Classifying the Wandering Mind: Revealing the Affective Content of Thoughts during Task-Free Rest Periods

Anita Tusche, Jonathan Smallwood, Boris C. Bernhardt, Tania Singer

PII: S1053-8119(14)00237-7
DOI: doi: 10.1016/j.neuroimage.2014.03.076
Reference: YNIMG 11252

To appear in: *NeuroImage*

Accepted date: 27 March 2014

Please cite this article as: Tusche, Anita, Smallwood, Jonathan, Bernhardt, Boris C., Singer, Tania, Classifying the Wandering Mind: Revealing the Affective Content of Thoughts during Task-Free Rest Periods, *NeuroImage* (2014), doi: 10.1016/j.neuroimage.2014.03.076

Classifying the Wandering Mind:

Revealing the Affective Content of Thoughts during Task-Free Rest Periods

Anita Tusche [1] *, Jonathan Smallwood [1, 2], Boris C. Bernhardt [1] & Tania Singer [1]

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Department of Social Neuroscience, Stephanstraße 1A, 04103 Leipzig, Germany

[2]University of York, Department of Psychology, Heslington, York, YO10 5DD, United Kingdom

Running title: Classifying the wandering mind

*** Corresponding author**

Anita Tusche, PhD

Max Planck Institute for Human Cognitive and Brain Sciences,

Department of Social Neuroscience

Stephanstrasse 1a, 04301 Leipzig, Germany

email: atusche@cbs.mpg.de

Phone: +(49) 341 9940 2690

Fax: +(49) 341 9940 2356

Abstract

Many powerful human emotional thoughts are generated in the absence of a precipitating event in the environment. Here, we tested whether we can decode the valence of internally driven, self-generated thoughts during task-free rest based on neural similarities with task-related affective mental states. We acquired functional magnetic resonance imaging (fMRI) data while participants generated positive and negative thoughts as part of an attribution task (Session A) and while they reported the occurrence of comparable mental states during task-free rest periods (Session B). With the use of multivariate pattern analyses (MVPA), we identified response patterns in the medial orbitofrontal cortex (mOFC) that encode the affective content of thoughts that are generated in response to an external experimental cue. Importantly, these task driven response patterns reliably predicted the occurrence of affective thoughts generated during unconstrained rest periods recorded one week apart. This demonstrates that at least certain elements of task-cued and task-free affective experiences rely on a common neural code. Furthermore, our findings reveal the role that the mOFC plays in determining the affective tone of unconstrained thoughts. More generally, our results suggest that MVPA is an important methodological tool for attempts to understand unguided subject driven mental states such as mind-wandering and daydreaming based on neural similarities with task-based experiences.

Keywords: fMRI, medial orbitofrontal cortex (mOFC), mind-wandering, multivariate pattern classification, resting-state, self-generated mental states

Highlights

- Task-cued emotional brain responses predict affective thought during task-free rest
- Elements of task-cued & task-free affective experiences rely on common neural code
- mOFC as key structure in determining the affective tone of unconstrained thoughts
- Task-based MVPA as a tool to study unguided mental states such as mind-wandering

Introduction

Although thoughts and feelings are often elicited by events in the here and now, not all experiences are generated from incoming sensory information. Indeed, many powerful human experiences (such as grief, anger or joy) can occur in the absence of any direct external referent. These subject driven emotional and cognitive states that are unrelated to events in the here and now are a core element of the human condition: They make up half of our waking thought (Killingsworth and Gilbert, 2010; Smallwood and Schooler, 2006) and are implicated in our emotional lives because of their robust links to state and trait unhappiness (Killingsworth and Gilbert, 2010; Smallwood et al., 2009; Smallwood and O'Connor, 2011; Smallwood et al., 2007; Watts et al., 1988).

As subject driven, self-generated mental states are a major element of our lives, it is surprising that we yet lack a rigorous understanding of how they occur and are processed in the brain. Given the evidence for their role for our emotional lives and happiness, the present study aimed to understand the neural basis of the affective dimension of subject driven mental states. Single cell recording studies in non-human primates as well as brain imaging studies in humans have frequently implicated the medial orbital frontal cortex (mOFC) and the adjacent ventromedial prefrontal cortex (vmPFC) in the processing of the affective valence and value of a wide-range of stimuli (Berridge and Kringelbach, 2013; Grabenhorst and Rolls, 2011; Kringelbach, 2013; Lebreton et al., 2009; Roy et al., 2012; Schoenbaum et al., 2011; Wallis, 2007; Wilson-Mendenhall et al., 2013). The mOFC is considered to be a central node in the brain's emotional circuits (Roy et al., 2012) and damage or dysfunction in this region has been linked to altered emotional responses and impaired emotion regulation (Bechara et al., 2000; Davidson et al., 2000; Izquierdo et al., 2005; Rolls et al., 1994), also see (Etkin et al., 2011; Etkin and Wager, 2007). Moreover, it has been proposed that the mOFC and adjacent portions of the vmPFC are involved in the

generation of affective meaning (Roy et al., 2012). In particular, this region was suggested to act as a hub that connects various systems such as sensory systems, interoceptive signals, long-term memory, and social cognition that contribute to the representation of conceptual information relevant to our well being and future prospects and the generation of affective responses. Based on this evidence, we hypothesized that the mOFC may play a general role in processing subject driven affective mental states that are generated in the absence of an external referent.

One difficulty in assessing the neural basis of internally driven affective and cognitive states is that their independence from external stimuli, as well as their spontaneous occurrence, makes these experiences hard to study using the classical experimental paradigms of cognitive science (Smallwood, 2013a, b). The present study explored whether we can take advantage of multivariate pattern analysis (MVPA) (Haxby et al., 2001; Haynes et al., 2007; Kriegeskorte et al., 2006; Norman et al., 2006), a machine-learning framework that that has been used to demonstrate that conceptually similar *task-related* mental states can be determined based on the spatial similarities within the evoked neural signals (Corradi-Dell'Acqua et al., 2011; Kahnt et al., 2011; Kassam et al., 2013; Lewis-Peacock and Postle, 2008; Poldrack et al., 2009). Some views of internally driven, task-unrelated thoughts have suggested they reflect fundamentally unique neural processes, with an extreme example being the notion of task-positive and task negative networks (Fox et al, 2005; see Spreng 2012 for a review of this issue). Recent neuroimaging work has provided a more nuanced perspective: There are, for example, spatial similarities between the neural networks engaged by external tasks and those revealed during periods of unconstrained rest (Smith et al., 2009). Such evidence suggests that subject driven thought may not differ from states elicited by external events in the neural processes they recruit, but rather may be unique in the manner that they are initiated (Smallwood, 2013a, b). The present study capitalized on

this assumption by using MVPA to overcome fundamental difficulties in determining the occurrence of subject driven mental states by decoding the affective content of unconstrained thought based on its neural similarities with emotional experiences that are generated as part of a well controlled experimental task.

In the current study, functional magnetic resonance imaging (fMRI) was acquired from healthy participants who completed two independent experimental sessions, separated by a one-week interval with the session order counterbalanced across subjects. In Session A, participants generated positive and negative emotional mental states in response to external cues in an attribution task (Figure 1A). In Session B, participants were allowed to rest and thought sampling was used to examine the valence of spontaneously occurring self-generated thought (Figure 1B). We expected brain responses in areas such as the mOFC to provide information on the affective tone of thoughts regardless of whether they occur as a consequence of an external task, or during unguided task-free processing. Using MVPA, we tested whether the neural representations of affective mental states elicited as part of a task could be used to decode the content of emotional experiences that are generated during periods of task-free rest. Finally we asked a subset of these participants to undergo a standard resting-state scan (Session C) that took place approximately three month apart from Session A and Session B. This last step allowed us to explore whether brain responses in the mOFC reflected participants tendency to self-generate positive thoughts in a context that had no external task nor was disrupted by thought sampling.

Materials and Methods

Participants

Thirty healthy volunteers (15 female, aged between 21 and 33 years, mean ± SD: 26 ± 3 years) participated in two fMRI sessions (Session A and B) and a behavioral posttest. A subset of 20 participants also took part in a separate task-free resting-state fMRI scan (Session C) (average time lag between Session B and Session C: 97 days ± 38 days SEM). Participants were German native speakers, right-handed, free of psychiatric or neurological history, and had normal or corrected-to-normal vision. They were paid €8 per hour for their participation in the experiment. Data of three participants had to be excluded because of head movement beyond 3 mm/3 degrees during scanning sessions A and B. Moreover, functional data of two participants of Session A and of one participant in Session B had to be discarded due to head movements of more than 3 mm within the session. No data had to be discarded from resting-state Session C. Written informed consent was obtained from all participants, and the local ethics committee approved our study.

Task

Participants took part in two scanning sessions (Session A and B) separated by a 1-week interval, with the session order counterbalanced across subjects. A subset of subjects participated in an independent resting-state scan (Session C) that was approximately three months apart from Session B.

Session A. To obtain task-related brain responses, participants performed blocks of a self-referential attribution task (based on (Addis et al., 2007)) in which they were consecutively presented with three trait adjectives that were either positive (positive attribution blocks) or negative (negative attribution

blocks) (Figure 1A). Trait adjectives were selected from previously published stimuli (Herbert et al., 2006) based on behavioral pilot studies using independent samples. Each of the 48 trait adjectives (24 positive, 24 negative) was presented once for the duration of 9 s and was the target of a self-referential attribution where participants were asked to imagine themselves being as described by the visual cue. Participants indicated the successful generation of a vivid attribution via a button press and were asked to maintain and to elaborate on this attribution for the remainder of the trial. Trait adjectives within a task block were separated by variable delays of 4 to 8 s during which a black fixation cross against a white background was shown. The presentation order of trait adjectives was randomized within and across task blocks. In each of the 8 runs in Session A, participants performed one positive and one negative attribution block, with the block order counterbalanced across runs and randomized across participants. Task blocks were followed by intervals of 30 to 90 s during which participants were asked to fixate on a centrally presented fixation cross. Subsequent to the scanning session, participants rated the valence of each attribution in a self-paced computerized task outside of the scanner to confirm results of the independent behavioral pretests.

Session B. To obtain functional data during unconstrained affective thought, participants took part in task-free rest periods (6 runs of 9 min each) during which they were asked to fixate on a black fixation cross centrally presented against a white background. At unpredictable intervals (minimum of 30 s), rest periods were intermitted and thought samples were obtained to asses participants types of thoughts (Figure 1B). Using a continuous rating scale (ranging from -3 to 3), participants first reported the affective valence of their thoughts prior to the respective thought sample. Self-reported valence (z-scored) was used to identify 'positive' or 'negative' thoughts in Session B. In addition, thought probes were used to assess whether participants' thoughts were related to something in the past or future and

to the self or others. Accounting for the possibility that thoughts may involve some elements relating to both anchors of the scale, participants were instructed to report the primary affective tone (positive or negative) and the focus (self or others; past or future) of the preceding thought. Each of the three rating scales that were presented in a fixed order was displayed for 8 s. Participants responded by pressing the left and right buttons of a button box that was placed in their right hand to move the courser from its randomized initial position. In total, we collected 36 thought samples in Session B for each participant. Please note that a smaller number of thought samples (24) were also collected in rest intervals between attribution blocks of Session A (minimum duration of rest intervals = 30 s) to assess intra-individual stability of the self-reported content of mental states during task-free periods. Functional runs of both scanning sessions were divided by breaks of approximately 1 min in which no brain responses were acquired. Presentation 14.9 (http://www.neurobs.com) was used for stimulus presentation and data collection.

Session C. An additional standard resting-state scan that was undisturbed by thought sampling was obtained in an independent Session C. Participants were instructed to keep their eyes open, to fixate a centrally presented white cross against a black background and to think of nothing in particular.

[insert Figure 1]

Functional image acquisition

Functional imaging was performed on a 3T Verio scanner (Siemens Medical Systems, Erlangen) equipped with a 12-channel head coil. T2*-weighted functional images were obtained using an echoplanar

imaging (EPI) sequence (TR = 2 s, TE = 27 ms, flip angle = 90°, 3 x 3 x 3 mm, 1 mm interslice gap, matrix size 70 x 70, 37 slices tilted at approximately 30° from axial orientation, ipat = 2). In each of the 8 runs of Session A , 225 volumes were acquired. In Session B, 258 volumes were measured for each of the 6 runs. Functional runs of both scanning sessions were divided by breaks of approximately 1 min when no functional images were obtained. Subsequent to the functional imaging of Session B, a T1-weigthed high-resolution anatomical image was collected using a MPRAGE sequence (TR = 2.3 s, TE = 2.98 ms, flip angle = 9°, 1 x 1 x 1 mm, matrix size 240 x 256, 176 sagittal slices, ipat = 2) using a 32-channel head coil. In an independent scanning Session C, we also obtained task-free fMRI time series for a subset of 20 participants. For this resting-state data, 150 volumes were acquired over 5 min using the same sequence as the task-related fMRI data. During Session C, participants were instructed to keep their eyes open, to fixate a centrally presented white cross against a black background and to think of nothing in particular.

fMRI Data analysis

Functional images of all scanning sessions were analyzed using the statistical parametric mapping software SPM8 (http://www.fil.ion.ucl.ac.uk/spm) implemented in Matlab. For each data set, preprocessing consisted of slice-time correction, spatial realignment and normalization to the Montreal Neurological Institute (MNI) brain template (voxels were resampled to 2 x 2 x 2 mm), and spatial smoothing using a Gaussian kernel of 8 mm FWHM. Preprocessed data of Session A and B were analyzed using a general linear model (GLM) as implemented in SPM8 (Friston et al., 1995). For every run, neural activation was modeled by distinct regressors convolved with a canonic hemodynamic response function (hrf). A 128 s high-pass cutoff filter was applied to eliminate low-frequency drifts in the data of Session A and B independently. For each participant, we estimated multiple GLMs.

*GLM1***.** To identify brain regions that encode the affective content of *externally cued, task-related thoughts* in Session A, we used a GLM that estimated two regressors of positive (R1) and negative attributions (R2) for each run, with the duration equal to the time period from the cue onset (trait adjective) to participants' button press for that trial. The remainder of the trial was modeled by two additional regressors, for positive and negative attributions separately (R3, R4). Attributions were defined as 'positive' and 'negative' based on valence ratings obtained in behavioral pretests using independent samples. In the rare cases (average of 2.58% of trials ± 0.49 SEM; Supplemental Figure 1) where participant's self-reported valence (obtained after Session A) differed from this a-priori assignment, attributions were assigned based on participant's stated affective experience. The GLM also modeled the rest periods between task blocks (30 to 90 s, R5), the three rating periods for the limited number of thought probe samples in Session A (8 s each, R6-R8), and three parametric regressors corresponding to self-reports on these scales (z-scored per rating scale, R9-R11) as regressors of no interest. Parameter estimates of the externally cued, task-related positive and negative attributions (R1, R2) were then applied to multivariate pattern analyses (MVPA, see below). Please note that GLM1 estimated regressors for each of the 8 functional runs, ensuring that the number of regressors applied to MVPA was balanced for positive and negative conditions.

GLM2. To identify brain regions that encode the affective content of *unconstrained self-generated thoughts* during task-free rest in Session B, we used a GLM that estimated two regressors corresponding to rest periods including positive (R1) and negative thoughts (R2) for each run. Both regressors were modeled using a fixed duration of 30 s prior to the respective thought sample, corresponding to the minimum duration of undisturbed rest between thought samples. Assignment of rest periods was based on self-reported valence in the thought probes (z-scored). Similar to GLM1, the model included the

three rating periods of the thought probe samples (8 s each, R3-R5) and two parametric regressors of the corresponding ratings (z-scored) that were not used to sort the prior rest periods (i.e., 'past-future', 'self-other', R6, R7) as coregressors of no interest. Parameter estimates of positive and negative thoughts during task-free rest (R1, R2) were then applied to MVPA (see below). Similar to GLM1, the number of regressors for positive and negative thoughts was balanced to prevent biases in the classification approach.

In addition to MVPA, we also tested for conventional univariate effects of affective valence for task-related attributions (GLM1, Session A) and thoughts during task-free rest (GLM2, Session B). For each participant, we computed contrast statistics against baseline for regressors of interest (R1, R2; for GLM1 and GLM2, respectively). Individual contrast images of the respective GLM were then used to compute random-effects group analysis using paired t-tests as implemented in SPM8. Significant clusters were identified using a statistical threshold of $p < 0.05$, FDR corrected for multiple comparisons at the cluster level (height threshold of $p < 0.005$).

Multivariate pattern analyses (MVPA).

Whole-brain MVPA searchlight decoding within Session A. In a first step, a 'searchlight' decoding approach was used to identify brain regions that predict the valence of externally cued, task-related thoughts in Session A (Figure 1A). This multivariate pattern classification approach does not depend on a priori assumptions about informative brain regions or prior voxel selection, avoiding the problem of circular analysis (or 'double dipping') (Kriegeskorte et al., 2009) and ensuring an unbiased analysis of neural activation patterns throughout the whole brain (Haynes et al., 2007; Kriegeskorte et al., 2006).

For each participant, a sphere with a radius of 3 voxels (Bode et al., 2011; Soon et al., 2008; Soon et al., 2013; Weygandt et al., 2012) was defined around a given voxel v_i of the measured volume. For every run of Session A (total of 8 runs), parameter estimates (R1 and R2, GLM1) were extracted for each of the N voxels within this spherical cluster and transformed in an N-dimensional pattern vector. Pattern vectors were created separately for positive and negative attributions. Pattern vectors of all runs but one ('training dataset') were then used to train a linear support vector machine classifier (SVM, http://www.csie.ntu.edu.tw/~cjlin/libsvm) using a fixed cost parameter C = 1. This provided the basis of the subsequent classification of the pattern vectors of the remaining run ('test dataset') as representing either positive or negative attributions. The procedure was repeated several times, always using a different run as test dataset to achieve a robust run-wise cross-validation (i.e.., 8-fold cross-validation for 8 runs of Session A). For each spherical cluster, the amount of predictive information on the valence of task-related thoughts was represented by the average percentage of correct classifications across all cross-validation steps and was assigned to the central voxel v_i of the sphere. This support vector classification (SVC) was successively carried out for all spherical clusters created around every measured voxel. Thereby, we obtained a three-dimensional map of average classification accuracies for each participant. Individual accuracy maps were then entered into a random-effects group analysis using a simple t-test as implemented in SPM8. This allowed identifying brain regions that reliably encoded the affective content of task-related mental states across participants. Classification was based on two alternatives ('positive' versus 'negative' attributions), resulting in a chance level of 50%. Only regions showing significant decoding accuracies above chance and passing the statistical threshold of $p < 0.05$ (FDR corrected for multiple comparisons at the cluster level, height threshold of $p < 0.005$) were considered relevant for information encoding. To test for biases in the classification, an additional control analysis was implemented that matched the described decoding analyses except for assigning randomly selected, permuted labels to the data (Tusche et al., 2010).

ROI-based MVPA within Session B. Next, we tested whether similar brain regions encode the valence of mental states during task-related attributions (Session A) and task-free rest periods (Session B). To address this issue, brain regions that predicted the valence of thoughts in Session A were defined as regions of interest (ROIs) for the decoding of task-unrelated positive and negative thoughts during rest periods in Session B. Thus, we used results of an independent scanning Session A to select features for the classification of the affective content of thoughts in Session B. Each ROI was defined by a 10 mm sphere around the statistical peak of the predictive cluster (Table 1), using the MarsBaR toolbox (see Figure 2D for an example). For each participant, we then investigated whether local activation patterns within a particular ROI (extracted from GLM2, R1 and R2) predict the valence of mental states during rest periods, as reported in subsequent thought probe samples. Except for the selection of activation patterns that was defined by the ROI, the classification procedure was identical to the searchlight decoding approach described above. For each participant, the ROI-specific predictive information on the valence of thoughts during task-free rest was represented by the average percentage of correct classification across the run-wise cross-validation steps. To assess the statistical significance of the predictive information across participants, we submitted the average predictive accuracy of each ROI to a t-test against the chance level of 50% as implemented in Matlab R2013a.

ROI-based MVPA across Session A and B. In a next step, we tested whether activation patterns obtained during externally cued, task-related thoughts in Session A can be used to reliably predict the affective content of unconstrained thoughts during rest periods in Session B. For each participant, we trained a classifier on ROI-specific activation patterns of all runs of Session A (‘positive’ versus ‘negative’ attributions, GLM1), and tested the classifier on ROI-specific activation patterns of one run of Session B

('positive' versus 'negative' thoughts during rest, GLM2). Please note that for the cross-session classification, each activation pattern was normalized to ensure that predictions are based on information encoded in the distributed response patterns and do not merely reflect average signal differences across conditions (as identified for positive and negative attributions in Session A). More precisely, we removed the mean response (across all voxels within the pattern) from each voxel and divided the resulting values by the standard deviation for that response pattern (Misaki et al., 2010). Importantly, this across-voxel normalization preserves the shape of the response patterns (Misaki et al., 2010). For each participant, we estimated the average ROI-specific classification accuracy (run-wise cross-validation) and submitted it to a t-test to assess the statistical significance of the predictive information across participants.

Session C (fMRI resting-state data) - Degree centrality.

In addition to our functional data from Session A and B, for a subset of 20 participants, we also acquired resting-state fMRI data in an independent Session C in which no thought probes occurred during task-free rest. Session C data were used to assess the relationship between functional network integration of the mOFC and individual differences in the affective tone of self-generated mental states during rest in Session B. Importantly, functional data undisturbed by thought sampling enabled us to explicitly test whether associations of the mOFC with affective processing of self-generated thought were an artefact of thought sampling, for instance due to metacognition, response preparation or biases in self-reports. We computed the degree centrality of the preprocessed resting-state data as a robust marker of functional network integration of a brain region (Buckner et al., 2009; Zuo et al., 2012). Resting-state data was preprocessed using the data processing assistant for resting-state fMRI (Song et al., 2011; Yan and Zang, 2010) (http://www.restfmri.net). Preprocessing parameters were chosen to closely

correspond to those for the processing of Session A and B, with the exception that data was corrected for nuisance signals (i.e., average signal of cerebro-spinal fluid and white matter, and motion parameters) and band-pass filtered to be within 0.01 and 0.1 Hz (Buckner et al., 2009; Zuo et al., 2012). Based on the cross-correlation matrix between the time series of all voxels in the brain, we computed the weighted degree centrality of every voxel to quantify the relevance of each particular voxel for the whole brain network. We followed previous recommendations and thresholded the correlation matrix at $r > 0.25$ prior to estimating the centrality (Buckner et al., 2009). For each participant, centrality data at each voxel were normalized using a z-transform that took into account centrality data across the entire brain (Buckner et al., 2009). Statistical analysis of centrality data was carried out using SurfStat (Worsley et al., 2009) (http://www.math.mcgill.ca/keith/surfstat/). We assessed the relationship between participants' average experienced valence of thought during task-free rest periods in Session B (as reported in the thought samples) and average degree centrality of the mOFC-ROI based on Session C. We also performed a supplemental whole-brain correlation analysis between participants' average valence of thoughts during task-free rest periods in Session B and voxel-wise degree centrality in Session C (thresholded at $p < 0.005$, uncorrected, $k > 20$ voxels).

Results

Behavioral

Attributions. Average attribution times were comparable for positive (mean ± SD: 4.00 ± 1.57 s) and negative task conditions (4.14 ± 1.54 s) (paired t-test, $p > 0.05$, $t = -1.55$), indicating that task-performance was comparable for both attribution blocks. Attribution-specific valence judgments (Likert scale, range: -3 to +3) obtained after Session A confirmed that attributions in positive task blocks were

experienced as more positive (1.98 ± 0.42) than attributions in negative task blocks (-1.73 ± 0.46; paired t-test, p = 0.001; t = 21.73; Figure 2A).

Thought samples. The average valence of task-related attributions in Session A (0.32 ± 0.26) was positively correlated with the average valence of thoughts during task-free rest periods in Session B (0.77 ± 0.52; Pearson's r = 0.34, p = 0.03, one-tailed), suggesting that individual tendencies to engage in positive or negative thoughts generalized across task-related and unconstrained affective thoughts (Figure 2B). No relationship of mean attribution valence was found with participant's tendency to think about the past or the future (0.32 ± 0.46; i.e. slightly future-oriented) or others or oneself (0.23 ± 0.53, i.e. slightly self-oriented) as assessed using thought samples in Session B (all p > 0.05). Within Session B, average ratings per thought sample scale were not significantly correlated (all p > 0.05). Subject-wise correlations of thought probe ratings and initial positions of the ratings confirmed that randomized initial positions of the scales did not cause response biases in Session B (all p > 0.05).

FMRI

Decoding the valence of task-related attributions in Session A.

A whole-brain searchlight decoding approach was used to identify brain regions that encode the valence of externally cued, task-related thought in Session A. Local activation patterns in the medial orbitofrontal cortex (mOFC; Figure 2C), anterior cingulate cortex (ACC), the posterior cingulate cortex (PCC), middle temporal gyrus and middle frontal gyrus were found to reliably predict the valence of self-referential attributions (p < 0.05, corrected for multiple comparisons at cluster level; see Table 1 for details and complete list of results). Supplemental permutation analyses confirmed that response

patterns in all clusters yielded classification accuracies that are highly unlikely to be obtained by chance only (Supplemental Table 1). Please note that the accuracy level of the prediction (see Table 1) is comparable to previous studies that applied a linear support vector machine classifier with a fixed cost parameter (C = 1) as used in our in a searchlight decoding approach (Bode et al., 2013; Bode and Haynes, 2009; Hampton and O'Doherty, 2007; Heinzle et al., 2012; Reverberi et al., 2012; Soon et al., 2008; Tusche et al., 2010) (see supplemental material for summary statistics of decoding accuracies).

[insert Table 1]

Supplemental univariate analysis identified clusters in the bilateral mOFC (peak at [MNI -6, 44, -16], t = 5.96), PCC (peak at [MNI -6, -62, 18], t = 4.56), ventral striatum (peak at [MNI 18, 6, 6], t = 5.24 and [MNI -18, 4, -4], t = 4.92), right cerebellum (peak at [MNI 8, -74, -30], t = 5.67) and left posterior superior temporal gyrus (peak at [MNI -48, -60, 22], t = 5.26) that were significantly more strongly activated for positive than for negative attributions ($p < 0.05$, corrected for multiple comparisons), indicating that predictive information in activation patterns of the mOFC and the PCC might be partly due to signal increase in these areas during positive attributions. The reverse contrast [negative > positive attributions] did not yield significant results. A supplemental parametric modulation analysis confirmed that univariate brain responses in the bilateral mOFC (peak at [MNI 14, -46, -16], t = 4.58) reflected the affective tone of task-related thoughts in Session A (see Supplemental Figure 2 for an illustration and a complete list of results).

Decoding the valence of thoughts during task-free rest periods in Session B.

ROI-based prediction. In a next step, we investigated whether *similar brain regions* encode the valence of thought during both externally cued attributions and unconstrained, task-free rest periods, respectively. To formally address this issue, we defined brain regions that predicted the valence of attributions in Session A (Table 1) as regions of interest (ROIs). When these ROI-based decoding analyses were applied to the data from Session B (*within*-Session B decoding), we found that distributed activation patterns in the mOFC (Figure 2D) provided information about positive and negative thoughts during rest (average of 60% decoding accuracy across participants, ± 2.99 SEM, $p = 0.003$). No effect of session order was found ($p = 0.71$; Figure 2E). Local response patterns in the ACC were also predictive of participant's task-free positive and negative thoughts prior to the respective thought sample (average of 57% decoding accuracy, ± 2.43 SEM, $p = 0.009$). No other ROI that encoded task-related positively and negatively valenced thoughts in Session A was found to predict the affective tone of mental states in Session B significantly above chance level (all $p > 0.17$).

Having established that the mOFC and the ACC encode the affective content in Session A and B respectively, we then tested whether task-based response patterns (Session A) can be used to predict the occurrence of emotionally valenced thoughts during unconstrained, task-free rest periods (Session B). More precisely, this cross-session decoding approach explicitly examined whether *similar neural codes* underlie both experiences and, hence, whether such a neural similarity can be used to identify the content of unconstrained thoughts during task-free rest. This cross-session classification revealed that individual response patterns in the mOFC obtained during attributions predicted the valence of participants' self-generated thoughts in Session B (average of 57% decoding accuracy, ± 2.94 SEM, $p = 0.030$), independent of session order ($p = 0.98$). No other ROI exhibited patterns that were predictive across sessions (all $p > 0.11$). To provide further evidence for the statistical significance of the mOFC-

based prediction we used permutation tests. To this end, we repeated the above decoding approach 1000 times with randomly assigned test labels (i.e. 'positive' versus 'negative') to the test data. This resulted in a permutation distribution of average decoding accuracies that would be achieved by chance only (i.e. a null distribution). The mOFC-based prediction of 57% (across run-wise cross-validation and across participants) corresponded to the 99.6 percentile of the permutation distribution, corresponding to a probability of alpha = 0.004 that the above prediction was realized by chance. Due to computational demands, supplemental permutation test was restricted to ROI-based analysis. Supplemental permutation tests also confirmed that clusters reported in Table 1 and Table 2 yielded classification accuracies that are unlikely to be obtained by chance (all probability of alpha < 0.05).

To ascertain the specificity of the information contained in the mOFC for the affective content of thoughts, we estimated two additional GLMs that were identical to GLM2 ('positive' or 'negative'), except that the rest periods prior to the thought samples in Session B were sorted based on their ratings of 'future' or 'past' (GLM3) and 'self' or 'other' (GLM4). ROI-based predictions on the activation patterns extracted from the parameter estimates of the respective GLM did not yield results for any of the ROIs (mOFC [past-future]: average of 49% decoding accuracy, p = 0.48; mOFC [self-other]: average of 51% decoding accuracy, p = 0.81; all other p > 0.16). Thus the mOFC activation pattern is specific to the emotional component of thought. Whole brain searchlight decoding on self-directed versus other directed thoughts in Session B (as reported in the thought samples) confirmed that a more dorsal cluster in the mPFC (peak at [MNI 26, 38, 14]) encoded the self-directedness of thoughts (Supplemental Figure 3).

[insert Figure 2]

In addition, we investigated whether task-based response patterns in the mOFC-ROI (Session A) encode the experienced valence of thoughts during task-free rest periods between attribution blocks (as reported in the thought samples obtained in Session A). Similar to GLM2, we re-estimated GLM1 with 2 regressors corresponding to the 30 s rest interval prior to 'positive' and 'negative' thought samples. As expected, task-related response patterns in the mOFC predicted the valence of thoughts between the task-blocks significantly above chance level (average of 56% decoding accuracy, $p = 0.026$). This provides a replication of the overlap between task driven and subject driven self-generated thoughts. However, due to the close temporal proximity of these thought probes to the attribution task, the paper focused on the task-based prediction of thought samples obtained in Session B.

Whole brain analyses. To provide further evidence for the role of the mOFC in unconstrained, self-generated positive and negative thoughts in Session B, we performed a whole brain searchlight decoding for each participant. Confirming the ROI-based decoding results (within-Session B classification), local activation patterns in the mOFC and the ACC were found to predict the valence of spontaneously generated thought during task-free rest periods ($p < 0.05$, corrected for multiple comparisons). See Table 2 for details. Supplemental permutation analyses confirmed that response patterns in all predictive clusters yielded classification accuracies that are highly unlikely to be obtained by chance only (Supplemental Table 1). Importantly, there was significant spatial overlap between the predictive clusters based on the task-related ROI (within Session A) and task-free thoughts (within Session B) as illustrated in Figure 3. Despite this spatial convergence of the predictive mOFC cluster for task-based and task-free mental states, a conventional univariate analysis testing for differences in activation between rest periods in Session B that involved positive and negative thoughts (GLM2) failed

to find significant results ($p < 0.05$, corrected). Similarly, mass-univariate average BOLD responses (percentage signal change) for the mOFC-ROI did not differ for positive versus negative rest periods (all $p > 0.68$). These findings indicate that the sensitivity of MVPA makes it a powerful tool to investigate subject driven mental states such as mind-wandering or daydreaming.

[insert Table 2]

[insert Figure 3]

Resting-state fMRI.

To further specify the functional role of the mOFC in internally driven affective processing, we explored whether individual differences in the degree of functional network integration of the mOFC during *undisturbed* rest (i.e., neither by tasks nor thought probes) relate to participants average valence of unguided thoughts in Session B. This analysis is important because the technique of thought sampling could potentially disrupt the 'normal' flow of self-generated experiences at rest. Thus, the independent resting-state fMRI session allowed examining variations in how affectively toned self-generated thought are associated with neural processes in uninterrupted rest. We acquired resting-state fMRI data from a subset of 20 participants in an independent Session C that was undisturbed by thought probe sampling. Using a ROI-based analysis (see Figure 2D for illustration of ROI in mOFC), we found a positive correlation between the average degree centrality of the mOFC and participants' average experienced valence in Session B ($p < 0.025$, $t = 2.51$; with a corresponding correlation coefficient of $r = 0.51$; Figure 2F). In other words, the mOFC exhibited greater hub like properties during a period of undisturbed rest in participants who tended to engage in more positively valenced thought and feeling states.

A supplemental voxel-wise correlation for the whole brain confirmed that the degree of functional network integration of the right mOFC (peak at [MNI 12, 48, -12], t = 3.65, k = 23 voxels) and the bilateral PCC (peak at [MNI -6, -39, 9], t = 4.01, k = 68 voxels) in Session C was positively correlated with the affective content of unguided subject driven mental states during task-free rest in Session B ($p < 0.005$, k = 20 voxels, whole-brain uncorrected; please note that uncorrected results are reported to emphasize the convergent evidence and the relative specificity of the mOFC findings). No other clusters were identified. These findings confirm a central role of the mOFC in the neural network underlying subject driven affective processing during rest, at the level of individual differences.

Discussion

The present study tested whether we can decode the affective content of unguided, subject driven thoughts during task-free rest based on response patterns of conceptually related experiences obtained in an emotional attribution task. Using advanced multivariate pattern analysis (MVPA) (Haxby et al., 2001; Haynes et al., 2007; Kriegeskorte et al., 2006; Norman et al., 2006), we demonstrated that the valence of unconstrained affective thoughts can be reliably predicted based on neural activation patterns in the mOFC identified when people are externally cued to engage in similar affective mental states. We found similar co-variance at the subjective level: participants' mean experienced valence of task-cued and spontaneously generated thought was positively correlated over a one-week interval[1]. Moreover, using resting-state data from an independent Session C, functional connectivity measures (i.e., degree centrality) demonstrated that an individual's tendency to self generate positive thoughts was associated with the hub like properties of the mOFC during task-free rest. Relating individual differences in affective tone with patterns of intrinsic functional connectivity, this result provided further support for the functional role of mOFC in the affective processing of internally driven thought and feeling states. Given that the rest periods in Session C were undisturbed by thought probes, this finding indicates that the association of the mOFC with the affective tone of unconstrained mental processes is not an artifact of the method of thought sampling and due to processes such as response preparation for the impending thought probe, metacognition or response biases in self-reports. Altogether these results can be parsimoniously accounted for by the hypothesis that there are important neural processes, encoded by the mOFC, that are common to task-cued and spontaneous, unguided affective states. This evidence for shared neural codes across both experiences forms the basis

[1] The rationale for expecting an association between the valence ratings in Session A and Session B is the same rationale as for our neural prediction. Just as we hypothesized that neural response patterns from task driven and subject driven examples of the same state are predictive, we also assumed that inducing cognitive states within an individual sheds light on the way that the experience unfolds in unconstrained settings.

for future MVPA investigations into purely subject driven brain states based on neural similarities with task-based experiences.

Our results indicate that the functional role of the mOFC is not limited to the processing of affective valence of a wide range of stimuli that occur as part of a task, or reflect the regulation of externally cued emotional responses (Bechara et al., 2000; Berridge and Kringelbach, 2013; Davidson et al., 2000; Etkin et al., 2011; Etkin and Wager, 2007; Grabenhorst and Rolls, 2011; Izquierdo et al., 2005; Kringelbach, 2013; Rolls et al., 1994; Roy et al., 2012). Instead, we demonstrate that this region also reflects the affective content of thoughts that occur in the absence of an external referent. An extension of the central role of the mOFC to the processing of unguided, internally driven affective thoughts is consistent with the proposal that this region plays a domain general role in generating affective meaning (Roy et al., 2012). The counterbalanced order of the first two scanning sessions (Session A and B) rules out the possibility that the experiential and neural associations identified by MVPA merely reflect the recollection of the previously performed external attribution task. Likewise, trivial accounts for shared representations such as priming of the affective content of thoughts during rest by preceding attributions are unlikely due to the one-week delay between scanning sessions. Furthermore, because supplemental decoding analyses demonstrated that the shared neural representations of the mOFC were specific to the valence of participants' self-generated thoughts, we are confident that this region is primarily involved in the affective dimension of experiences like daydreaming or mind-wandering. More precisely, activation patterns in the mOFC obtained in an affective attribution task contained reliable information about the affective content of spontaneously self-generated thoughts, but not whether these thoughts were more related to self or others and future or past. Although episodes of self-generated thought are commonly regarded as a form of mental time travel (Stawarczyk et al., 2011),

they are also an emotional experience. Our data suggests that the mOFC likely plays a crucial role in this affective element of unconstrained thought.

In practical terms, our observation that affective experiences are mediated by the mOFC provides important clues into the specific mechanism through which cognition can influence happiness and ultimately psychological well-being in daily life. Given that internally driven, self-generated thoughts are more frequent in individuals with depression (Smallwood and O'Connor, 2011; Smallwood et al., 2007; Watts et al., 1988), chronic unhappiness could be reflected by changes in the manner through which the mOFC contributes to the emotional characteristics of conscious thought. More generally, because engaging in negative subject driven thoughts is linked to premature aging and stress (Epel et al., 2012), our results have implications for how thought can directly impact on health and well-being. We suspect that understanding the role of mOFC in internally driven thought and feeling states could ultimately shed an important light on the disruptive influence of such experiences on the health of many members of todays' society.

There are a number of limitations that should be borne in mind when considering the current results. One issue concerns the time window we used to explore the functional data from Session B. We employed a thirty second window as this was the longest time frame common to *all* thought sampling probes. Although a shorter window (Christoff et al., 2009; Hasenkamp et al., 2012) would provide a more targeted result, it would also provide less power in our analysis, and would also be chosen in an arbitrary fashion. The issue of the appropriate window size is a methodological challenge in the study of self-generated thought. The spontaneous occurrence of the episode means that we cannot reliably identify the onset – or the duration – for the experience (Smallwood, 2013a), at least not using thought

probes obtained with considerable time lags. This lack of a clear indicator of their onsets means that self-generated thoughts cannot be explored with the precision with which we can understand other aspects of cognition and affective experience. Yet it is important to recognize that the capacity to classify subject driven mental states using MVPA provides an important methodological advance that in the future may determine when these experiences begin. Thus, although the window size used in the current work is a limitation, our demonstration that MVPA can be used to classify self-generated affective thoughts will likely provide an independent approach that can help optimize the selection of window size in future studies investigating these states. It is also important to recognize the value of the self-reported information in understanding subject driven mental states. Although we were able to employ MVPA to predict the affective content of thought, these forms of supervised classification are valid in as much as the labels employed are themselves valid. At present, first-person experience can only be assessed using introspection and thus it would be inappropriate to conclude that our technique is equivalent to so-called mind reading. Instead our study demonstrates that MVPA could be subsequently employed to test, and understand, the complex varieties of self-generated thoughts that are experienced in daily life. In addition to the increased precision that MVPA provides (such as its potential capacity to determine the onset more accurately) it could also be used to test theories of self-generated thought. In the current study, we focused on examining the specific aspects of neural activity that are specific to affective thought. However, future work could also examine neural processes that may be common to different forms of experience. In our current study, for example, we used a single Likert scale reflecting a single dimension from self to other. Following up on behavioral evidence of a statistical decomposition of these two components (Ruby et al., 2013a; Ruby et al., 2013b), future studies might use separate scales to access shared and unique elements of self-related and other-related thoughts. Conceivably, MVPA could be easily used to provide a neural basis for identifying the

ontology of self-generated thought much as it has been proposed for other aspects of cognition such as cognitive control (Lenartowicz et al., 2010) or for psychiatric disorders (Bilder et al., 2009).

Our results make two general points regarding how to understand states of unconstrained self-generated thought such as daydreaming or mind-wandering. First, at least in the affective domain, greater care is needed when attempting to probe the differences between unconstrained subject driven experiences and mental states that are initiated in response to an external task. Traditionally, self-generated states of mind-wandering or daydreaming have been viewed as unique because of the supposition that they recruit unique neural substrates (Fox et al., 2005). However, our results suggest that this view is overly simplistic. At a process-level, both forms of experience can be understood by neural changes that occur in the mOFC, demonstrating that at least certain elements of both experiences rely on a common neural code. These two classes of experience are nonetheless fundamentally different: the occurrence of subject driven affective thought must be influenced directly by intrinsic changes in the brain, while similar mental states that arise as part of a task occur in response to an external perceptual cue (Smallwood, 2013a, b). Second, our data suggest that one reason why task-based and resting-state networks exhibit spatial overlap (Smith et al., 2009) is because at rest, participants could be actively engaged in cognitive and emotional processes that recruit the same neural process used by task-based paradigms. Although we focused on emotional processes in the current study, many brain networks exhibit similarities between task and rest (Smith et al., 2009). Based on our data, future research should aim to identify which networks exhibit coactivity between rest and task-related processing because of the cognitive operations that participants engage in during rest and those which do not.

Finally, our results demonstrate the efficiency of approaches like MVPA in decoding subject driven mental events in task-free resting-state sessions. While multi-voxel activation patterns in the mOFC encoded the affective content of unconstrained thoughts, conventional univariate contrasts failed to reveal differential neural effects during rest, perhaps because of the increased sensitivity of MVPA relative to mass-univariate approaches (Kragel et al., 2012; Kriegeskorte et al., 2007). MVPA has previously been applied successfully to investigate shared neural representations across experimental conditions for a variety of affective and cognitive tasks (Corradi-Dell'Acqua et al., 2011; Kahnt et al., 2011; Kassam et al., 2013; Lewis-Peacock and Postle, 2008; Poldrack et al., 2009). For example, activation patterns in the medial prefrontal cortex and left superior temporal sulcus were found to encode basic affective states such as happiness, anger or fear across different sensory stimulus modalities (Peelen et al., 2010). These emotion-specific representations in the brain were found when participants were asked to make judgments about the emotional intensity conveyed by the stimuli (i.e., pictures of faces, body movements or voices). The present study extends previous findings by showing that neural activation patterns that encode the affective content of externally cued, task-related mental states are reinstated when these experiences are spontaneously self-generated in the absence of an environmental cue. Thus, MVPA allowed us to exploit the spatial similarity between task-related and unconstrained neural activity and to reliably predict the affective content of unguided subject driven mental states. As understanding intrinsic activity in the brain is a key aim of cognitive neuroscience (Zhang and Raichle, 2010), our data suggests that task-based decoding together with subjective self-reports are an important tool in studying unconstrained self-generated cognitive and affective information processing with greater rigor than before.

References

Addis, D.R., Wong, A.T., Schacter, D.L., 2007. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45, 1363-1377.

Bechara, A., Damasio, H., Damasio, A.R., 2000. Emotion, decision making and the orbitofrontal cortex. Cerebral Cortex 10, 295-307.

Berridge, K.C., Kringelbach, M.L., 2013. Neuroscience of affect: brain mechanisms of pleasure and displeasure. Curr Opin Neurobiol 23, 294-303.

Bilder, R.M., Sabb, F.W., Parker, D.S., Kalar, D., Chu, W.W., Fox, J., Freimer, N.B., Poldrack, R.A., 2009. Cognitive ontologies for neuropsychiatric phenomics research. Cognitive Neuropsychiatry 14, 419-450.

Bode, S., Bogler, C., Haynes, J.-D., 2013. Similar neural mechanisms for perceptual guesses and free decisions. Neuroimage 65, 456-465.

Bode, S., Haynes, J.-D., 2009. Decoding sequential stages of task preparation in the human brain. Neuroimage 45, 606-613.

Bode, S., He, A.H., Soon, C.S., Trampel, R., Turner, R., Haynes, J.D., 2011. Tracking the unconscious generation of free decisions using ultra-high field fMRI. PLoS One 6, e21612.

Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., Andrews-Hanna, J.R., Sperling, R.A., Johnson, K.A., 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J Neurosci 29, 1860-1873.

Christoff, K., Gordon, A.M., Smallwood, J., Smith, R., Schooler, J.W., 2009. Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. Proceedings of the National Academy of Sciences 106, 8719-8724.

Corradi-Dell'Acqua, C., Hofstetter, C., Vuilleumier, P., 2011. Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. J Neurosci 31, 17996-18006.

Davidson, R.J., Putnam, K.M., Larson, C.L., 2000. Dysfunction in the neural circuitry of emotion regulation--a possible prelude to violence. Science 289, 591-594.

Epel, E.S., Puterman, E., Lin, J., Blackburn, E., Lazaro, A., Mendes, W.B., 2012. Wandering Minds and Aging Cells. Clinical Psychological Science 1, 75-83.

Etkin, A., Egner, T., Kalisch, R., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn Sci 15, 85-93.

Etkin, A., Wager, T.D., 2007. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. Am J Psychiatry 164, 1476-1488.

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. Proc Natl Acad Sci U S A 102, 9673-9678.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical Parametric Maps in Functional Imaging: A General Linear Approach. Hum Brain Mapp 2, 189-210.

Grabenhorst, F., Rolls, E.T., 2011. Value, pleasure and choice in the ventral prefrontal cortex. Trends Cogn Sci 15, 56-67.

Hampton, A.N., O'Doherty, J.P., 2007. Decoding the neural substrates of reward-related decision making with functional MRI. Proceedings of the National Academy of Sciences of the United States of America 104, 1377-1382.

Hasenkamp, W., Wilson-Mendenhall, C.D., Duncan, E., Barsalou, L.W., 2012. Mind wandering and attention during focused meditation: A fine-grained temporal analysis of fluctuating cognitive states. Neuroimage 59, 750-760.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425-2430.

Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. Current Biology 17, 323-328.

Heinzle, J., Anders, S., Bode, S., Bogler, C., Chen, Y., Cichy, R.M., Hackmack, K., Kahnt, T., C., K., Reverberi, C., Soon, S.C., Tusche, A., M., W., Haynes, J.D., 2012. Multivariate decoding of fMRI data - Towards a content-based cognitive neuroscience. e-Neuroforum 3, 1-16.

Herbert, C., Kissler, J., Junghofer, M., Peyk, P., Rockstroh, B., 2006. Processing of emotional adjectives: Evidence from startle EMG and ERPs. Psychophysiology 43, 197-206.

Izquierdo, A., Suda, R.K., Murray, E.A., 2005. Comparison of the effects of bilateral orbital prefrontal cortex lesions and amygdala lesions on emotional responses in rhesus monkeys. J Neurosci 25, 8534-8542.

Kahnt, T., Heinzle, J., Park, S.Q., Haynes, J.D., 2011. Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. Neuroimage 56, 709-715.

Kassam, K.S., Markey, A.R., Cherkassky, V.L., Loewenstein, G., Just, M.A., 2013. Identifying Emotions on the Basis of Neural Activation. PLoS One 8, e66032.

Killingsworth, M.A., Gilbert, D.T., 2010. A Wandering Mind Is an Unhappy Mind. Science 330, 932-932.

Kragel, P.A., Carter, R.M., Huettel, S.A., 2012. What makes a pattern? Matching decoding methods to data in multivariate pattern analysis. Front Neurosci 6, 162.

Kriegeskorte, N., Formisano, E., Sorger, B., Goebel, R., 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A 104, 20600-20605.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc Natl Acad Sci U S A 103, 3863-3868.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12, 535-540.

Kringelbach, M.L., 2013. Limbic Forebrain: The Functional Neuroanatomy of Emotion and Hedonic Processing. In: DW, P. (Ed.), Neuroscience in the 21st Century. Springer, New York, pp. 1335-1363.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., Pessiglione, M., 2009. An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. Neuron 64, 431-439.

Lenartowicz, A., Kalar, D.J., Congdon, E., Poldrack, R.A., 2010. Towards an Ontology of Cognitive Control. Topics in Cognitive Science 2, 678-692.

Lewis-Peacock, J.A., Postle, B.R., 2008. Temporary activation of long-term memory supports working memory. J Neurosci 28, 8765-8771.

Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 53, 103-118.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10, 424-430.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal Representations of Perceived Emotions in the Human Brain. Journal of Neuroscience 30, 10127-10134.

Poldrack, R.A., Halchenko, Y.O., Hanson, S.J., 2009. Decoding the large-scale structure of brain function by classifying mental States across individuals. Psychological Science 20, 1364-1372.

Reverberi, C., Görgen, K., Haynes, J.-D., 2012. Compositionality of Rule Representations in Human Prefrontal Cortex. Cerebral Cortex 22, 1237-1246.

Rolls, E.T., Hornak, J., Wade, D., McGrath, J., 1994. Emotion-related learning in patients with social and emotional changes associated with frontal pole damage. J Neurol Neurosurg Psychiatry 57, 1518-1524.

Roy, M., Shohamy, D., Wager, T.D., 2012. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. Trends Cogn Sci 16, 147-156.

Ruby, F.J.M., Smallwood, J., Engen, H., Singer, T., 2013a. How Self-Generated Thought Shapes Mood—The Relation between Mind-Wandering and Mood Depends on the Socio-Temporal Content of Thoughts. PLoS One 8, e77554.

Ruby, F.J.M., Smallwood, J., Sackur, J., Singer, T., 2013b. Is Self-Generated Thought a means of Social Problem Solving? Frontiers in Psychology 4.

Schoenbaum, G., Takahashi, Y., Liu, T.-L., McDannald, M.A., 2011. Does the orbitofrontal cortex signal value? Annals of the New York Academy of Sciences 1239, 87-99.

Smallwood, J., 2013a. Distinguishing how from why the mind wanders: a process-occurrence framework for self-generated mental activity. Psychol Bull 139, 519-535.

Smallwood, J., 2013b. Searching for the elements of thought: reply to Franklin, Mrazek, Broadway, and Schooler (2013). Psychol Bull 139, 542-547.

Smallwood, J., Fitzgerald, A., Miles, L.K., Phillips, L.H., 2009. Shifting Moods, Wandering Minds: Negative Moods Lead the Mind to Wander. Emotion 9, 271-276.

Smallwood, J., O'Connor, R.C., 2011. Imprisoned by the past: unhappy moods lead to a retrospective bias to mind wandering. Cogn Emot 25, 1481-1490.

Smallwood, J., O'Connor, R.C., Sudbery, M.V., Obonsawin, M., 2007. Mind-wandering and dysphoria. Cognition & Emotion 21, 816-842.

Smallwood, J., Schooler, J.W., 2006. The restless mind. Psychol Bull 132, 946-958.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A 106, 13040-13045.

Song, X.W., Dong, Z.Y., Long, X.Y., Li, S.F., Zuo, X.N., Zhu, C.Z., He, Y., Yan, C.G., Zang, Y.F., 2011. REST: a toolkit for resting-state functional magnetic resonance imaging data processing. PLoS One 6, e25031.

Soon, C.S., Brass, M., Heinze, H.J., Haynes, J.D., 2008. Unconscious determinants of free decisions in the human brain. Nat Neurosci 11, 543-545.

Soon, C.S., He, A.H., Bode, S., Haynes, J.-D., 2013. Predicting free choices for abstract intentions. Proceedings of the National Academy of Sciences 110, 6217-6222.

Spreng, R.N., 2012. The fallacy of a “task-negative” network. Frontiers in Psychology 3.

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., D'Argembeau, A., 2011. Mind-wandering: phenomenology and function as assessed with a novel experience sampling method. Acta Psychol (Amst) 136, 370-381.

Tusche, A., Bode, S., Haynes, J.D., 2010. Neural Responses to Unattended Products Predict Later Consumer Choices. Journal of Neuroscience 30, 8024-8031.

Wallis, J.D., 2007. Orbitofrontal cortex and its contribution to decision-making. Annu Rev Neurosci 30, 31-56.

Watts, F.N., Macleod, A.K., Morris, L., 1988. Associations between Phenomenal and Objective Aspects of Concentration Problems in Depressed-Patients. British Journal of Psychology 79, 241-250.

Weygandt, M., Blecker, C.R., Schäfer, A., Hackmack, K., Haynes, J.-D., Vaitl, D., Stark, R., Schienle, A., 2012. fMRI pattern recognition in obsessive–compulsive disorder. Neuroimage 60, 1186-1193.

Wilson-Mendenhall, C.D., Barrett, L.F., Barsalou, L.W., 2013. Neural Evidence That Human Emotions Share Core Affective Properties. Psychological Science 24, 947-956.

Worsley, K.J., Taylor, J.E., Carbonell, F., Chung, M.K., Duerden, E., Bernhardt, B., Lyttelton, O., Boucher, M., Evans, A.C., 2009. SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory. Neuroimage 47, S102.

Yan, C.-G., Zang, Y.-F., 2010. DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front Syst Neurosci 4, 13.

Zhang, D., Raichle, M.E., 2010. Disease and the brain's dark energy. Nat Rev Neurol 6, 15-28.

Zuo, X.N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F.X., Sporns, O., Milham, M.P., 2012. Network centrality in the human functional connectome. Cerebral Cortex 22, 1862-1875.

Figures

Figure1. Experimental stimulation in Session A and Session B.

Participants took part in two scanning sessions (A and B) separated by a 1-week interval, with the session order counterbalanced across subjects. In Session A, participants performed 8 blocks of positive and negative self-referential attributions that were externally cued by centrally presented trait adjectives. In Session B, they took part in 6 runs of task-free rest periods of 9 min each during which they were asked to fixate on a centrally presented fixation cross. At unpredictable intervals, rest periods were intermitted and thought samples were obtained to assess participants' types of thoughts. Thereby, the affective content of mental states during rest could be assessed. In addition, thought probes were used to assess whether participants' thoughts were related to something in the past or the future and to the self or others.

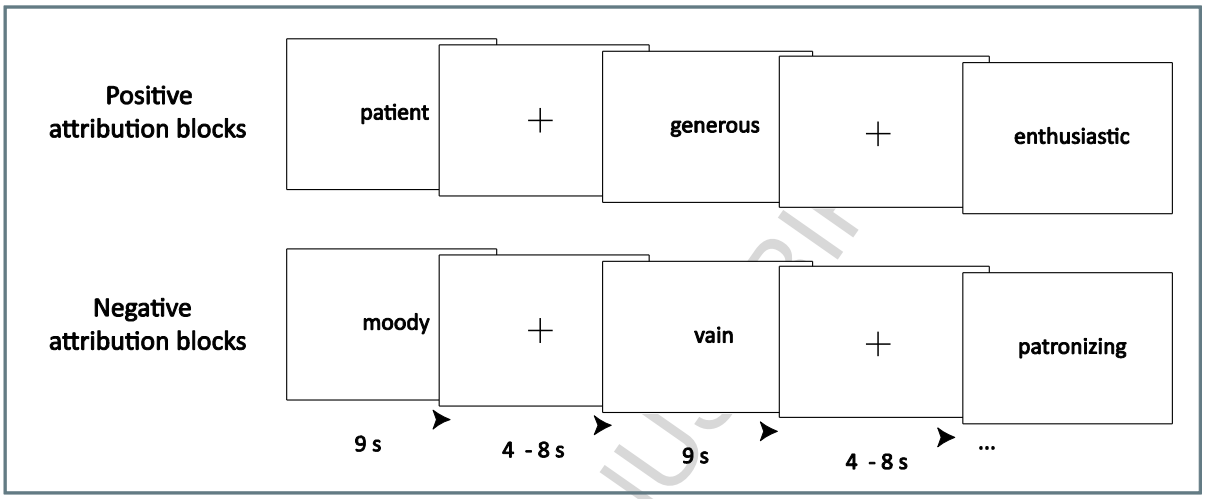
Figure 2. Results.

Self-reported valence of externally cued attributions obtained subsequent to Session A confirmed that participants experienced the attributions in positive task-blocks as more positive than those in negative task-blocks ($p = 0.001$; bars display means and SEM). B. The average experienced valence in the attribution task (Session A) was positively correlated with participants' average self-reported valence during task-free rest periods in Session B, separated by a one-week interval (Pearson's $r = 0.34$, $p = 0.03$, one-tailed). C. MVPA searchlight decoding was used to identify brain regions that encode the valence of task-related mental states in Session A. Activation patterns in the mOFC predicted the affective content of self-referential attributions (average decoding accuracy of 62%). D. We created a spherical ROI around the statistical peak of the mOFC cluster that encoded the valence of task-related attributions in Session A. E. Multi-voxel response patterns in the mOFC-ROI reliably predicted the valence of self-generated mental states during task-free rest prior to thought samples (Session B). This was found to hold true independent of whether participants first performed Session A or Session B ($p = 0.71$, illustrated in dark grey). Supplemental ROI-based decoding using random labels yielded predictions close to the chance level of 50% (illustrated in light grey), indicating that our decoding approach was unbiased. Mean and SEM are displayed. F. ROI-based analysis found that the average degree centrality in the mOFC derived from an independent resting-state Session C (without thought probes) was positively correlated with participants' average valence of task-free thoughts during rest in Session B ($p < 0.025$, corresponding to $r = 0.51$, z-transformed values are displayed).

Figure 3. Neural prediction of the valence of thoughts during task-free rest periods in Session B.

Using a whole-brain searchlight decoding approach, activation patterns in the mOFC predicted the valence of self-generated mental states during rest periods in Session B (59% average decoding accuracy; $p < 0.05$, corrected). The figure displays the overlap (illustrated in yellow) of the predictive searchlight cluster illustrated in green) and the mOFC-ROI based on task-related attributions in Session A (illustrated in red).

A Session A. Task-related positive and negative thoughts.



B Session B. Task-free positive and negative thoughts.

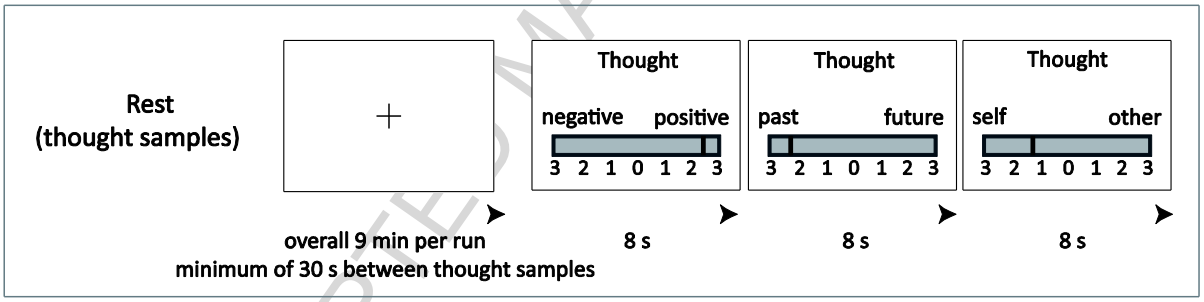


Figure 1

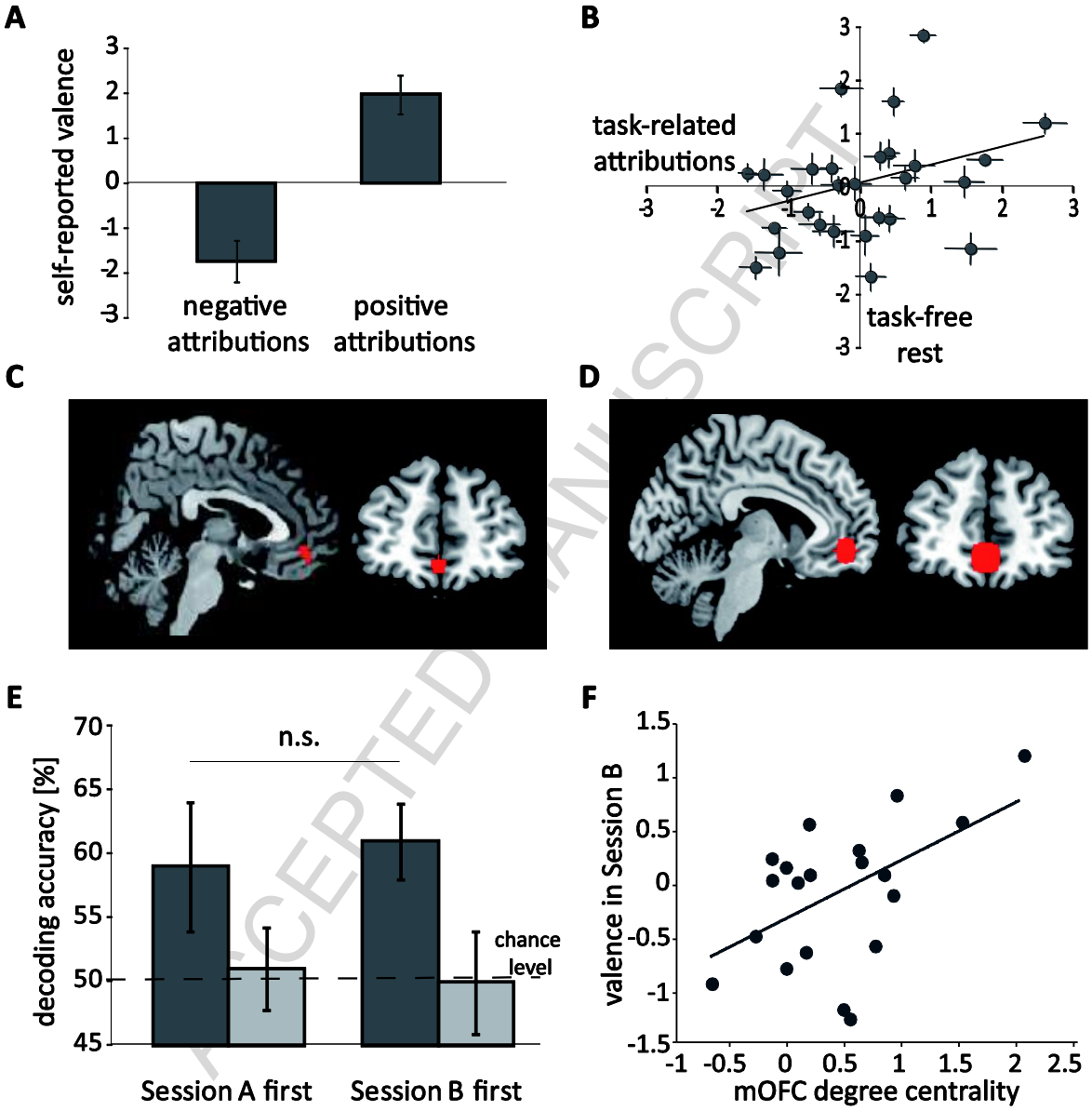


Figure 2

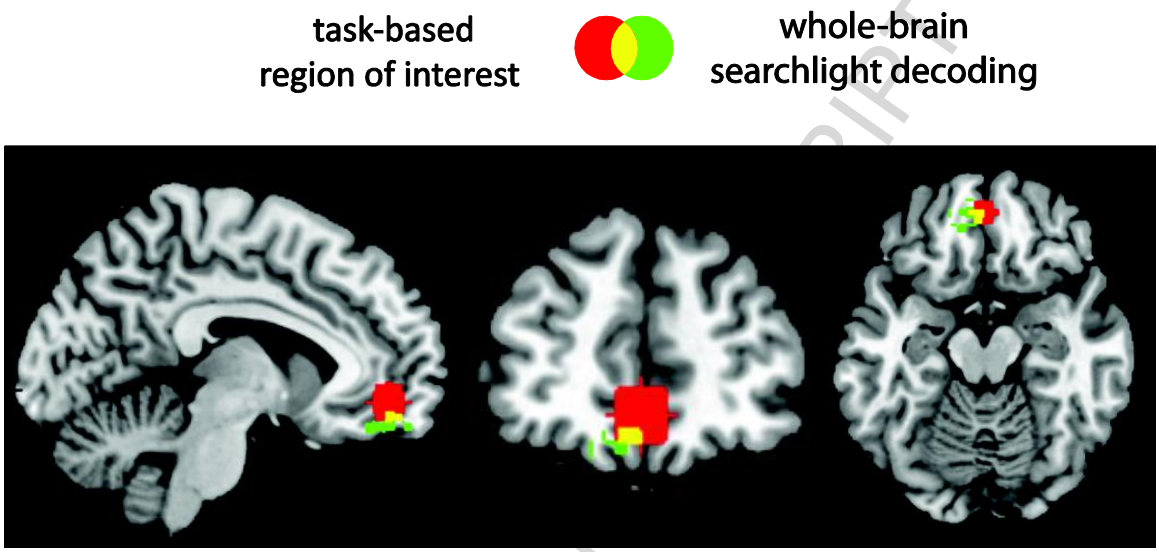


Figure 3

Tables

Table 1: MVPA searchlight decoding in Session A. Brain regions that encode the valence of task-related positive and negative mental states during self-referential attributions.

						MNI		
		BA	Accuracy	T	K	X	Y	Z
mOFC	L/R	11	62	5.60	69	0	48	-8
ACC *	R	32/11	61	4.99	108	14	36	2
Middle frontal gyrus *	L	8	62	5.20	108	-30	6	70
Middle frontal gyrus	L	8	62	4.74	67	-24	28	50
PCC/Precuneus *	L/R	29	63	4.72	118	-8	-46	8
Temporal pole	R	21	63	4.10	97	50	8	-28
Middle temporal gyrus *	L	21/20	62	6.34	262	-56	-36	-10
Middle temporal gyrus	L	21/20	61	4.31	87	-56	-12	-22
Hippocampus	L		61	4.68	77	26	-38	-2
Cerebellum	R		63	5.51	78	28	-76	-30
Cerebellum	L		61	3.89	75	-26	-67	-34

Results are reported at statistical threshold of $p < 0.05$, corrected for multiple comparisons at cluster level using FDR, * indicates clusters that are significant after FWE correction at $p < 0.05$; only peak activations of clusters are reported; mOFC = medial orbitofrontal cortex; ACC = anterior cingulate cortex; PCC = posterior cingulate cortex; L = left hemisphere, R = right hemisphere, BA = Brodmann area, K = cluster size, MNI = Montreal Neurological Institute.

Table 2: MVPA searchlight decoding in Session B. Brain regions that encode valence of task-free mental states during rest periods.

						MNI		
	Side	BA	Accuracy	T	K	X	Y	Z
mOFC *	L/R	11	59	4.51	138	-8	48	-16
ACC/subgenual *	L/R	11/25	63	4.32	161	8	34	-6
Precentral gyrus	R	6	59	4.16	66	22	-16	70
Postcentral gyrus*	L	4	63	4.47	239	-24	-28	60
Inferior parietal lobule *	R	40	59	4.77	183	26	-38	46
Superior temporal gyrus	R	22	59	4.72	102	66	-12	12
Angular gyrus	R	39	60	4.54	69	44	-76	28
Occipital cortex*	R	17	61	4.83	294	14	-90	8
Occipital cortex	L	19	60	5.43	99	-32	-70	36

Results are reported at statistical threshold of $p < 0.05$, corrected for multiple comparisons at cluster level using FDR, * indicates clusters that are significant after FWE correction at $p < 0.05$; only peak activations of clusters are reported; mOFC = medial orbitofrontal cortex; ACC = anterior cingulate cortex; L = left hemisphere, R = right hemisphere, BA = Brodmann area, K = cluster size, MNI = Montreal Neurological Institute.