NeuroImage 211 (2020) 116585

Contents lists available at ScienceDirect

NeuroImage

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

Reproducibility of amygdala activation in facial emotion processing at 7T

Nicole Geissberger [1], Martin Tik [1], Ronald Sladky, Michael Woletz, Anna-Lisa Schuler, David Willinger, Christian Windischberger [*]

MR Center of Excellence, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria

ARTICLE INFO

Keywords:
Emotion processing
Amygdala
Quality assessment
Test-retest
Stability
Habituation

ABSTRACT

Despite its importance as the prime method for non-invasive assessment of human brain function, functional MRI (fMRI) was repeatedly challenged with regards to the validity of the fMRI-derived brain activation maps. Amygdala fMRI was particularly targeted, as the amygdala's anatomical position in the ventral brain combined with strong magnetic field inhomogeneities and proximity to large vessels pose considerable obstacles for robust activation mapping.

In this high-resolution study performed at ultra-high field (7T) fMRI, we aimed at (1) investigating systematic replicability of amygdala group-level activation in response to an established emotion processing task by varying task instruction and acquisition parameters and (2) testing for intra- and intersession reliability.

At group-level, our results show statistically significant activation in bilateral amygdala and fusiform gyrus for each of the runs acquired. In addition, while fusiform gyrus activations are consistent across runs and sessions, amygdala activation levels show habituation effects across runs. This amygdala habituation effect is replicated in a session repeated two weeks later. Varying task instruction between matching emotions and matching persons does not change amygdala activation strength. Also, comparing two acquisition protocols with repetition times of either 700 ms or 1400 ms did not result in statistically significant differences of activation levels.

Regarding within-subject reliability of amygdala activation, despite considerable variance in individual habituation patterns, we report fair to good inter-session reliability for the first run and excellent reliability for averages over runs.

We conclude that high-resolution fMRI at 7T allows for robust mapping of amygdala activation in a broad range of variations. Our results of amygdala 7T fMRI are suitable to inform methodology and may encourage future studies to continue using emotion discrimination paradigms in clinical and non-clinical applications.

1. Introduction

Despite its tremendous success in revealing details on brain functioning, functional MRI has faced several challenges regarding validity in general and replicability in particular, see Nat Neurosci 20 (2017). Most prominent among them are the possible use of invalid statistical thresholds yielding false positive study outcomes (Eklund et al., 2016) and a general publication bias, referring to a preference for reporting only statistically significant results, often in combination with low statistical power that are potentially overestimating the effect size (Button et al., 2013).

These circumstances not only question the results of cognitive fMRI experiments, they are also of particular concern when fMRI experiments are used for investigating the efficacy of drug treatment schemes in psychiatry (Cremers et al., 2016; Duff et al., 2015; McGonigle et al., 2017; Nathan et al., 2014; Sheline et al., 2001; Sladky et al., 2015; Szczepanik et al., 2016), psychotherapy (Huang et al., 2014; Marini et al., 2015; Siegle et al., 2006; Straub et al., 2015) or are investigated as therapeutic themselves like in neuro-feedback fMRI studies (Young et al., 2017). Some fMRI activation patterns are even considered as biomarkers (e.g. Swartz et al. (2015)). Misinterpreting fMRI study results could hinder detecting the real mechanism of action or identification of potential biomarkers leaving treatment chances for patients neglected.

Therefore studies that systematically vary conditions in order to investigate the exact boundaries of an fMRI task and its expected neural response are strongly needed.

According to Nichols et al. (2017), the terms reproducibility, replicability and reliability are not used with consistent definitions in the

* Corresponding author.
E-mail address: christian.windischberger@meduniwien.ac.at (C. Windischberger).
[1] Authors contributed equally to the manuscript.

https://doi.org/10.1016/j.neuroimage.2020.116585
Received 14 November 2018; Received in revised form 24 November 2019; Accepted 23 January 2020
Available online 26 January 2020

community. In order to avoid confusion and misunderstandings in this regard, we use reproducibility as an umbrella term and would like to determine further terminology in this manuscript: When varying parameters (i.e. task instructions or acquisition protocol) and comparing group results, we are using the term replicability, in the sense of systematic replication as proposed by Schmidt (2009). Reliability is referring to single-subject data compared over repeated sessions and its degree of concordance.

The amygdala has recently been in the focus of critique after amygdalar activation patterns have even been attributed to vascular artefacts (Boubela et al., 2015). FMRI studies in health and disease concerning consistency of amygdala activations using fMRI led to heterogeneous results ranging between low and excellent test-retest reliability (Alcauter et al., 2014; Fournier et al., 2014; Johnstone et al., 2005; Manuck et al., 2007; Nord et al., 2017; Plichta et al., 2012; Schacher et al., 2006). A large number of fMRI studies provided evidence for clear amygdalar involvement during emotional tasks (Breiter et al., 1996; Derntl et al., 2009; Siegle et al., 2002; Sladky et al., 2013). This heterogeneity might be partly explained by the amygdala's strong adaptiveness, resulting for example in quicker and short-lived responses not being captured by long repetition times (Larson et al., 2006).

Another challenge lies in the location of the amygdala in the medial part of the temporal pole, close to bone tissue and air-filled sinuses (Amunts et al., 2005; Nieuwenhuys et al., 2007) which – if not treated properly – renders it prone to massive susceptibility artefacts via a combination of geometric distortions and signal loss from intra-voxel dephasing (Merboldt et al., 2001; Robinson et al., 2004). In small-sample group studies, it remains unclear if the reported amygdala effects might be caused by only those subjects whose individual anatomy allowed for sufficient amygdala signal quality. In a previous study we have shown an increase in signal quality (improved signal-to-noise-ratio) and percental signal change in the amygdala for 7 T vs 3 T fMRI. By using optimised imaging parameters it became possible to avoid intra-voxel dephasing effects which are due to magnetic field inhomogeneities around the sinuses (Sladky et al., 2013).

Furthermore, the reported amygdala activation changes could also be partly explained by an interaction between task-responses and general context-dependent effects, e.g. the amygdalar reactivity could be higher in novel situations (Balderston et al., 2011), be transiently altered by stressful live situations (Van Wingen et al., 2012) and adopt gradually to the situation of the scanning session (Waraczynski, 2016).

Besides the general critique on fMRI test-retest reliability, several factors may therefore be a source of variable outcomes in studies aiming for amygdala activation (Fischer et al., 2003; Merboldt et al., 2001; Sauder et al., 2013): a) variability in subject compliance b) differences in image acquisition protocols, c) psychological effects related to task-repetition (e.g. habituation) and d) susceptibility artefacts.

In this study, we aimed to investigate on such sources of variance by employing an adapted version of the well-established emotion discrimination task (EDT) where subjects had to match facial emotion by selecting one of two faces that showed the same emotion as a probe face (Hariri et al., 2002).

First, we systematically varied task instructions by including two different sets of instructions on the way they should process facial stimuli. In one condition, subjects were instructed to match emotions of the faces presented (similar to the original paradigm of Hariri and colleagues). In the other condition, subjects were required to match persons instead of emotions. Thereby, we tested for possible differences in amygdala activation across varying processing levels, which is especially important in psychiatric populations where emotion recognition and attentional levels might be compromised or when instructions could be misunderstood.

Second, we examined the potential impact of sequence parameter choice, i.e. timing. We measured half of runs with a TR of 1.4s and whole-brain coverage and the other half at 0.7s and a smaller field of view (though still covering the temporal lobes).

Third, in order to investigate the influence of task repetition within and across sessions on amygdala activity, subjects had to perform six runs (three for each TR) within a session and underwent a re-test session two weeks later. Using a total of 12 runs per subject we assess changes in amygdala activation over three repeated runs, two sessions and across varying instruction and acquisition parameters. Finally, we were employing 7-T fMRI using a sequence that was shown to overcome perisinal susceptibility issues (Sladky et al., 2013).

Within this systematic replication study, we aimed to go beyond direct replication by controlling procedural factors that could lead to a change in study outcomes. We therefore employed the following factors of systematic variability: (1) task instruction, (2) repetition time, and (3) repeated runs and one re-test session.

2. Materials and methods

2.1. Subjects

Fourteen healthy volunteers (7f/7m, mean age: 25.3 ± 3.0 years) were recruited from the general public to participate in this study. All subjects were right-handed, with no self-reported history of neurological or psychiatric disorders. Participants were asked to refrain from caffeine, alcohol and nicotine 6 h prior to the examinations. All subjects were financially reimbursed and did not face any negative consequences from early withdrawal from the study. The study protocol was approved by the institutional review board of the Medical University of Vienna and in accordance with the Declaration of Helsinki. After giving informed written consent to participate in the study, all subjects completed a PANAS questionnaire directly before and then after each examination session. All subjects were scanned twice within a time interval of approximately two weeks (mean ± standard deviation: 14.5 ± 2.3 days) in the afternoon.

2.2. Task

Subjects were examined using an extended version of the emotion discrimination task (EDT), a paradigm well-known to activate brain regions associated with face recognition and emotion processing (Sabatinelli et al., 2005). Subjects are presented with three emotional faces, one on top and two on the bottom (in a triangular configuration), and are required to match the top face to one of the bottom faces during that task.

In the standard EDT task (further referred to as MATCH EMOTIONS, Hariri et al. (2002)), the subjects are presented with three images showing faces, and are asked to match the facial emotional expressions. An object discrimination task (MATCH OBJECTS) is used as control condition. In this task, subjects are presented with shapes of polygons on a skin-toned background of blurred facial information and asked to match the number of corners of the polygons to the target one. In this study, we added an additional variation of the EDT referred to as MATCH PERSONS condition, where subjects were instructed to match facial identities instead of emotional expressions, while being presented with the same visual stimuli as in the MATCH EMOTIONS condition. The task was set up as a block-design, where a fixation cross was shown for 8 s as baseline condition, followed by a screen explaining the subsequent task block (MATCH EMOTIONS, MATCH PERSONS, MATCH OBJECTS) for 3 s and the corresponding task block with a duration of 20 s. The task setup is shown schematically in Fig. 1. Each condition block was repeated three times per run.

During the task, images were displayed using a video projector on a screen at the head end of the scanner. Face triplets were randomly combined at runtime and presented using a Python script. Subjects were able to respond via button-press using an MRI-compatible response box. After button-press or if the subject failed to respond within 5 s, subjects were presented with a new triplet of images.

The face images were taken from the Radboud Faces Database (Langner et al., 2010) and showed faces of Caucasian males and females.

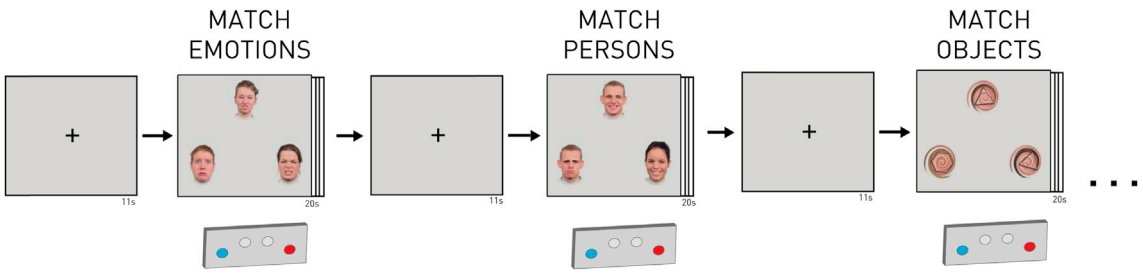

Fig. 1. Flow chart of emotions, persons & objects matching trials. During the task paradigm, participants were presented with three images and instructed to match one of the bottom images to the target image at the top. They were asked either to match emotions, match persons or match objects. Each task block had a duration of 20 s and was followed by 11 s of rest, where a fixation cross was shown to the subjects for 8 s, and the instruction for the next task block was shown for 3 s (MATCH EMOTIONS, MATCH PERSONS, MATCH OBJECTS).

The presented faces expressed the seven basic emotions according to Ekman (1992), namely anger, disgust, fear, happiness, sadness, surprise, contempt, as well as a neutral expression.

Each of the three task blocks was repeated three times per run. The order of task blocks was randomized for each subject to avoid order effects.

2.3. Data acquisition

Data was acquired on a SIEMENS Magnetom 7 T whole-body MR scanner, using a 32-channel head coil with the CMRR multiband EPI (MB-factor = 3) sequence (Moeller et al., 2010), and two different parameter sets with varying times of repetition (TR) and fields of view (FOV). For simplicity, we refer to the different parameter sets by their TR, i.e. TR = 1.4s and TR = 0.7s. For TR = 1.4s, echo time (TE) was 23 ms and 78 slices with a spatial resolution of $1.5 \times 1.5 \times 1$ mm^3 and 0.25 mm slice gap (FOV = 192x192x97.5 mm^3) were acquired at a flip angle of 62° and a bandwidth of 1447 Hz/px. For TR = 0.7s, TE and in-plane resolution were kept identical to the TR = 1.4s runs, but the number of slices was reduced to 39, with slice gaps increased to 0.5 mm (FOV = 192x192x58.5 mm^3). As demonstrated in Fig. 2, while TR = 1.4s allowed for whole-brain imaging, TR = 0.7s was optimised for speed but still allowed for covering all regions of interest (inferior temporal lobe up to corpus callosum). The sequence order was counterbalanced across subjects and sessions.

Runs were repeated three times for each of the two acquisition parameter sets (TR = 1.4 s/TR = 0.7s) and two sessions (session 1/session 2), therefore resulting in a total of 12 runs for each subject.

2.4. Pre-processing

Functional data pre-processing included de-spiking (AFNI), slice-timing correction (FSL) (Sladky et al., 2011), distortion correction (FSL fugue), bias-field correction (ANTs), realignment (FSL), normalization to MNI space (ANTs), and smoothing with a 6 mm FWHM Gaussian kernel (FSL). For optimal normalization results and to ensure comparability over runs, all runs for each subject were co-registered (ANTs) to a mean EPI over all runs with whole-brain coverage (TR = 1.4s); the normalization parameters were calculated using this mean image and applied to all runs. Finally, the first 5 principal components of white matter (WM) and ventricle (CSF) time courses were extracted for nuisance regression. For further information on detailed preprocessing parameters, see Supplementary Information (S5).

2.5. Data analysis

First-level and second-level analyses were conducted in SPM12, post-hoc tests and correlations on ROI results were calculated using MATLAB. For single-subject (first-level) analysis, linear regression was performed at each voxel, using generalized least squares with a global approximate AR(1) autocorrelation model, drift-fit with Discrete Cosine Transform basis (128s cutoff). Task blocks of 20s each were convolved with SPM's canonical HRF and used as regressors of interest. Realignment, CSF and WM parameters obtained from the previous preprocessing steps were included in the model as nuisance regressors.

Resulting single-subject beta maps of BOLD responses were used for group analysis. Separate *flexible factorial designs* (as implemented in SPM12) were computed for TR = 1.4s and TR = 0.7s, each representing 2 factors, i.e. 6 *runs* (3 per session) and 3 *conditions* (MATCH EMOTIONS, MATCH PERSONS, MATCH OBJECTS). The final design matrices therefore contained 18 regressors. Linear regression was performed at each voxel, using generalized least squares with a global repeated measures correlation model. Contrasts were calculated for MATCH EMOTIONS and MATCH PERSONS, respectively, versus the MATCH OBJECTS condition, i.e. MATCH EMOTIONS > MATCH OBJECTS and MATCH PERSONS > MATCH OBJECTS. For direct comparison of both task instructions, the contrast MATCH EMOTIONS > MATCH PERSONS was additionally calculated. Threshold of the t-statistics was set to $p < 0.05$, cluster-wise

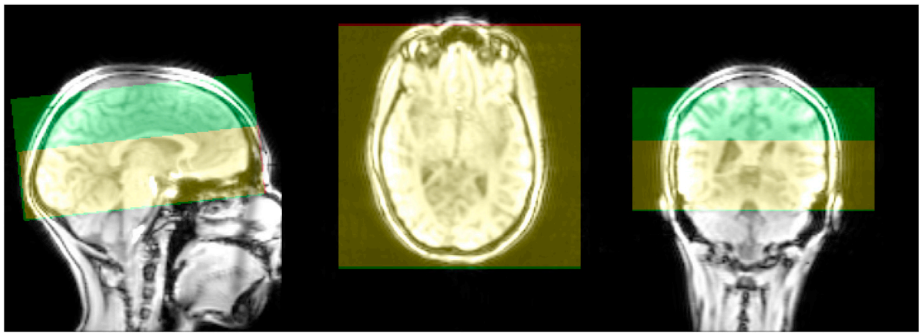

Fig. 2. Placement of Field of View. Example of measurement volumes for one subject. TR = 1.4s allowed for whole-brain coverage (yellow + green). For TR = 0.7s, the number of slices was reduced and the limited FOV was positioned to cover all regions of interest (yellow).

FWE correction (initial cluster defining threshold $p < 0.001$).

Anatomical labelling of activation pattern was based on the TT_Daemon atlas in AFNI (whereami function). Based on the contrast MATCH EMOTIONS > MATCH OBJECTS (all runs of TR = 1.4s) we identified the peak activation coordinates of our regions of interest (ROIs): amygdalae and fusiform gyri. Around the coordinates identified, we extracted parameter estimates, normalised to percent signal change (divided by mean), for all conditions within a sphere of four voxel diameter. Using these values, mean and standard error of the mean (SEM) were calculated for all runs and sessions. The standard error was estimated as the standard deviation of the sample divided by the square root of the sample size, i.e. number of subjects.

Habituation effects across runs were assessed by calculating linear regression slopes of the effect sizes (parameter estimates) for each subject and testing these slopes for group-wise statistical significance using one-sided, one-sample t-tests.

For analysing the influence of task repetition on group-level results, we calculated the spatial correlation of whole-brain second-level activation maps for each run over both sessions. The degree of similarity between first and second session was quantified by calculating the correlation coefficient of the whole-brain beta contrast maps (MATCH EMOTIONS > MATCH OBJECTS) of the first, second and third run in the first session with their respective counterpart in the second session.

For investigations on within-subject reliability, we calculated voxel-based ICC maps and ROI-based ICC values based on means within four-voxel diameter spheres (see above). ICC values (Shrout and Fleiss, 1979) are widely used for reliability analysis. Specifically, we used the ICC(3,1) implementation to assess activation reliability across sessions. This was done for each of the three runs per session (i.e. first run of first session compared to first run of second session, etc.) and for run-averaged maps (mean contrast estimates of first and second run of first session compared to mean contrast estimates of first and second run of second session, etc.). In accordance with Fleiss (1986), we denote ICC values below 0.4 as poor, from 0.4 to 0.59 as fair, 0.60 to 0.75 as good, and above 0.75 as excellent. In addition, Pearson correlation values were calculated in the same manner and included in the supplement (Table S3). We furthermore added ICCs for behavioural reaction times and habituation effects in the supplementary material (S1 and S2).

In order to investigate potential differences related to sequence parameter choice, i.e. TR = 1.4s vs TR = 0.7s, we compared beta-values, normalised to percent signal change, within the first run of the first session for the MATCH EMOTIONS > MATCH OBJECTS contrast. The data was visualised using MATLAB's boxplot function. We additionally calculated contrast-to-noise ratio (CNR) according to Geissler et al. (2007) in order to evaluate data quality for both TRs. We therefore calculated the square root of residual sum-of-squares as an estimate for the standard deviation. The CNR was computed dividing the beta contrast of interest by estimated standard deviation for each subject. We report CNR mean and standard deviation over subjects.

3. Results

3.1. Behavioural data

Mean Reaction times for the different task conditions per run and session are given in Table 1.

Table 1
Response times (s): Time until button press as response to image triplet for each run and session given as mean and standard deviation (both acquisition protocols accumulated).

Condition	Session 1			Session 2		
	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
MATCH EMOTIONS	1.50 ± 0.55	1.48 ± 0.55	1.47 ± 0.58	1.44 ± 0.63	1.47 ± 0.68	1.41 ± 0.6
MATCH PERSONS	1.05 ± 0.44	0.99 ± 0.48	0.97 ± 0.42	1.00 ± 0.47	0.93 ± 0.43	0.91 ± 0.38
MATCH OBJECTS	1.06 ± 0.32	1.02 ± 0.34	0.99 ± 0.3	0.99 ± 0.34	0.98 ± 0.38	0.95 ± 0.31

3.2. Across task instructions

Contrasts of MATCH EMOTIONS and MATCH PERSONS vs control (MATCH OBJECTS) yielded similar activation patterns in the core task regions amygdala and fusiform gyrus (Fig. 3). Importantly, no significant differences in activation levels were found in these core regions between conditions. Regarding DLPFC, however, the MATCH EMOTIONS contrast showed stronger activation (left DLPFC peak coordinate: −51, 27.5, 21.5 [MNI], $T_{peak} = 10.21$; right DLPFC: 51, 29, 18.5, $T_{peak} = 11.14$, all $p_{FWEc} < .05$) compared to the MATCH PERSONS contrast (left DLPFC: −52.5, 33.5, 14 [MNI], $T_{peak} = 3.91$; right DLPFC: 49.5, 29, 17, $T_{peak} = 9.36$, all $p_{FWEc} < .05$). Direct comparison of MATCH EMOTIONS > MATCH PERSONS (Fig. 4) revealed statistically significant differences in both hemispheres (left DLPFC: −54, 24.5, 27.5 [MNI], $T_{peak} = 8.10$; right DLPFC: 55.5 29 21.5 [MNI], $T_{peak} = 6.00$, all $p_{FWEc} < .05$). Fig. 4 further shows the ROI-based percent signal changes in DLPFC activation for both conditions.

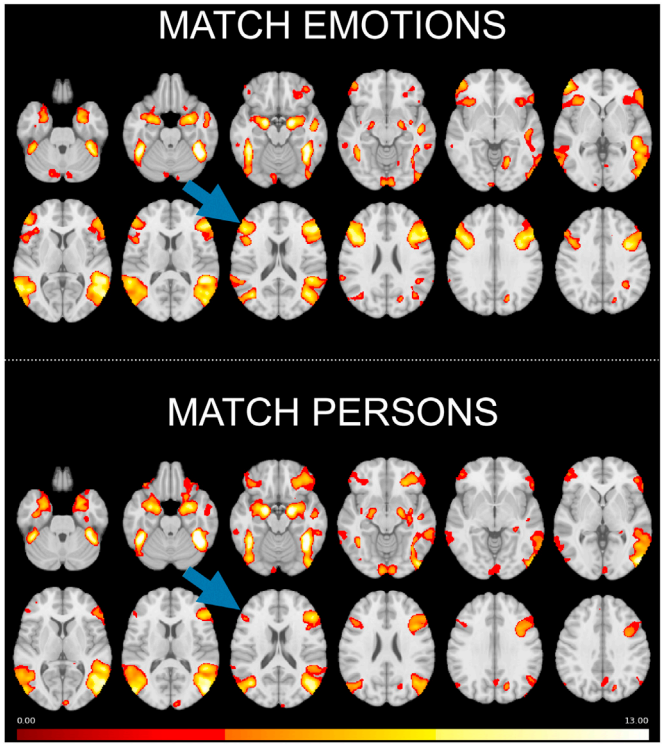

Fig. 3. Activation patterns across task instructions, t-map over whole brain. Amygdala and fusiform cortex activation remained unaffected by the choice of task instruction, while DLPFC showed significantly higher activation only for MATCH EMOTIONS > MATCH OBJECTS compared to MATCH PERSONS > MATCH OBJECTS, pFWEc < .05.

3.3. Across imaging sequences

To account for possible influence of EPI acquisition parameter choices, we compared single-subject contrasts (MATCH EMOTIONS > MATCH OBJECTS, first run of each imaging parameter set in the first session) across imaging parameter sets (TR = 1.4s, TR = 0.7s). Activation in bilateral amygdalae and fusiform gyri was detected in each data set, irrespective of acquisition parameter choice, and no significant difference in response amplitudes was found (Fig. 5). Contrast-to-noise-ratio

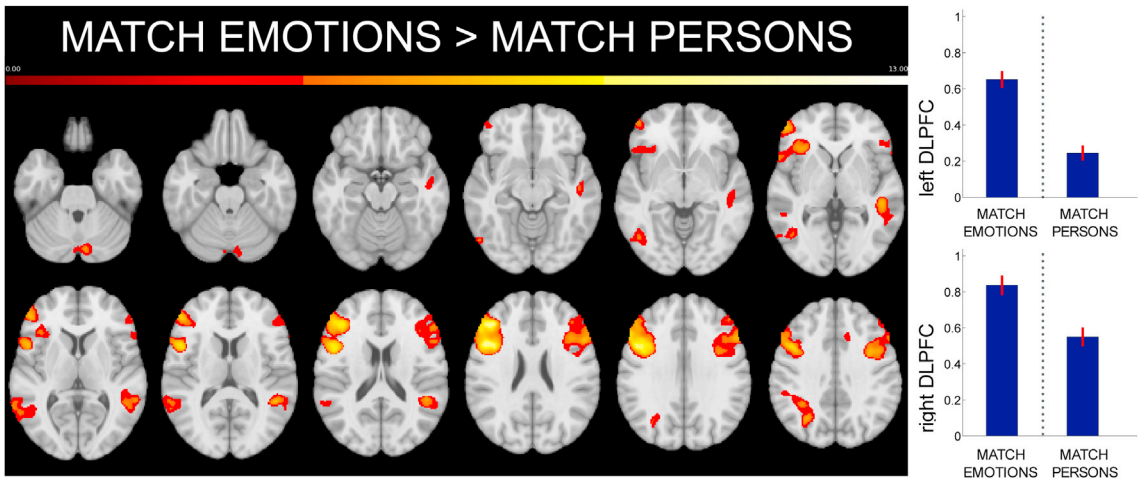

Fig. 4. Contrast MATCH EMOTIONS > MATCH PERSONS, mean of all runs and sessions. Left: t-map over whole brain, pFWEc < .05. No significant difference in activation could be found for amygdala and fusiform gyrus. Significantly higher activation can be seen in bilateral DLPFC (pFWEc < .05) in the MATCH EMOTIONS condition compared to the MATCH PERSONS condition, this effect is particularly high in left DPLFC. Right: bar charts comparing effect sizes (contrasts of MATCH EMOTIONS > MATCH OBJECTS and MATCH PERSONS > MATCH OBJECTS respectively) and standard errors of ROIs (spheres with 4 voxel diameter, coordinates as described in Table 2), over subjects.

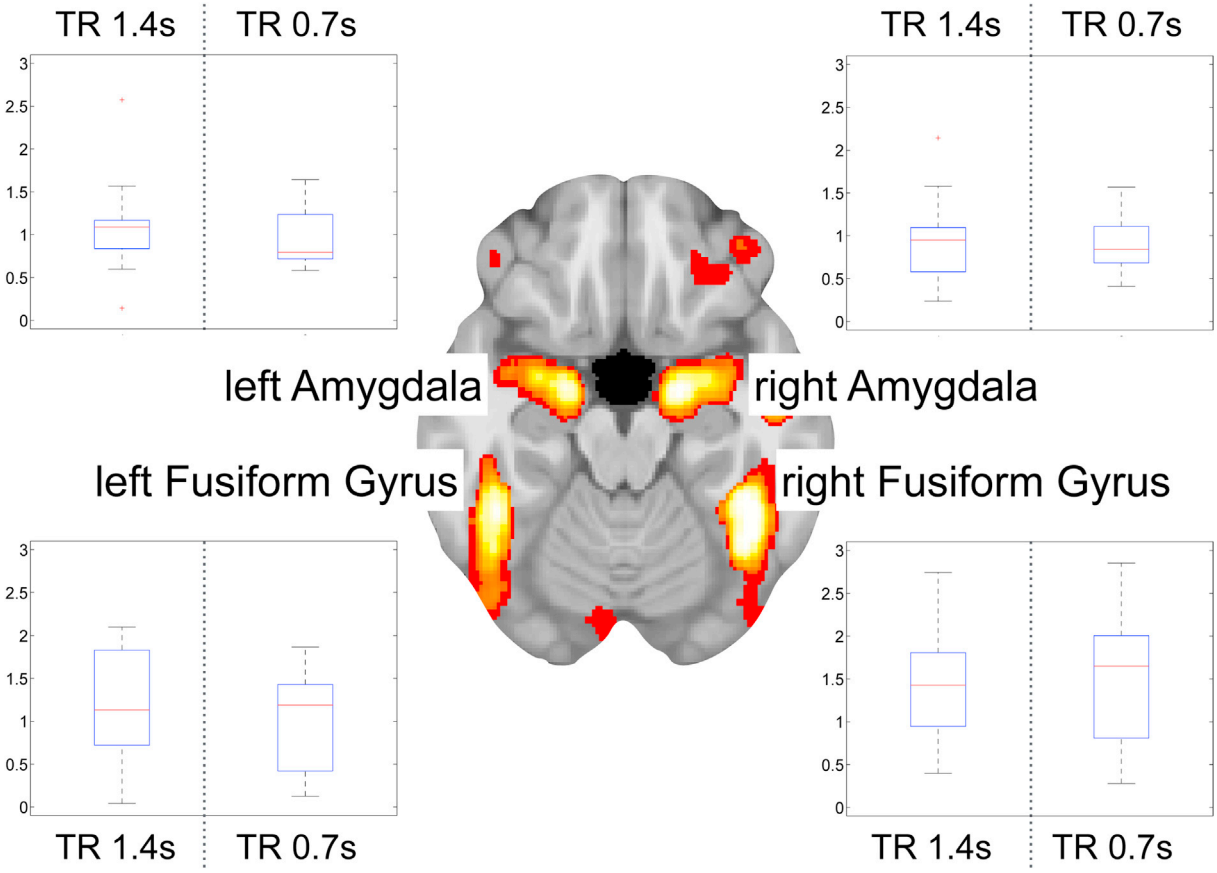

Fig. 5. Comparison of different imaging parameter sets (TR = 1.4s, TR = 0.7s). Response amplitudes are shown as median and quartile plots. No statistically significant differences across acquisition parameter choices were found.

(cnr) was mean ± standard deviation: 2.440 ± 1.560 (TR = 0.7) vs. 3.822 ± 1.704 (TR = 1.4) at the left amygdala (−19.5 -5.5 -17.5 [MNI]); 2.553 ± 1.589 (TR = 0.7) vs. 4.250 ± 1.557 (TR = 1.4) at the right amygdala (19.5 -5.5 -17.5 [MNI]); 5.207 ± 3.073 (TR = 0.7) vs. 6.947 ± 3.105 (TR = 1.4) at the left fusiform gyrus (−43.5 -46 -23.5 [MNI]) and 7.013 ± 3.731 (TR = 0.7) vs. 8.737 ± 3.399 (TR = 1.4) at the right fusiform gyrus (43.5–52 -22 [MNI]).

3.4. Across runs and sessions

Using 7T ultra-high field whole brain EPI (TR = 1.4s), we revealed statistically significant activation changes in bilateral amygdalae and

Table 2
Increase in brain activation during emotion discrimination (MATCH EMOTIONS > MATCH OBJECTS) showing coordinates, t-values, p-values and cluster sizes (k) for bilateral amygdalae, fusiform gyri, and dorsolateral prefrontal cortices (DLPFC). All regions show statistically significant activation for all runs and sessions. Threshold of the t-statistic was set to $p < 0.05$, cluster-wise FWE correction (initial cluster-defining threshold $p < 0.001$).

			Session 1			Session 2		
region		All runs	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
rAmy	coordinate	[19.5 -5.5 -17.5]	[18 -7 -18]	[18 -5.5 -17.5]	[19.5 -4 -19]	[18 -5.5 -19]	[19.5 -5.5 -17.5]	[19.5 -5.5 -19]
	t-value	14.05	11.09	9.17	7.77	9.95	7.92	7.07
	p FWEc p uncorr	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000
	k	3182	4105	2090	739	2585	754	835
lAmy	coordinate	[-19.5 -5.5 -17.5]	[-20 -6 -18]	[-19.5 -5.5 -17.5]	[-19.5 -5.5 -17.5]	[-19.5 -5.5 -17.5]	[-19.5 -4 -19]	[-19.5 -4 -19]
	t-value	12.45	10.6	7.79	7.61	8.67	6.48	5.61
	p FWEc	0.000 0.000	0.000 0.000	0.000	0.000	0.000	0.000	0.001
	p uncorr			0.000	0.000	0.000	0.000	0.018
	k	2536	2507	1489	724	1989	452	390
rFusGyr	coordinate	[43.5–52 -22]	[44 -52 -22]	[43.5–52 -22]	[43.5–52 -22]	[43.5–53.5 -22]	[42 -53.5 -20.5]	[42 -53.5 -20.5]
	t-value	18.05	11.23	11.25	11.05	12.17	11.73	11.14
	p FWEc	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p uncorr	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	k	13429	1732	1711	1915	2217	1965	1799
lFusGyr	coordinate	[-43.5 -46 -23.5]	[-43.5 -46 -22]	[-43.5 -47.5 -22]	[-43.5 -46 -23.5]	[-43.5 -47.5 -22]	[-45 -47.5–23.5]	[-43.5 -58 -17.5]
	t-value	14.99	9.07	8.8	9.51	10.21	9.89	6.31
	p FWEc	0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000
	p uncorr	0.000						
	k	2646	1372	1473	1751	1760	1516	1527
rDLPFC	coordinate	[51 29 18.5]	[51 32 15.5]	[51 27.5 20]	[51 27.5 20]	[52.5 27.5 20]	[51 29 18.5]	[54 32 17]
	t-value	11.14	7.81	6.75	7.11	6.75	7.97	6.88
	p FWEc	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p uncorr	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	k	9540	5581	5802	6449	5461	5540	4768
lDLPFC	coordinate	[-51 27.5 21.5]	[-51 27.5 23]	[-52.5 24.5 26]	[-51 27.5 21.5]	[-51 26 20]	[-49.5 27.5 21.5]	[-51 29 21.5]
	t-value	10.21	7.58	6.38	6.44	6.54	6.81	5.67
	p FWEc p uncorr	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000	0.000 0.000
	k	9278	6723	3622	5018	4711	4623	4163

fusiform cortices in all recorded runs for the contrast MATCH EMOTIONS > MATCH OBJECTS on the group-level (see Fig. 6 and Table 2). While there was a significant response in the bilateral orbito-frontal cortex (OFC) during the first run of the first session, this effect is not present in all the following sessions and runs.

The degree of similarity between first and second session, quantified as correlation over sessions of the second-level whole-brain contrast maps of the MATCH EMOTIONS > MATCH OBJECTS condition, revealed

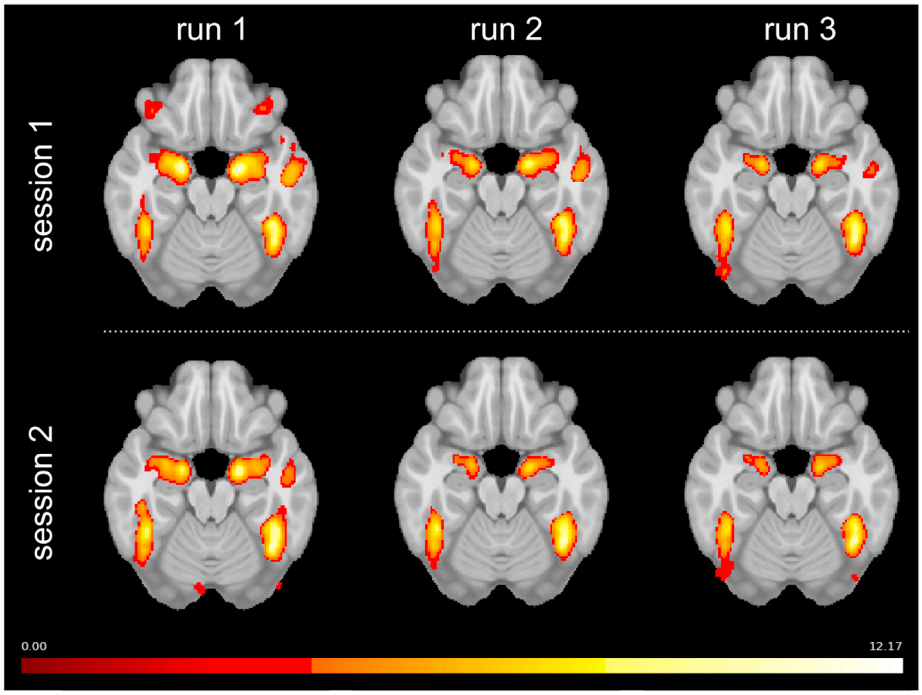

Fig. 6. Reproducibility of amygdalar & fusiform response: fMRI group results across runs and sessions. Importantly, MATCH EMOTIONS > MATCH OBJECTS revealed statistically significant amygdalar and fusiform cortex activation for all runs. The threshold of the t-statistics was set to $p < 0.05$, cluster-wise FWE correction (initial cluster defining threshold $p < 0.001$).

excellent reproducibility with values of r = 0.863 for the first run, r = 0.898 in the second run and r = 0.866 for the third run, respectively.

To further investigate reproducibility of whole-brain task-induced activation changes across runs, we binarised each of the six group-level SPMs (corresponding to the six runs). The sum over these binarised maps resulted in an overlap map of all 6 runs representing the number of runs (colour-coded) where each voxel was activated above threshold (Fig. 7).

As illustrated, amygdalae and fusiform gyri show supra-threshold activation in each run. Effect sizes across runs and sessions for both ROIs are shown in Fig. 8. Both right and left amygdala activation levels exhibit statistically significant habituation for both sessions, while fusiform gyrus shows habituation effects for the right side and the second session only. All details are given in Table 3. Notably, changes in amygdala activation levels are remarkably similar for both hemispheres and sessions (all between 0.12 and 0.16 per run).

As shown in Table 4, there is no significant inter-session habituation for single runs (slope between run 1 session 1 and run 1 session 2) in bilateral amygdala, fusiform gyrus and DLPFC. However, a comparison between the slopes for the mean activation changes between session 1 and 2 revealed a significant slope for the left amygdala.

3.5. Within-subject reliability

Within-subject reliability was assessed by calculating ICC values over runs and sessions. ICC values for each ROI, based on one, two or three runs are shown in Table 5 and Fig. 9. A comparison of first run of both sessions shows a good reliability for right amygdala (ICC: 0.603) and fair reliability for left amygdala (ICC: 0.484). Correlating an average of first two runs already leads to good ICC values in both amygdalae (0.6565 and 0.6825), and reaching excellent reliability for averages of 3 runs (0.7912 and 0.7530). Voxel-wise ICC brain-maps are displayed in Fig. 10.

Results for fusiform gyrus and DLPFC also show an increase in ICC values over the number of averaged runs. Pearson correlation coefficients (please see supplement S3) show very similar results.

4. Discussion

Within this study, we investigated the systematic replicability and reliability of fMRI activation during facial emotion processing, with particular focus on the amygdala. We acquired 7T fMRI data using two task instruction approaches, two image acquisition protocols and multiple repetitions across runs and sessions. In short, statistically significant group-level activation for amygdala and fusiform gyrus was found in each of the twelve runs. Inter-session reliability (single-subject level) was fair to good for the first run and excellent for averages over runs.

4.1. Influence of task instruction

By varying task instructions, we accounted for possible differences in amygdala activation due to different cognitive processing strategies resulting from misunderstandings or even cognitive limitations. While presented with the same visual stimuli, subjects were instructed to either MATCH EMOTIONS, i.e. evoking intended emotional processing, or MATCH PERSONS, i.e. implying unintended emotional processing of face stimuli.

Former studies have used variations of the emotion discrimination task by introducing conditions such as age matching (Habel et al., 2007), gender matching (Lange et al., 2003), person matching (Dannlowski et al., 2012; Lois et al., 2018; Pezawas et al., 2005; Smith et al., 2017; Swartz et al., 2015) as well as passive viewing of emotional faces (Lange et al., 2003; Lois et al., 2018).

Lois et al. (2018) found robust amygdala activation in both person matching and passive viewing, while in contrast, Lange et al. (2003) reported increased amygdala involvement during passive viewing compared to both age matching and emotion matching.

A large corpus of studies has shown amygdala activity in different age groups and neuro-psychiatric conditions (Dannlowski et al., 2012; Pezawas et al., 2005; Smith et al., 2017; Swartz et al., 2015) for person matching condition using fearful and angry faces exclusively.

In our study, we have found highly significant activation changes in amygdala and fusiform gyrus not only for emotion matching, but also for the person matching condition over basic emotions according to Ekman (1992). While this corroborates the results of Lois et al. (2018), it is in contrast to Habel et al. (2007) who reported higher activation in the amygdala during emotion attribution compared to age attribution. One explanation could be found in the fact that we used the identical visual stimuli in both instruction conditions, whereas seeing specific emotions as words may act as primer, trigger more emotional responses and therefore lead to stronger activation (Schlochtermeier et al., 2013).

In contrast to Habel et al. (2007) and Lange et al. (2003), which both reported differences in activation between conditions, we showed identical visual stimuli and only varied the instructions on how to process images, i.e. match persons versus match emotions. Between both conditions, we have found neither significant difference in amygdala activation nor in fusiform gyrus activation. Given that identical stimuli with

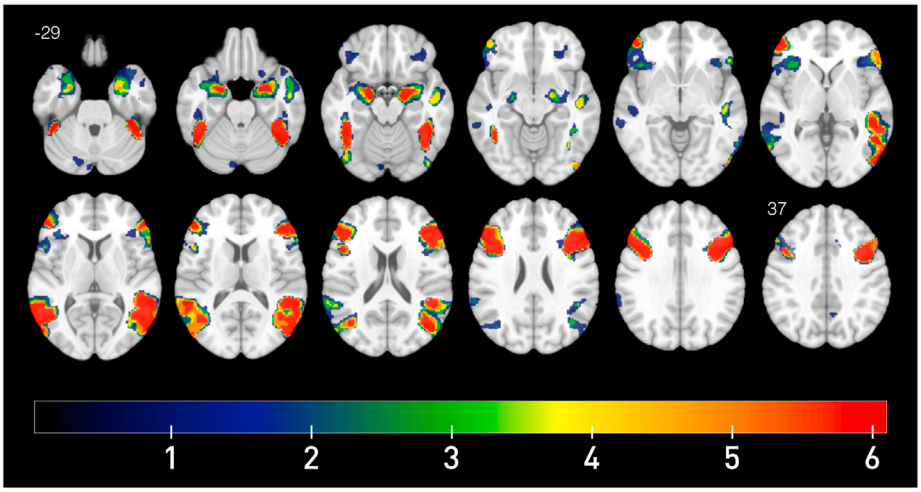

Fig. 7. Reproducibility map across all six runs. Group SPMs for separate runs were binarised ($p < 0.05$ $FWE_{cluster\text{-}wise}$) and added. It can be seen that the MATCH EMOTIONS > MATCH OBJECTS contrast shows statistically significant activation in bilateral amygdala, fusiform cortex, and DLPFC for all six runs (red areas). Insular and orbitofrontal areas were activated in some runs only.

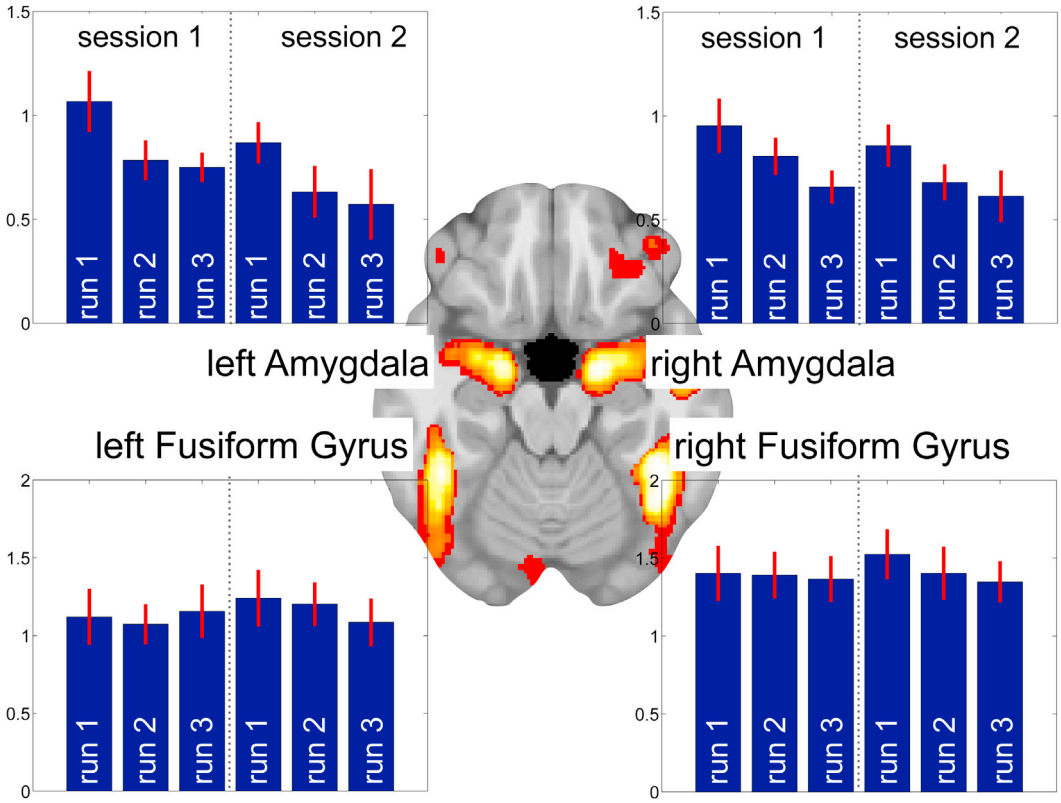

Fig. 8. Response variability across runs and sessions. ROI analyses for amygdala and fusiform gyrus peak activations over 2 sessions and 3 runs, respectively: mean and standard error of percent signal change for MATCH EMOTIONS > MATCH OBJECTS. All four regions show statistically significant activation in each run. Statistically significant habituation effects over runs within sessions were found for bilateral amygdalae and both sessions, whereas fusiform gyrus showed habituation for the second session only (for details refer to Table 3).

Table 3
Intra-session habituation: Mean linear regression slopes across runs. Statistically significant slopes ($p < .05$) are printed in bold. Bilateral amygdalae show statistically significant habituation over runs within sessions. For fusiform gyrus, statistically significant habituation was found only in the second session.

	session 1		session 2	
	slope	p-value	slope	p-value
Amygdala right	**−0.1478**	**0.0375**	**−0.1218**	**0.0102**
Amygdala left	**−0.1587**	**0.0256**	**−0.1480**	**0.0140**
Fusiform gyrus right	−0.0181	0.3822	**−0.0884**	**0.0259**
Fusiform gyrus left	0.0183	0.3865	−0.0777	0.0554
DLPFC right	−0.0166	0.4089	−0.0321	0.3650
DLPFC left	−0.0460	0.1947	−0.0394	0.2553

varying task instructions results in similar activation patterns (see Fig. 3) we conclude that the EDT implementation used in this study should be applicable also in populations with limited compliance, such as trials with subjects suffering from psychiatric disorders, children and elderly, as well as populations with cognitive impairment or neurological problems (Pereira et al., 2016; Vanderwal et al., 2015).

Regarding DLPFC, we found a significantly higher DLPFC activation during the MATCH EMOTIONS task compared to the MATCH PERSONS task, indicating that emotion discrimination is a cognitively more demanding task. The DLPFC has been associated with emotion regulation (Banks et al., 2007; Goldin et al., 2008; Mauss et al., 2007) and shown to be down-regulated in unipolar depression (Siegle et al., 2007). We therefore suggest that DLPFC activation during the EDT is an indicator for intentional emotion processing and possible regulation strategies while amygdala activity increase might be a more basal response to emotive faces.

Table 4
Inter-session habituation: Mean linear regression slopes across sessions: comparing first run of each session, and mean of all runs across sessions. Statistically significant slopes ($p < .05$) are printed in bold. Right amygdala does not show statistically significant habituation over sessions. Left amygdala shows no significant difference in the first run over sessions, however significant habituation was found when comparing means of all 3 runs over sessions. For fusiform gyrus and DLPFC no statistically significant difference was found between session.

	first run over sessions		mean of runs over sessions	
	slope	p-value	slope	p-value
Amygdala right	−0.0961	0.1925	−0.0887	0.0786
Amygdala left	−0.1989	0.0691	**−0.1759**	**0.0195**
Fusiform gyrus right	0.1221	0.1687	0.0383	0.2855
Fusiform gyrus left	0.1203	0.1455	0.0594	0.1568
DLPFC right	−0.0812	0.2884	−0.0132	0.4078
DLPFC left	−0.1071	0.1533	−0.0401	0.2606

4.2. *Influence of imaging sequence*

An additional aspect in our study targeted the effects of differences in imaging protocols. We first chose an EPI protocol combining good temporal (1.4s) and high spatial resolution (78 slices, 1.5 × 1.5 × 1 mm^3) comparable to the current state-of-the-art for high-resolution 7T studies (Moeller et al., 2010; Triantafyllou et al., 2005). A second protocol has been chosen to account for variations in imaging parameters throughout different neuroscience studies from different labs. Former studies have argued that shorter TRs at high spatial resolution increases SNR per time unit and increases BOLD sensitivity (Yoo et al., 2018). To our knowledge this, however, has not been directly tested in an emotion discrimination paradigm yet.

Table 5
Between-session ICC values of MATCH EMOTIONS > MATCH OBJECTS contrasts in ROIs based on the individual runs or averaged over several runs.

	Intra-class correlation coefficient (ICC)				
	first run	second run	third run	Mean of run1 & run2	Mean of all runs
Amygdala right	0.6034	0.4327	0.0395	0.6565	0.7912
Amygdala left	0.4843	0.4678	0.1484	0.6825	0.7530
Fusiform gyrus right	0.7322	0.6993	0.7968	0.8820	0.9006
Fusiform gyrus left	0.7769	0.8615	0.8779	0.8705	0.9236
DLPFC right	0.4066	0.5550	0.3678	0.8247	0.8713
DLPFC left	0.5940	0.7001	0.6542	0.8516	0.8252

As the necessity for altering imaging protocols towards higher sampling rates is particularly indicated in applications such as real-time fMRI (Lorenz et al., 2018) and brain computer interfaces (BCI), we have compared task-induced activation across two different acquisition parameter sets. By changing TR from 1400 ms to 700 ms, we doubled image acquisition frequency, but had to reduce the number of slices. A slight increase of slice gaps allowed us to ensure that slabs still covered temporal and occipital lobes. It must be noted that our comparison employed image acquisition protocols with high spatial resolution (voxel size 1.5x1.5 $\times$ 1mm^3) to achieve high amygdala SNR through reduced intra-voxel dephasing effects (Merboldt et al., 2001; Robinson et al., 2004). Such a combination of high-resolution acquisition and whole-brain coverage (including temporal lobes) in just 1.4s requires the use of multiband (MB) or simultaneous multi-slice (SMS) sequences.

As theorized by Murphy et al. (2007), Yoo et al. (2018) demonstrated higher effect sizes by reduction of TR in the motor-network. Comparing lower and higher TRs in our study did not yield significant differences in fusiform gyrus and amygdala ROI activation levels. At the individual level, contrast-to-noise ratio (CNR) were higher for longer vs shorter TR. Systematic variation of EPI parameters might help in the future to identify optimal sequence parameters for amygdalar activation acquisition. Importantly, as shown in Fig. 5 group-level variance is similar for the two acquisition protocols investigated. Systematic reproducibility of emotional paradigms between different scanner site was demonstrated by Gee et al. (2015) for amygdala activations with travelling participants.

4.3. Influence of task repetition across runs and sessions

Each subject underwent fMRI scanning in two sessions, and each session comprised six runs of facial emotion processing. Our results clearly show statistically significant amygdala activation in each run for both sessions at group level. In addition, activation in the amygdala region does not follow the trajectory of the basal vein of Rosenthal (BVR). This result clearly indicates that our 7T activation maps are not a product of the venous confounds recently reported by Boubela et al. (2015).

From our results we argue that the EDT task – as used herein – has the potential to increase amygdala activity and is thus suitable for clinical applications aiming at between and within-group comparisons, such as investigations on effects of pharmacotherapy (e.g. Duff et al. (2015)) or psychotherapy (e.g. Marini et al. (2015)).

Nevertheless, we also found that parameter estimate amplitudes decreased across runs within a session indicating intra-session habituation effects. Habituation, i.e. a response decrease after repeated exposition is a highly adaptive mechanism, and habituation abnormalities are hypothesized to be at the core of various psychiatric disorders (McDiarmid et al., 2017).

In the literature there is a differentiation between intra- and inter-session habituation. In healthy subjects, while some studies do not find a decrease of intra-session amygdala activation (Gur et al., 2007; Infantolino et al., 2018; Sladky et al., 2012), others do report amygdala habituation (Breiter et al., 1996; Fischer et al., 2003; Plichta et al., 2014). Habituation patterns were also investigated in patient populations. In contrast to patients suffering from child sexual abuse related PTSD (van den Bulk et al., 2016) and social anxiety disorder (Sladky et al., 2012), where increased amygdala habituation compared to healthy control groups was reported, the contrary effect was found in populations with autism spectrum disorder (Kleinhans et al., 2009; Swartz et al., 2013) and patients suffering from schizophrenia (Holt et al., 2005; Williams et al., 2013). Gur et al. (2007) describe no habituation effects for schizophrenic subjects. Different patterns in habituation might partly be explained by the psychiatric condition itself.

Inter-session habituation effects have also been proposed. Johnstone et al. (2005) suggest that a distance of two weeks is sufficient to avoid habituation effects in the second session. In this regard, Spohrs et al. (2018) could recently not show any significant inter-session habituation effect in a re-test session after around three weeks. Also, Plichta et al. (2014) could not show any inter-session habituation effect in a re-measuring session after around two weeks. In our study we could also not show consistent inter-session habituation effect in the amygdala after

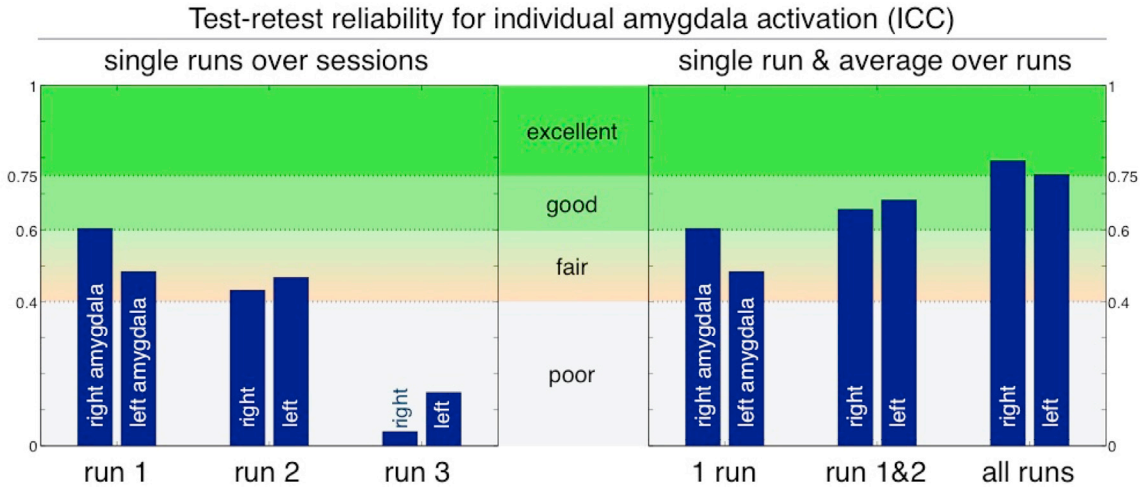

Fig. 9. Bar charts of amygdala ICC values. Run-based ICC for each of the three runs showing changing reliability over runs (left) and ICCs over sessions based on the number of runs averaged (right). Reliability for right/left amygdala increases averaging from good/fair (one run) to good/good (two runs averaged) and excellent/excellent (three runs averaged). For details see Table 5.

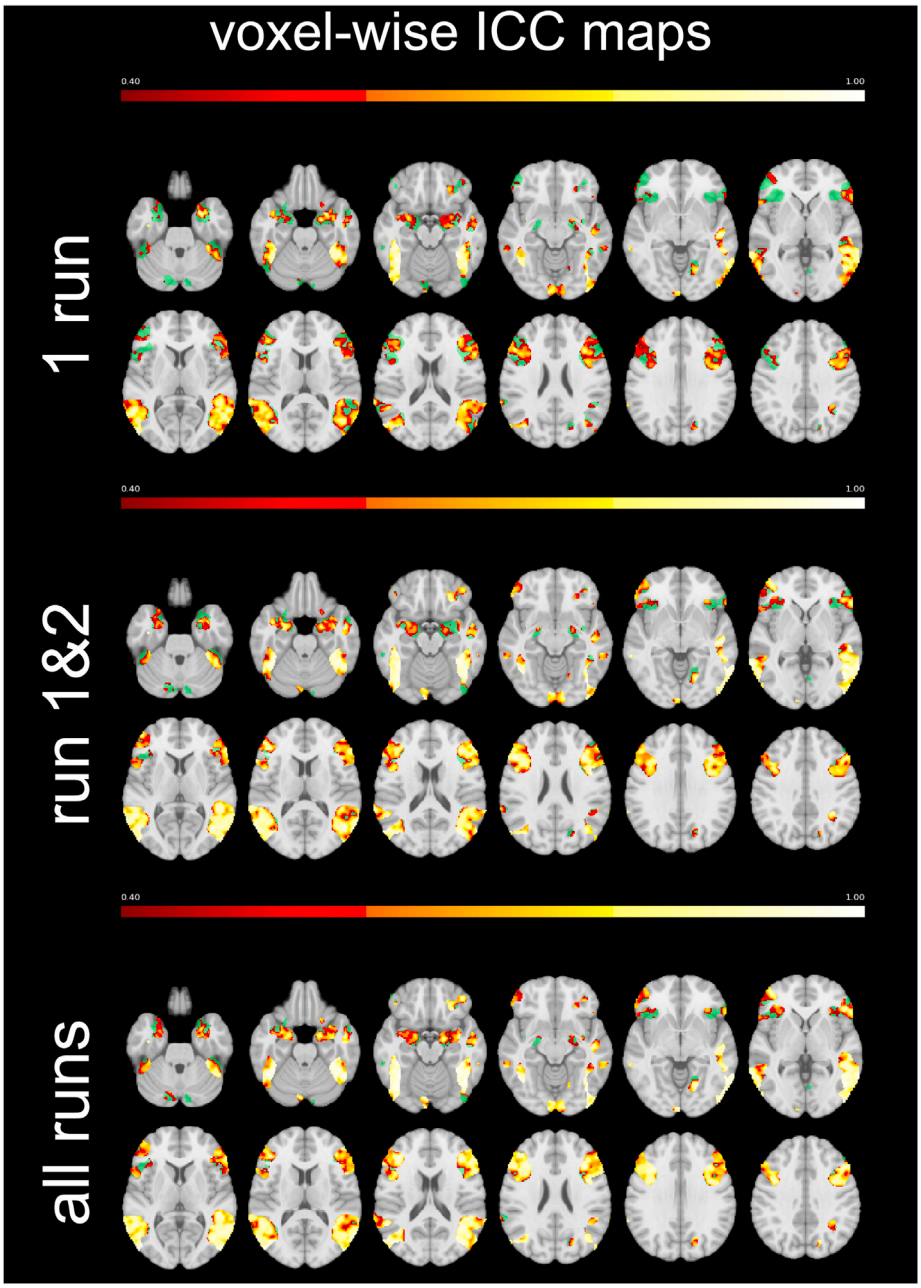

Fig. 10. Voxel-wise ICC maps over sessions based on first run (top), an average over the first two runs (middle) and all three runs (bottom) respectively. ICC values are shown thresholded above 0.40 (fair reliability and above). Masked by activation map of MATCH EMOTIONS > MATCH OBJECTS over all runs (pFWEc < .05), masks shown as green underlay.

a pause of two weeks when comparing first runs of each session.

Regarding reliability, on a single subject level habituation effects were not consistent indicated by variable slopes. This suggests that individual habituation effects might vary between sessions possibly due to differences in emotion regulation strategies (Morawetz et al., 2016). Animal studies further suggest different habituation patterns for different subsets of neurons in the amygdala (Ben-Ari and La Salle, 1974; Herry et al., 2007). Future research has to account for specific influencing factors on habituation such as the above-mentioned pathologies and personal differences per se. Furthermore, fMRI at high resolution might allow for insights in sub-nuclei specific habituation patterns.

Regarding within-subject reliability across sessions, we performed comprehensive correlation analyses. Based on the ICC results, we may conclude that the first run shows good/fair reliability for right/left amygdala activation levels. Averaging over two runs yields good reliability in both hemispheres. Even excellent reliability may be achieved by averaging across all three runs.

Behavioural data as reflected in reaction times did reveal a poor to good reliability in the match emotion conditions and a good to excellent reliability in the match persons condition. Concerning the relationship between reliability in behavioural and neuronal data White et al. (2016) suggests a mismatch between reliability of amygdalar connectivity and behavioural response in a dot probe paradigm. They thereby stress the dissociation between neural reactivity and response. Based on our finding, their results might be relativized in terms of task dependency. We could show here that matching faces results in better behavioral reliability than emotion matching.

Speculating on the reasons for generally higher ICCs in this study compared to previous reports, we find three possible causes. First, this study was performed at ultra-high magnetic field (7 T). Higher magnetic

fields offer not only higher MR signals but also increased BOLD-related signal changes, effectively increasing signal-to-noise and contrast-to-noise ratios (Ugurbil et al., 2003). Second, imaging data was acquired using a protocol employing high spatial resolution (1.5x1.5 × 1mm^3). Only such high-resolution approaches will compensate the increased signal losses from dephasing effects along susceptibility borders (Merboldt et al., 2001) and enable us to fully benefit from the higher sensitivity at 7T (Sladky et al., 2013). Third, we want to suggest that randomization of stimuli might be a factor increasing re-test responses. In an animal study, Herry et al. (2007) found decrease habituation patterns in specific amygdalar neurons for unpredictable stimuli. It can be summarized that all the three factors together might have contributed to increased reliability. Future work has to be done to isolate the three factors in order to explain how strong each of them contributes to improved reliability measures.

4.4. Summary

Taken together, at group level we have found replicable activation in the amygdala during facial emotion processing across all premises, including repeated measurements, task instructions and imaging parameters. In addition, we have also shown habituation effects in amygdala response amplitude within sessions that were remarkably similar across repeated sessions. We conclude that using high-resolution multiband imaging methods it is possible to obtain replicable activation maps in the amygdala at 7 T. These results are suitable to inform methods on the feasibility and validity of amygdala fMRI and encourage future studies using this facial emotion processing in clinical and non-clinical applications.

Acknowledgements

This research is supported by the CREAM project that has been funded by the European Commission under Grant Agreement no 612022 (FP7 ICT 2013-10). This publication reflects the views only of the authors, the European Commission cannot be held responsible for any use that may be made of the information contained herein. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116585.

References

Alcauter, S., Lin, W., Smith, J.K., Short, S.J., Goldman, B.D., Reznick, J.S., Gilmore, J.H., Gao, W., 2014. Development of thalamocortical connectivity during infancy and its cognitive correlations. J. Neurosci. 34, 9067–9075.

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N., Habel, U., Schneider, F., Zilles, K., 2005. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat. Embryol. 210, 343–352.

Balderston, N.L., Schultz, D.H., Helmstetter, F.J., 2011. The human amygdala plays a stimulus specific role in the detection of novelty. Neuroimage 55, 1889–1898.

Banks, S.J., Eddy, K.T., Angstadt, M., Nathan, P.J., Phan, K.L., 2007. Amygdala–frontal connectivity during emotion regulation. Soc. Cognit. Affect Neurosci. 2, 303–312.

Ben-Ari, Y., La Salle, G.L.G., 1974. Lateral amygdala unit activity: II. Habituating and non-habituating neurons. Electroencephalogr. Clin. Neurophysiol. 37, 463–472.

Boubela, R.N., Kalcher, K., Huf, W., Seidel, E.M., Derntl, B., Pezawas, L., Nasel, C., Moser, E., 2015. fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. Sci. Rep. 5, 10499.

Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S.E., Rosen, B.R., 1996. Response and habituation of the human amygdala during visual processing of facial expression. Neuron 17, 875–887.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365.

Cremers, H., Lee, R., Keedy, S., Phan, K.L., Coccaro, E., 2016. Effects of escitalopram administration on face processing in intermittent explosive disorder: an fMRI study. Neuropsychopharmacology 41, 590.

Dannlowski, U., Stuhrmann, A., Beutelmann, V., Zwanzger, P., Lenzen, T., Grotegerd, D., Domschke, K., Hohoff, C., Ohrmann, P., Bauer, J., 2012. Limbic scars: long-term consequences of childhood maltreatment revealed by functional and structural magnetic resonance imaging. Biol. Psychiatr. 71, 286–293.

Derntl, B., Windischberger, C., Robinson, S., Kryspin-Exner, I., Gur, R.C., Moser, E., Habel, U., 2009. Amygdala activity to fear and anger in healthy young males is associated with testosterone. Psychoneuroendocrinology 34, 687–693.

Duff, E.P., Vennart, W., Wise, R.G., Howard, M.A., Harris, R.E., Lee, M., Wartolowska, K., Wanigasekera, V., Wilson, F.J., Whitlock, M., 2015. Learning to identify CNS drug action and efficacy using multistudy fMRI data. Sci. Transl. Med. 7, 274ra216-274ra216.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. In: Proceedings of the National Academy of Sciences, 201602413.

Ekman, P., 1992. Are There Basic Emotions?.

Fischer, H., Wright, C.I., Whalen, P.J., McInerney, S.C., Shin, L.M., Rauch, S.L., 2003. Brain habituation during repeated exposure to fearful and neutral faces: a functional MRI study. Brain Res. Bull. 59, 387–392.

Fleiss, J.L., 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons Inc.

2017 Fostering reproducible fMRI research. Nat. Neurosci. 20, 298.

Fournier, J.C., Chase, H.W., Almeida, J., Phillips, M.L., 2014. Model specification and the reliability of fMRI results: implications for longitudinal neuroimaging studies in psychiatry. PloS One 9, e105169.

Gee, D.G., McEwen, S.C., Forsyth, J.K., Haut, K.M., Bearden, C.E., Addington, J., Goodyear, B., Cadenhead, K.S., Mirzakhanian, H., Cornblatt, B.A., 2015. Reliability of an fMRI paradigm for emotional processing in a multisite longitudinal study. Hum. Brain Mapp. 36, 2558–2579.

Geissler, A., Gartus, A., Foki, T., Tahamtan, A.R., Beisteiner, R., Barth, M., 2007. Contrast-to-noise ratio (CNR) as a quality parameter in fMRI. J. Magn. Reson. Imag. 25, 1263–1270. An Official Journal of the International Society for Magnetic Resonance in Medicine.

Goldin, P.R., McRae, K., Ramel, W., Gross, J.J., 2008. The neural bases of emotion regulation: reappraisal and suppression of negative emotion. Biol. Psychiatr. 63, 577–586.

Gur, R.E., Loughead, J., Kohler, C.G., Elliott, M.A., Lesko, K., Ruparel, K., Wolf, D.H., Bilker, W.B., Gur, R.C., 2007. Limbic activation associated with misidentification of fearful faces and flat affect in schizophrenia. Arch. Gen. Psychiatr. 64, 1356–1366.

Habel, U., Windischberger, C., Derntl, B., Robinson, S., Kryspin-Exner, I., Gur, R.C., Moser, E., 2007. Amygdala activation and facial expressions: explicit emotion discrimination versus implicit emotion processing. Neuropsychologia 45, 2369–2377.

Hariri, A.R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M.F., Weinberger, D.R., 2002. Serotonin transporter genetic variation and the response of the human amygdala. Science 297, 400–403.

Herry, C., Bach, D.R., Esposito, F., Di Salle, F., Perrig, W.J., Scheffler, K., Lüthi, A., Seifritz, E., 2007. Processing of temporal unpredictability in human and animal amygdala. J. Neurosci. 27, 5958–5966.

Holt, D.J., Weiss, A.P., Rauch, S.L., Wright, C.I., Zalesak, M., Goff, D.C., Ditman, T., Welsh, R.C., Heckers, S., 2005. Sustained activation of the hippocampus in response to fearful faces in schizophrenia. Biol. Psychiatr. 57, 1011–1019.

Huang, X., Huang, P., Li, D., Zhang, Y., Wang, T., Mu, J., Li, Q., Xie, P., 2014. Early brain changes associated with psychotherapy in major depressive disorder revealed by resting-state fMRI: evidence for the top-down regulation theory. Int. J. Psychophysiol. 94, 437–444.

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.

Johnstone, T., Somerville, L.H., Alexander, A.L., Oakes, T.R., Davidson, R.J., Kalin, N.H., Whalen, P.J., 2005. Stability of amygdala BOLD response to fearful faces over multiple scan sessions. Neuroimage 25, 1112–1123.

Kleinhans, N.M., Johnson, L.C., Richards, T., Mahurin, R., Greenson, J., Dawson, G., Aylward, E., 2009. Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. Am. J. Psychiatr. 166, 467–475.

Lange, K., Williams, L.M., Young, A.W., Bullmore, E.T., Brammer, M.J., Williams, S.C.R., Gray, J.A., Phillips, M.L., 2003. Task instructions modulate neural responses to fearful facial expressions. Biol. Psychiatr. 53, 226–232.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A., 2010. Presentation and validation of the Radboud faces database. Cognit. Emot. 24, 1377–1388.

Larson, C.L., Schaefer, H.S., Siegle, G.J., Jackson, C.A., Anderle, M.J., Davidson, R.J., 2006. Fear is fast in phobic individuals: amygdala activation in response to fear-relevant stimuli. Biol. Psychiatr. 60, 410–417.

Lois, G., Kirsch, P., Sandner, M., Plichta, M.M., Wessa, M., 2018. Experimental and methodological factors affecting test-retest reliability of amygdala BOLD responses. Psychophysiology e13220.

Lorenz, R., Violante, I.R., Monti, R.P., Montana, G., Hampshire, A., Leech, R., 2018. Dissociating frontoparietal brain networks with neuroadaptive Bayesian optimization. Nat. Commun. 9, 1227.

Manuck, S.B., Brown, S.M., Forbes, E.E., Hariri, A.R., 2007. Temporal stability of individual differences in amygdala reactivity. Am. J. Psychiatr. 164, 1613–1614.

Marini, S., Cinosi, E., Lupi, M., Acciavatti, T., Corbo, M., Di Tizio, L., Dezi, S., Di Giannantonio, M., 2015. Psychoanalysis and functional magnetic resonance imaging: a systematic review. Eur. Psychiatr. 30, 332.
Mauss, I.B., Bunge, S.A., Gross, J.J., 2007. Automatic emotion regulation. Soc. Pers. Psychol. Compass 1, 146–167.
McDiarmid, T.A., Bernardos, A.C., Rankin, C.H., 2017. Habituation is altered in neuropsychiatric disorders—a comprehensive review with recommendations for experimental design and analysis. Neurosci. Biobehav. Rev. 80, 286–305.
McGonigle, J., Murphy, A., Paterson, L.M., Reed, L.J., Nestor, L., Nash, J., Elliott, R., Ersche, K.D., Flechais, R.S., Newbould, R., 2017. The ICCAM platform study: an experimental medicine platform for evaluating new drugs for relapse prevention in addiction. Part B: fMRI description. J. Psychopharmacol. 31, 3–16.
Merboldt, K.-D., Fransson, P., Bruhn, H., Frahm, J., 2001. Functional MRI of the human amygdala? Neuroimage 14, 253–257.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63, 1144–1153.
Morawetz, C., Kellermann, T., Kogler, L., Radke, S., Blechert, J., Derntl, B., 2016. Intrinsic functional connectivity underlying successful emotion regulation of angry faces. Soc. Cognit. Affect Neurosci. 11, 1980–1991.
Murphy, K., Bodurka, J., Bandettini, P.A., 2007. How long to scan? The relationship between fMRI temporal signal to noise ratio and necessary scan duration. Neuroimage 34, 565–574.
Nathan, P.J., Phan, K.L., Harmer, C.J., Mehta, M.A., Bullmore, E.T., 2014. Increasing pharmacological knowledge about human neurological and psychiatric disorders through functional neuroimaging and its application in drug discovery. Curr. Opin. Pharmacol. 14, 54–61.
Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci. 20, 299–303.
Nieuwenhuys, R., Voogd, J., Van Huijzen, C., 2007. The Human Central Nervous System: a Synopsis and Atlas. Springer Science & Business Media.
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.
Pereira, R.F., Heinsfeld, A.S., Franco, A.R., Buchweitz, A., Meneguzzi, F., 2016. Detecting task-based fMRI compliance using plan abandonment techniques. GigaScience 5, 21–22.
Pezawas, L., Meyer-Lindenberg, A., Drabant, E.M., Verchinski, B.A., Munoz, K.E., Kolachana, B.S., Egan, M.F., Mattay, V.S., Hariri, A.R., Weinberger, D.R., 2005. 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: a genetic susceptibility mechanism for depression. Nat. Neurosci. 8, 828.
Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A.B., Sauer, C., Tost, H., Esslinger, C., 2012. Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. Neuroimage 60, 1746–1758.
Plichta, M.M., Grimm, O., Morgen, K., Mier, D., Sauer, C., Haddad, L., Tost, H., Esslinger, C., Kirsch, P., Schwarz, A.J., 2014. Amygdala habituation: a reliable fMRI phenotype. Neuroimage 103, 383–390.
Robinson, S., Windischberger, C., Rauscher, A., Moser, E., 2004. Optimized 3 T EPI of the amygdalae. Neuroimage 22, 203–210.
Sabatinelli, D., Bradley, M.M., Fitzsimmons, J.R., Lang, P.J., 2005. Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. Neuroimage 24, 1265–1270.
Sauder, C.L., Hajcak, G., Angstadt, M., Phan, K.L., 2013. Test-retest reliability of amygdala response to emotional faces. Psychophysiology 50, 1147–1156.
Schacher, M., Haemmerle, B., Woermann, F., Okujava, M., Huber, D., Grunwald, T., Krämer, G., Jokeit, H., 2006. Amygdala fMRI lateralizes temporal lobe epilepsy. Neurology 66, 81–87.
Schlochtermeier, L.H., Kuchinke, L., Pehrs, C., Urton, K., Kappelhoff, H., Jacobs, A.M., 2013. Emotional picture and word processing: an fMRI study on effects of stimulus complexity. PloS One 8, e55619.
Schmidt, S., 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. Rev. Gen. Psychol. 13, 90–100.
Sheline, Y.I., Barch, D.M., Donnelly, J.M., Ollinger, J.M., Snyder, A.Z., Mintun, M.A., 2001. Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study. Biol. Psychiatr. 50, 651–658.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.
Siegle, G.J., Steinhauer, S.R., Thase, M.E., Stenger, V.A., Carter, C.S., 2002. Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. Biol. Psychiatr. 51, 693–707.
Siegle, G.J., Carter, C.S., Thase, M.E., 2006. Use of FMRI to predict recovery from unipolar depression with cognitive behavior therapy. Am. J. Psychiatr. 163, 735–738.
Siegle, G.J., Thompson, W., Carter, C.S., Steinhauer, S.R., Thase, M.E., 2007. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: related and independent features. Biol. Psychiatr. 61, 198–209.
Sladky, R., Friston, K.J., Tröstl, J., Cunnington, R., Moser, E., Windischberger, C., 2011. Slice-timing effects and their correction in functional MRI. Neuroimage 58, 588–594.
Sladky, R., Höflich, A., Atanelov, J., Kraus, C., Baldinger, P., Moser, E., Lanzenberger, R., Windischberger, C., 2012. Increased neural habituation in the amygdala and orbitofrontal cortex in social anxiety disorder revealed by fMRI. PloS One 7, e50050.
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. Eur. J. Radiol. 82, 728–733.
Sladky, R., Spies, M., Hoffmann, A., Kranz, G., Hummer, A., Gryglewski, G., Lanzenberger, R., Windischberger, C., Kasper, S., 2015. (S)-citalopram influences amygdala modulation in healthy subjects: a randomized placebo-controlled double-blind fMRI study using dynamic causal modeling. Neuroimage 108, 243–250.
Smith, S., Almagro, F., Miller, K., 2017. UK Biobank Brain Imaging Documentation. Version 1.3.
Spohrs, J., Bosch, J.E., Dommes, L., Beschoner, P., Stingl, J.C., Geiser, F., Schneider, K., Breitfeld, J., Viviani, R., 2018. Repeated fMRI in measuring the activation of the amygdala without habituation when viewing faces displaying negative emotions. PloS One 13, e0198244.
Straub, J., Plener, P., Sproeber, N., Sprenger, L., Koelch, M., Groen, G., Abler, B., 2015. Neural correlates of successful psychotherapy of depression in adolescents. J. Affect. Disord. 183, 239–246.
Swartz, J.R., Wiggins, J.L., Carrasco, M., Lord, C., Monk, C.S., 2013. Amygdala habituation and prefrontal functional connectivity in youth with autism spectrum disorders. J. Am. Acad. Child Adolesc. Psychiatry 52, 84–93.
Swartz, J.R., Williamson, D.E., Hariri, A.R., 2015. Developmental change in amygdala reactivity during adolescence: effects of family history of depression and stressful life events. Am. J. Psychiatr. 172, 276–283.
Szczepanik, J., Nugent, A.C., Drevets, W.C., Khanna, A., Zarate Jr., C.A., Furey, M.L., 2016. Amygdala response to explicit sad face stimuli at baseline predicts antidepressant treatment response to scopolamine in major depressive disorder. Psychiatr. Res. Neuroimaging 254, 67–73.
Triantafyllou, C., Hoge, R., Krueger, G., Wiggins, C., Potthast, A., Wiggins, G., Wald, L., 2005. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. Neuroimage 26, 243–250.
Ugurbil, K., Adriany, G., Andersen, P., Chen, W., Garwood, M., Gruetter, R., Henry, P.G., Kim, S.G., Lieu, H., Tkac, I., Vaughan, T., Van De Moortele, P.F., Yacoub, E., Zhu, X.H., 2003. Ultrahigh field magnetic resonance imaging and spectroscopy. Magn. Reson. Imaging 21, 1263–1281.
van den Bulk, B.G., Somerville, L.H., van Hoof, M.-J., van Lang, N.D., van der Wee, N.J., Crone, E.A., Vermeiren, R.R., 2016. Amygdala habituation to emotional faces in adolescents with internalizing disorders, adolescents with childhood sexual abuse related PTSD and healthy adolescents. Dev. Cognit. Neurosci. 21, 15–25.
Van Wingen, G., Geuze, E., Vermetten, E., Fernandez, G., 2012. The neural consequences of combat stress: long-term follow-up. Mol. Psychiatr. 17, 116.
Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellanos, F.X., 2015. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122, 222–232.
Waraczynski, M., 2016. Toward a systems-oriented approach to the role of the extended amygdala in adaptive responding. Neurosci. Biobehav. Rev. 68, 177–194.
White, L.K., Britton, J.C., Sequeira, S., Ronkin, E.G., Chen, G., Bar-Haim, Y., Shechner, T., Ernst, M., Fox, N.A., Leibenluft, E., 2016. Behavioral and neural stability of attention bias to threat in healthy adolescents. Neuroimage 136, 84–93.
Williams, L.E., Blackford, J.U., Luksik, A., Gauthier, I., Heckers, S., 2013. Reduced habituation in patients with schizophrenia. Schizophr. Res. 151, 124–132.
Yoo, P.E., John, S.E., Farquharson, S., Cleary, J.O., Wong, Y.T., Ng, A., Mulcahy, C.B., Grayden, D.B., Ordidge, R.J., Opie, N.L., 2018. 7T-fMRI: faster temporal resolution yields optimal BOLD sensitivity for functional network imaging specifically at high spatial resolution. Neuroimage 164, 214–229.
Young, K.D., Siegle, G.J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2017. Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. Am. J. Psychiatr. 174, 748–755.