation during emotion processing in unaffected siblings of patients with schizophrenia. *Biological Psychiatry*, *70*(1), 81–87. https://doi.org/10.1016/j.biopsych.2011.03.011

Van den Boom, M. A., Vansteensel, M. J., Koppeschaar, M. I., Raemaekers, M. A. H., & Ramsey, N. F. (2019). Towards an intuitive communication-BCI: Decoding visually imagined characters from the early visual cortex using high-field fMRI. *Biomedical Physics & Engineering Express*, *5*(5), 55001. https://doi.org/10.1088/2057-1976/ab302c

Van Rooij, S. J. H., Kennis, M., Vink, M., & Geuze, E. (2015). Predicting treatment outcome in PTSD: A longitudinal functional MRI study on

trauma-unrelated emotional processing. *Neuropsychopharmacology*, 41, 1156–1165. https://doi.org/10.1038/npp.2015.257

Van Rooij, S. J. H., Rademaker, A. R., Kennis, M., Vink, M., Kahn, R. S., & Geuze, E. (2014). Neural correlates of trauma-unrelated emotional processing in war veterans with PTSD. *Psychological Medicine*, 45, 575–587. https://doi.org/10.1017/S0033291714001706

Van Rossum, G., & Drake, F. L., Jr. (1995). *Python tutorial*. Centrum voor Wiskunde en Informatica (CWI). https://doi.org/10.1017/9781108653947.010

Varkevisser, T., Gladwin, T. E., Heesink, L., van Honk, J., & Geuze, E. (2017). Resting-state functional connectivity in combat veterans suffering from impulsive aggression. *Social Cognitive and Affective Neuroscience*, 12(12), 1881–1889. https://doi.org/10.1093/scan/nsx113

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145, 166–179. https://doi.org/10.1016/j.neuroimage.2016.10.038

Vink, M., Derks, J. M., Hoogendam, J. M., Hillegers, M., & Kahn, R. S. (2014). Functional differences in emotion processing during adolescence and early adulthood. *NeuroImage*, 91, 70–76. https://doi.org/10.1016/j.neuroimage.2014.01.035

Wottschel, V., Chard, D. T., Enzinger, C., Filippi, M., Frederiksen, J. L., Gasperini, C., Giorgio, A., Rocca, M. A., Rovira, A., De Stefano, N., Tintoré, M., Alexander, D. C., Barkhof, F., & Ciccarelli, O. (2019). SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes

suggestive of multiple sclerosis. *NeuroImage: Clinical*, 24, 102011. https://doi.org/10.1016/j.nicl.2019.102011

Young, K. D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W. C., & Bodurka, J. (2014). Real-time fMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PLoS One*, 9(2), e88785. https://doi.org/10.1371/journal.pone.0088785

Yuen, K. S. L., Johnston, S. J., De Martino, F., Sorger, B., Formisano, E., Linden, D. J., & Goebel, R. (2012). Pattern classification predicts individuals' responses to affective stimuli. *Translational Neuroscience*, 3(3), 278–287. https://doi.org/10.2478/s13380-012-0029-6

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Varkevisser, T., Geuze, E., van den Boom, M. A., Kouwer, K., van Honk, J., & van Lutterveld, R. (2023). Pattern classification based on the amygdala does not predict an individual's response to emotional stimuli. *Human Brain Mapping*, 44(12), 4452–4466. https://doi.org/10.1002/hbm.26391