A causal test of the mechanisms by which affect state biases affective perception

Authors
Keith A. Bush*, Clinton D. Kilts

Affiliation
Brain Imaging Research Center, Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR 72205

Address correspondence to:
Keith A. Bush, Ph.D.
Brain Imaging Research Center
Department of Psychiatry
University of Arkansas for Medical Sciences
4301 W. Markham St. #554
Little Rock, AR 72205
Email: kabush@uams.edu

Abstract
In this study, we merged methods from machine learning and human neuroimaging to causally test the role of self-induced affect states in biasing the affective perception of subsequent image stimuli. To test this causal relationship, we developed a novel paradigm in which (n=40) healthy adult participants observed multivariate neural decodings of their real-time functional magnetic resonance image (rtfMRI) responses as feedback to guide explicit regulation of their brain (and corollary affect processing) state towards a positive valence goal state. By this method, individual differences in affect regulation ability were controlled. Attaining this brain-affect goal state triggered the presentation of pseudo-randomly selected affectively congruent (positive valence) or incongruent (negative valence) image stimuli drawn from the International Affective Picture Set. Separately, subjects passively viewed randomly triggered positively and negatively valenced image stimuli during fMRI acquisition. Multivariate neural decodings of the affect processing induced by these stimuli were modeled using the task trial type (state- versus randomly-triggered) as the fixed-effect of a general linear mixed effects model. Random effects were modeled subject-wise. We found that self-induction of a positive affective valence state significantly positively biased the perceived valence of subsequent stimuli. As a manipulation check, we validated affective state induction achieved by the image stimuli using independent psychophysiological response measures of hedonic valence and autonomic arousal. We also validated the predictive fidelity of the trained neural decoding models for brain states induced by an out-of-sample set of image stimuli. Beyond its contribution to our understanding of the neural mechanisms that bias affect processing, this work demonstrated the viability of novel experimental paradigms triggered by pre-defined affective cognitive states. This line of individual differences experimentation potentially provides scientists with a valuable tool for causal exploration of the roles and identities of intrinsic cognitive processing mechanisms that shape our perceptual processing of sensory stimuli.

Introduction

Our capacity to process and regulate emotions is central to our ability to optimize psychosocial functioning and quality of life[1]. As a corollary, disruptions in emotion processing regulation are broadly ascribed to psychiatric illnesses including borderline personality disorder, depression, anxiety disorders, PTSD, and substance-use disorders[2] which negatively impact quality of life and functioning[3,4]. In light of this, a primary focus of cognitive behavioral therapy (CBT), an effective treatment for disorders involving emotion dysregulation[5], is the development of mental strategies for identifying and volitionally reducing negatively biased emotional states that are the product of maladaptive emotion processing and regulation. Neuroimaging has provided great insight into the functional neurocircuits involved in CBT-based emotion regulation strategies[6]; however, the causal neurobiological mechanisms by which these strategies induce adaptive emotion processing over time remain elusive.

Research into the effects of temporal context on affect and emotion processing may have implications for increasing our understanding of the neural bases of emotion regulation. Prior work has demonstrated that changing affective context prior to an emotional target shapes the processing of that target. Such priming effects both accelerate and weaken the emotional response to affectively congruent target stimuli[7]. Manipulations of affect state impact the temporal structure of the neural responses to subsequent affective image stimuli[8] as well as the corollary psychophysiological responses to those stimuli[9,10]. Further, stimulus-cued emotional states bias the self-reported perception of successive emotional stimuli[11].

These findings are consistent with effects that would be predicted by the deployment of situational and attentional modification strategies according to the process model of emotion regulation[12] and point to potential underlying mechanisms driving CBT-related changes to emotion processing. However, the ability of affective cognitions related to these strategies to bias subsequent emotional responses has not yet been causally tested. Thus, the primary aim of this work was to contribute to our knowledge of the mechanisms underlying emotion regulation by experimentally demonstrating that self-induced and verified emotional states causally bias the affective perception of image stimuli.

Real-time functional magnetic resonance imaging (rtfMRI), when used to generate brain activation feedback[13] (i.e., rtfMRI-guided neuromodulation or neurofeedback), reflects a promising methodology that has not to our knowledge been applied for mechanistic testing of how the specific context related to such feedback-induced affect states causally bias affective perceptions. Here, the applied advantage of rtfMRI is that self-induced neurocognitive states (achieved with the aid of rtfMRI guidance) can be verified and used as independent experimental variables to trigger subsequent emotional stimulus-response characterizations. Yet, a challenge to rtfMRI-guided neuromodulation studies, and brain computer interface (BCI) research in general, is the large individual variation observed in subjects' ability to volitionally modulate their cognitive states – the well-known "BCI-illiteracy phenomenon"[14].

Within BCI studies, neurophysiological and psychological variables (e.g., self-confidence and concentration) were shown to significantly predict performance variation[15–17]. However, very little is known about individual differences in the ability to volitionally regulate emotional states. Therefore, the secondary aim of this project was to characterize individual variation in the ability to self-induce emotional states using neurofeedback according to the subjects' unguided self-induction ability. This research has direct clinical relevance to informing our understanding of the

neuroregulation capabilities of psychiatric patients to identify those most or least capable of emotion regulation.

To explore our aims, we developed a novel task in which healthy adult participants utilized rtfMRI feedback to explicitly regulate their brain and corollary affect processing states towards a goal of extreme pleasantness (positive valence). Reaching this brain-affect state triggered the presentation of an affectively congruent (positive valence) or incongruent (negative valence) image stimulus drawn from the International Affective Picture Set[18] (IAPS). Between regulation trials, participants passively viewed (without regulation) IAPS stimuli associated with either positive or negative valence. We then compared image stimulus-cued brain and emotional responses arising from explicitly feedback-facilitated and self-induced positively valent emotional states versus random emotional states (passive viewing) and causally tested the ability of self-induced positive affective states to bias the affective perception of image stimuli.

Our results reveal that self-induction of positive affect causally biases affect processing responses to image stimuli in a manner similar to viewing affectively laden image stimuli[11] suggesting a potential mechanism by which CBT-based mental strategies may work to reduce negatively biased emotional processing states. However, we also found that individual differences in the intrinsic ability to precisely self-induce affect processing states without guidance did not generalize to the achievement of self-induced positive affect in the presence of rtfMRI-feedback, potentially suggesting inherent affect regulation ability separate from that of concentration, e.g., the ability to accurately perceive temporally proximal affect processing states or temporally distal goal states. Additional research will be necessary to characterize the latent neurobiological and psychological factors driving these individual differences.

Methods

Ethics Statement

All participants provided written informed consent after receiving written and verbal descriptions of the study procedures, risks, and benefits. We performed all study procedures and analysis with approval and oversight of the Institutional Review Board at the University of Arkansas for Medical Sciences (UAMS) in accordance with the Declaration of Helsinki and relevant institutional guidelines and policies.

Participants

We enrolled subjects (n=40) having the following demographic characteristics: age [mean(s.d.)]: 38.8(13.3), range 20–65; sex: 22 (55%) female; race/ethnicity: 28 (70.%) self-reporting as White or Caucasian, 9 (22.5%) as Black or African-American, 1 (2.5%) as Asian, and 2 (5%) self-reporting as other; education [mean(s.d.)]: 16.8(2.2) years, range 12–23; WAIS-IV IQ [mean(s.d.)]: 102.5(15.3), range 73–129. All of the study's participants were right-handed (assessed via Edinburgh Handedness Inventory[19]) native-born United States citizens who were medically healthy and exhibited no current Axis I psychopathology, including mood disorders, as assessed by the SCID-IV clinical interview[4]. All participants reported no current use of psychotropic medications and produced a negative urine screen for drugs of abuse (cocaine, amphetamines, methamphetamines, marijuana, opiates, and benzodiazepines) immediately prior to both the clinical interview and MRI scan. When necessary, we corrected participants' vision to 20/20 using an MRI compatible lens system (MediGoggles™, Oxforshire, United Kingdom), and

we excluded all participants endorsing color blindness.

Experiment Design.

Following the provision of informed consent, subjects visited the Brain Imaging Research Center of the University of Arkansas for Medical Sciences on two separate days. On Study Day 1 a trained research assistant assessed all subjects for major medical and psychiatric disorders as well as administered instruments to collect the following data to be used as either secondary variables hypothesized to explain individual variance in emotion regulation-related neural activity, covariates of no interest, or to assess inclusion/exclusion criteria: demographics (BIRC demographic collection form), verbal IQ (Receptive and Expressive One-Word Picture Vocabulary Test[20]), working memory (Wechsler Adult Intelligence Scale, Digit Span task[21]), current and past psychiatric disorders and drug use history (Structure Clinical Interview for DSM-IV[4], depression symptom severity (Beck Depression Inventory[22]), history of childhood abuse and neglect (Childhood Trauma Questionnaire[23]), emotion dysregulation (Difficulties in Emotion Regulation Scale[24]), handedness (Edinburgh Handedness Inventory[19]), personality (NEO Five-Factor Inventory[25]), anxiety (State-Trait Anxiety Inventory[26] – trait assessed on Study Day 1 and state assessed on Study Day 2), and emotional invalidation (Perceived Invalidation of Emotion Scale[27], PIES). The participant returned to the BIRC for Study Day 2 within 30 days after Study Day 1 to complete the MRI acquisition. During this day, the participant received training and completed the full MRI acquisition protocol, depicted in Figure 1.

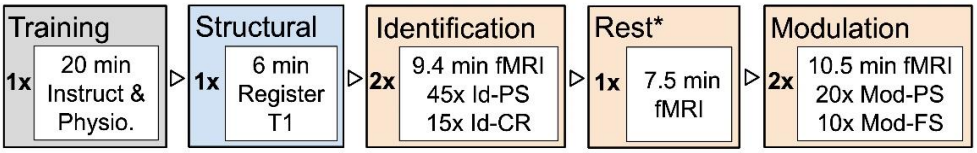

Figure 1: *Study Day 2 Experimental tasks: order, number of repetitions, duration, and stimuli. Tasks are colored by role. Gray depicts task training and application of psychophysiology recording apparatus. Blue depicts brain structural image acquisition. Orange depicts functional image acquisition. Identification and Modulation blocks of the fMRI acquisition summarize the relevant trial types used within that task (see Neuroimaging section for abbreviations). *Training of real-time multivariate pattern analysis predictive models was performed concurrently with the Resting State task of the fMRI acquisition.*

Training: Each participant received a video-based overview of the experiment to be performed on that day as well as training on the study's task variations and trial types. The participant was offered the opportunity to use the restroom and then was moved to the MRI scanner room and fully outfitted with psychophysiological recording equipment.

Neuroimaging: For each subject we captured a registration scan and detailed T1-weighted structural image. We then acquired functional MRI data for three task variations: identification, resting state, and modulation. Identification (Id) task acquisition consisted of 2 x 9.4 min fMRI scans during which the participant was presented with 120 images drawn from the International Affective Picture System[18] (IAPS) to support one of two trial types (see Figure 2): 90 passive stimulus (PS) trials and 30 cued-recall (CR) trials. Identification task PS trials (abbreviated Id-PS)

presented an image for 2 s (cue) succeeded by a fixation cross for a random inter-trial interval (ITI) sampled uniformly from the range 2–6 s. Identification task cued-recall (Id-CR) trials were multi-part: a cue image was presented for 2 s followed by an active cue response step for 2 s (the word "FEEL" overlaying the image) followed by the word FEEL alone for 8 s, which signaled the participant to actively recall and re-experience the affective content of the cue image, followed by a 2–6 s ITI. During pre-scan training on the Id-CR task's recall condition, subjects were instructed to "Imagine the last picture you saw as best you can. Try to make yourself feel exactly how you felt when you saw this picture the first time. Hold that feeling the whole time you see the word FEEL." Within each scan, Id-PS and Id-CR trials were pseudo-randomly sequentially ordered to minimize correlations between the hemodynamic response function (HRF)-derived regressors of the tasks. The order was fixed for all subjects.

During resting state acquisition, we acquired 7.5 min of fMRI data in which the subject performed mind-wandering with eyes open while observing a fixation cross. During training, subjects were instructed to "Keep your eyes open, look at the cross in front of you, and let your brain think whatever it wants to." Concurrently with the resting state task, the real-time variant of the MVPA prediction model (see below) was fit using data drawn from the Identification task fMRI data to define individual brain state representations of the affect goal.

Modulation (Mod) task acquisition consisted of 2 x 10.5 min fMRI scans during which the participant was presented with 60 IAPS images according to two trial types (see Fig 2): 40 passive stimulus (Mod-PS) trials, which were identically formatted to the Id-PS trials, and 20 feedback-triggered stimulus (Mod-FS) trials. Mod-FS trials used real-time fMRI feedback of the subject's affective state to guide them in self-inducing affective brain states associated with their representations of extreme positive valence. The computer system monitored the subject's neural representation of their valence levels within each acquisition volume of fMRI data and if that representation met pre-defined criteria (i.e., the goal state, which we defined as hyperplane distance ≥ 0.8 for 4 consecutive EPI volumes) then a positively (congruent) or negatively (incongruent) valent image stimulus was triggered. The brain state criteria representing the affect goal state were determined by the results of an initial pilot of the experiment to identify parameters that were challenging but consistently reachable. Within each scan, Mod-PS and Mod-FS trials were pseudo-randomly sequentially ordered to minimize correlations between the hemodynamic response function (HRF)-derived regressors of the tasks. The order was fixed for all subjects.

We provided real-time visual feedback during Mod-FS trials by manipulating the transparency of the word FEEL, which was the cue to volitionally regulate affect to an extreme positive valence. The transparency of the text was scaled to reflect real-time estimates of subject's represented affective valence with respect to the desired hyperplane distance threshold. This was achieved by mapping MVPA prediction model hyperplane distances (see below) from their base range [-1.25,1.25] to the range of possible transparencies, $\alpha \in [0,1]$. Fully transparent text ($\alpha=0$) appeared as a black screen and denoted poor affect regulation performance, i.e., highly negative valence. Fully opaque text ($\alpha=1$) appeared bright yellow and denoted good performance. The transparency of the text was reset every 2 s (reflecting the momentary hyperplane distance prediction based upon the current EPI volume). The transparency was adjusted (approximately 20 frames-per-second) to present smooth transitions towards that brain-affect goal. The initial hyperplane distance threshold was fixed for 20 seconds. If the subject had not attained the threshold (i.e. triggered the test stimulus) by this time then the threshold was linearly and

continuously lowered to 0 over the subsequent 18 s at which point the stimulus was automatically triggered even if the threshold had not been attained (Fig. 2).

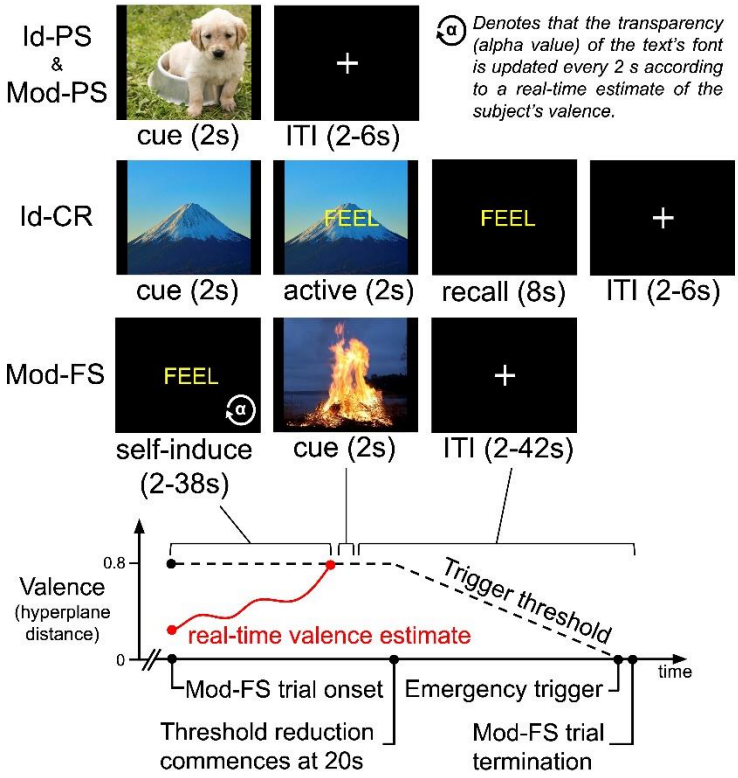

Figure 2: *Summary of experimental task trial designs. (Id-PS): Identification task passive stimulus trials, which were identical to Modulation task passive stimulus (Mod-PS) trials. (Id-CR): Identification task cued-recall trials. (Mod-FS): Modulation task feedback-triggered stimulus trials. (Bottom): depiction of a hypothetical Mod-FS trial for the experimental design.*

Stimulus Selection: We sampled 180 IAPS images to use as emotion processing induction stimuli. Identification task stimuli were sampled computationally using a previously published algorithm[28] that selects images such that the subspace of the valence-arousal plane for normative scores within the IAPS dataset is maximized (see Fig 3). We performed this sampling process first for the 90 images used in Id-PS trials. The IAPS identifiers of these images were previously reported[29]. We then separately (but similarly) sampled an additional 30 images used in Id-CR trials. The IAPS identifiers of these images were also previously reported[30]. Next we constructed extreme polar subsets of positively and negatively valenced image stimuli by constructing thresholds of permissible valence and arousal scores. Valence (v) was constrained such that: $v \geq 7$ or $v \leq 2.6$. We then iteratively constrained the permissible arousal scores until we identified positively and negatively valenced subsets that did not exhibit a group mean difference in arousal scores (found to be $4.6 < A < 6.8$) thereby controlling for arousal response as a stimulus subset variable. We then sampled 30 images each from these subsets and uniformly randomly assigned these images to Mod-PS trials (n=40) and Mod-FS trials (n=20), respectively. The outcome of this sampling and assignment process is presented in Figure 3. The specific IAPS identities of these

images are reported in the Appendix.

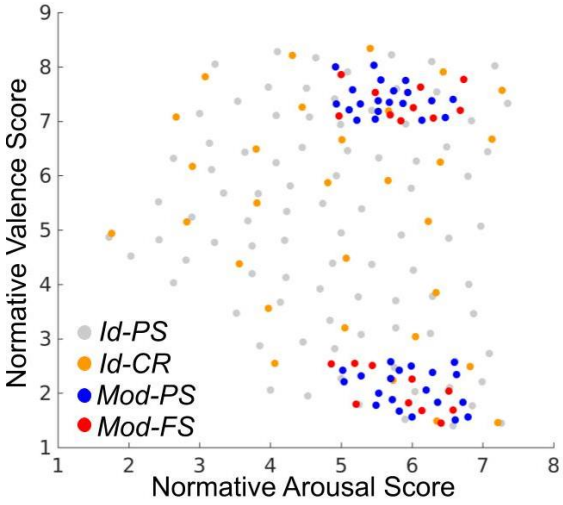

Figure 3: *Normative valence and arousal scores for stimuli selected for each of the four experimental trial types. Summary statistics for Identification task stimuli are as follows: Id-PS valence [mean (std. dev)] 5.04 (1.95); Id-PS arousal [mean (std. dev)] 4.95 (1.40); Id-CR valence [mean (std. dev)] 5.30 (1.95); Id-CR arousal [mean (std. dev)] 4.99 (1.51). There were no significant differences in affect properties between the Id-PS and Id-CR cue stimuli for either valence (p=.49; signrank, h_0: $\mu_1 = \mu_2$) or arousal (p=.86; ranksum, h_0: $\mu_1 = \mu_2$). Summary statistics for the Modulation task stimuli are as follows. Mod-PS (pos. valence cluster) valence [mean (std. dev)] 7.41 (.30); Mod-PS (neg. valence cluster) valence [mean (std. dev)] 2.08 (.36); Mod-FS (pos. valence cluster) valence [mean (std. dev)] 7.35 (0.32); Mod-FS (neg. valence cluster) valence [mean (std. dev)] 2.03 (0.41). Between the Mod-PS and Mod-FS stimuli in the positive valence cluster, there were no significant difference in valence (p=.60; ranksum; h_0: $\mu_1 = \mu_2$) nor arousal (p=.25; ranksum; h_0: $\mu_1 = \mu_2$). There were also no significant group differences in affect properties between the Mod-PS and Mod-FS stimuli in the negative valence cluster, either for valence (p=.74; ranksum; h_0: $\mu_1 = \mu_2$) or arousal (p=.54 ranksum; h_0: $\mu_1 = \mu_2$).*

MR Image Acquisition

We acquired all imaging data using a Philips 3T Achieva X-series MRI scanner (Philips Healthcare, Eindhoven, The Netherlands) with a 32-channel head coil. We acquired anatomic images using an MPRAGE sequence (matrix = 256 x 256, 220 sagittal slices, TR/TE/FA = 8.0844/3.7010/8°, final resolution =0.94 x 0.94 x 1 mm^3. We acquired functional images using the following EPI sequence parameters: TR/TE/FA = 2000 ms/30 ms/90°, FOV = 240 x 240 mm, matrix = 80 x 80, 37 oblique slices, ascending sequential slice acquisition, slice thickness = 2.5 mm with 0.5 mm gap, final resolution 3.0 x 3.0 x 3.0 mm^3.

Real-time MRI Preprocessing and Multivariate Pattern Classification

We implemented custom code that acquired each raw fMRI volume as it was written to disk by the MRI's computer system (post-reconstruction). Each volume underwent a preprocessing

sequence using AFNI[31] in the following order: motion correction using rigid body alignment (corrected to the first volume of Identification task Run 1), detrending (re-meaned), spatial smoothing using a 8 mm FWHM Gaussian filter, and segmentation. To construct a multivariate pattern classifier to apply to the real-time data we partitioned the Id-PS stimuli into groups of positive and negative valence (according to the middle Likert normative score) and formed time-series by convolving the hemodynamic response function with the respective stimuli's onset times (scaling the HRF amplitude according to the absolute difference between the stimuli's normative scores and the middle Likert score). We then thresholded these time-series to construct class labels {-1,+1} (as well as unlabeled) for each volume of the Identification task scans. We then trained a linear support vector machine[32] (SVM) to classify the valence property of each fMRI volume. Note, during the Modulation task the classification hyperplane output of the SVM was linearly detrended in real-time as follows. A hyperplane distance, h, was computed for each volume, i. For h_i, $i \geq 40$, the sequence of hyperplane distances $h_1,...,h_{i-1}$ was used to compute a linear trend (via the Matlab detrend function) which was subtracted from the hyperplane distance, h_i. In summary, the described system achieved real-time preprocessing and generated affect state predictions for each EPI volume acquired in the Modulation task of the experiment. Total processing time of each volume was less than the TR=2.0s parameter of the EPI sequence, allowing the real-time processing to maintain a consistent (reconstruction speed determined) latency throughout real-time acquisition.

Post-hoc MRI Preprocessing, Multivariate Pattern Classification, and Platt-Scaling

We used fmriprep[33] (version 20.0.0) software to conduct skull stripping, spatial normalization to the MNI152 atlas, and (fMRI only) despiking, slice-time correction, deobliquing, and alignment to normalized anatomical images. We then used fmriprep's motion parameter outputs to complete the preprocessing using AFNI, including regression of the mean time courses and temporal derivatives of the white matter (WM) and cerebrospinal fluid (CSF) masks as well as a 24-parameter motion model[34,35], spatial smoothing (8 mm FWHM), detrending, temporal filtering (.0078 Hz high-pass), and scaling to percent signal change. For resting state functional images we took the additional step of global mean signal subtraction prior to smoothing.

We then conducted high-accuracy post-hoc multivoxel pattern analysis (MVPA) of affect processing. We first extracted beta-series[36] neural activation maps associated with Id-PS trials from fully preprocessed fMRI data recorded during Identification task runs 1 and 2 according to well-documented methods[28]. We indexed these maps according to their corresponding stimulus, x. Therefore, the maps, $\beta(x)$, were paired with their respective normative scores $\{\beta(x), v(x), a(x)\}$ to form training data for multivoxel pattern classification implemented via linear SVM. For classification training, valence and arousal scores were each converted into positive (+1) or negative (-1) class labels according to their relation to the middle Likert score. Classification hyperplane distances were then converted to probabilities (i.e., the probability of the positive class label) via Platt-scaling[37]. These probabilities served as the affective decodings of the subjects' brain states for further analysis.

Cued-Recall, Passive Stimulus, and Feedback-Triggered Stimulus Modeling

We extracted beta-series for the cue and recall steps of the Id-CR trials, the cue step of the Mod-PS trials, and the cue step of the Mod-FS trials. We then used our fit SVM models to decode the

valence and arousal properties of the experiment at these steps. For the Mod-PS trials, we also constructed beta-series for the moment of trial onset as well as 2 s prior to the cue step of the Mod-FS trials – these allowed us to validate the triggers for affective stimulus presentations as well as to measure (post-hoc) the relative change of affect processing induced by the real-time fMRI feedback.

Surrogate Cued-Recall Task Modeling

Using previously reported methodology[38], we decoded the valence and arousal properties of each volume of Resting State fMRI data. We then uniformly randomly sampled 30 onset times for surrogate Id-CR trials and extracted the affect properties of the respective cue and recall steps of these surrogate trials to be used as within-subject controls during analysis of the actual Id-CR trials.

Psychophysiology Data Acquisition and Preprocessing

All MRI acquisitions included concurrent psychophysiological recordings conducted using the BIOPAC MP150 Data Acquisition System and AcqKnowledge software combined with the EDA100C-MRI module (skin conductance), TSD200-MRI pulse plethysmogram (heart rate), TSD221-MRI belt (respiration), and EMG100C-MRI module (facial electromyography). In line with prior work[39,40], we measured arousal independently based on skin conductance response (SCR) and valence based on facial electromyography (fEMG) response, specifically activity in the corrugator supercilli muscle, which was shown in prior work to capture the full affective valence range of our affect processing induction design[30]. We have extensively reported on our SCR electrode placement and preprocessing methods[29], and we recently reported our fEMG placement and preprocessing methods[30].

Results

Psychophysiological Response Validation of Affect Processing Induction via Image Stimuli.

We first verified the ability of the Identification task passive stimulus (Id-PS) trials to induce corollary psychophysiological responses[41] associated with affect processing that our machine learning approach would then be independently trained to detect within temporally concurrent affect processing brain states. We modeled normative scores of the cue stimuli of Id-PS trials using psychophysiological response measures within a GLMM framework, respectively, for valence and arousal properties. Normative hedonic valence scores of the stimuli were modeled according to facial electromyographic responses in the corrugator supercilli (cEMG) as the fixed effects. Normative autonomic arousal scores to the cue stimuli were modeled according to skin conductance responses as the fixed effects. In both models, we controlled for age and sex effects. Slope and intercept random-effects were modeled subject-wise. Both validation models detected significant stimulus-related induction of the desired physiological responses. Moreover, our cEMG-derived model of hedonic valence (β=.11; p=0.001; F-test; h_0: β=0) was selective for the valence property of affect – a cEMG-derived model of autonomic arousal was not significant (p=0.75; F-test; h_0: β=0). Similarly, our SCR-derived model was selective for the autonomic arousal property of affect (β=.07; p=.004; F-test; h_0: β=0) – applied to hedonic valence the SCR response associations were not significant (β=0.02; p=0.61; F-test; h_0: β=0).

Affect Processing Measurement

We then demonstrated that our prediction models accurately decoded affect processing within neural activation patterns associated with Id-PS trials, reproducing earlier work using similar modeling methodology[28]. Our tabulated prediction accuracy (averaged over 39 subjects completing the experiment) over the full stimulus set was highly significant for both valence ($p<0.001$; signrank; h_0: $\mu=.5$) and arousal ($p<0.001$; signrank; h_0: $\mu = .5$). We observed prediction performance comparable to the best known examples of classification of affect processing across the valence and arousal dimensions[28,42] when our measurements were restricted to those image stimuli exhibiting reliable brain state activations , i.e., the reliable stimulus set (Table 1), which were determined according to previously published methods[28].

Table 1: *Multivariate Neural Decoding Performance*

	Valence	**Arousal**
	Grp. Avg. Acc. (95% CI)	Grp. Avg. Acc. (95% CI)
Full Stimulus Set	.55 (.53,.57)	.61 (.59,.63)
Reliable Stimulus Set	.79 (.76,.82)	.75 (.72,.79)

Validation of Affect Decoding of Novel Stimuli

Prior to applying our decoding models to novel task domains, we first tested that these models (originally fit to Id-PS features and labels) well-generalized to novel image stimuli. To perform this test we modeled, via GLMM, the normative affect scores of cue stimuli in Id-CR and Mod-PS trials. However, each test was unique. First, we modeled Id-PS stimuli's normative scores as a function of decoded affect (separately for valence and arousal) controlling for the age and sex of the subjects and modeling random effects of affect decoding subject-wise. In Id-CR trials we found that decoded valence was significantly positively associated with the valence normative score ($\beta=.30$; $p<.001$; F-test; h_0: $\beta=0$). Similarly, we found that decoded arousal was significantly associated with the arousal normative score ($\beta=.17$; $p=.001$; F-test; h_0: $\beta=0$). Age and sex effects in both cases were not significant and random effects did not significantly improve the model's explained variance, which was very small for both valence ($R^2_{adj}=.02$) and arousal ($R^2_{adj}=.01$), respectively.

Next, we modeled Mod-PS stimuli's normative scores as a function of decoded affect (separately for valence and arousal normative scores). However, in this case we controlled for age and sex effects as well as the decoding of the complementary affective decoding in order to control for the bias of the sampling of the stimuli in this task (see Fig 3). In Mod-PS trials we found that Mod-PS decoded valence was significantly positively associated with the stimuli's normative valence scores ($\beta=.58$; $p<0.001$; F-test; h_0: $\beta=0$). However, decoded arousal was significantly negatively associated with normative valence scores ($\beta=-.20$; $p=0.02$; F-test; h_0: $\beta=0$). Age and sex effects were not significant but random effects did significantly improve the model's explained variance ($R^2_{adj}=.04$). In contrast, we found no significant associations between decoded arousal and the stimuli's normative arousal scores, which confirmed that the restriction of our sampling of the Mod-PS and Mod-FS stimuli to a narrow range of normative arousal was essential as a control for this confounding variable.

Real-time Stimulus Triggering

We next validated that our real-time feedback and brain-affect state triggering process functioned as designed. To test this we extracted the feedback signal calculated at the moment of stimulus trigger (including emergency triggering). The median feedback at the moment of trigger was μ = .93 ($p<.001$; signrank; h_0: $\mu=0$). Nearly three-quarters (see Figure 4) of all trials triggered at or above the design threshold.

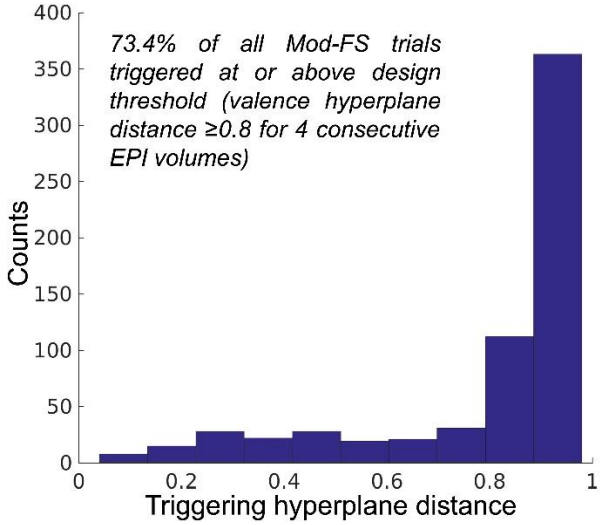

Figure 4: *Distribution of average feedback scores at the moment of FT-PO trial stimulus trigger.*

Real-time fMRI-Guided Self-Induction of Positive Valence States

We next demonstrated that our experimental condition, volitionally-induced positive valence, was truly achieved at the moment of stimulus triggering. As a reminder, the Mod-FS trials were triggered using low-quality real-time affect decoding models. Here we applied post-hoc high-accuracy models to decode affect processing within the fMRI volume immediately prior to the stimulus trigger as a best possible measure of the experimental condition. To test this measure, we bootstrapped random variants of the trigger predictions (randomly sampling within each subject before pooling predictions to incorporate random effects). We found that the mean predicted valence was significantly elevated ($\mu=.515$; $p=.02$; 1-sided bootstrap [n=10000]; h_0: $\mu<.5$).

Causal Effect of Positive Valence Self-Induction on Perceived Affect of Visual Stimuli

Building upon our confidence in the decoding measurement and our experimental condition, we next tested the study's primary hypothesis – that self-induced valence states bias the affective perception of image stimuli. Here, using a GLMM, we tested decoded perceived affect as a function of trial type, Mod-PS or Mod-FS, while controlling for the image stimuli's associated normative valence and arousal properties as well as the subject's age and sex. We modeled random slope and intercept effects of the trial type, subject-wise. Indeed, we found that volitional self-induction of positive valence prior to a stimulus significantly increased its perceived valence ($\beta=.024$; $p=.007$; F-test; h_0: $\beta=0$). Normative valence was also a significant positive predictor ($\beta=.06$; $p<.001$; F-test; h_0: $\beta=0$). Sex effects were not significant but age effects were found to

have a small but significant negative impact on perceived valence (β=-.001; p=.03; F-test; h_0: β=0). Finally, the stimuli's normative arousal scores were found not to be a significant predictor of valence (β=-.06; p=.09; F-test; h_0: β=0). Overall model performance was R^2_{adj}=.065 and random effects significantly impacted the model's explained variance.

Measurement of Explicit Affect Regulation

We next tested our ability to confirm affect self-induction using explicit affect regulation within the Id-CR trials. We first decoded the valence and arousal properties for both the cue and recall steps of the Id-CR trials. We then tested for group effects of affect regulation toward a known goal, using a GLMM by modeling, separately for valence and arousal, the decoded affect of the four recall steps of the Id-CR trials (4 volumes, 2 seconds each) as a function of the decoded affect of the cue stimuli (i.e. the affect regulation goal) as well as the control duration and the age and sex of the subject (see Figure 5). We found that the subjects significantly regulated affective valence (β=.33; p<.001; F-test; h_0: β=0). Random effects significantly improved the model's effect-size (p<.05; likelihood ratio test; h_0: observed responses generated by fixed-effects only) and cued-recall affect regulation effects were significantly greater than that of surrogate effects (p=.001; signrank; h_0: β_{IN}-β_{RST}=0). The fixed-effect of control duration was also significant (β=.01; p<.001; F-test; h_0: β=0) and overall model prediction performance was good (R^2_{adj}=.10). Further, we found that subjects significantly regulated affective arousal and that random effects significantly improved effect-size (β=.33; p<0.05; likelihood ratio test; h_0: observed responses generated by fixed-effects only); however, these cued-recall affect regulation effects were not significantly greater than that of surrogate effects (p=.10; signrank; h_0: β_{IN}- β_{RST}=0).

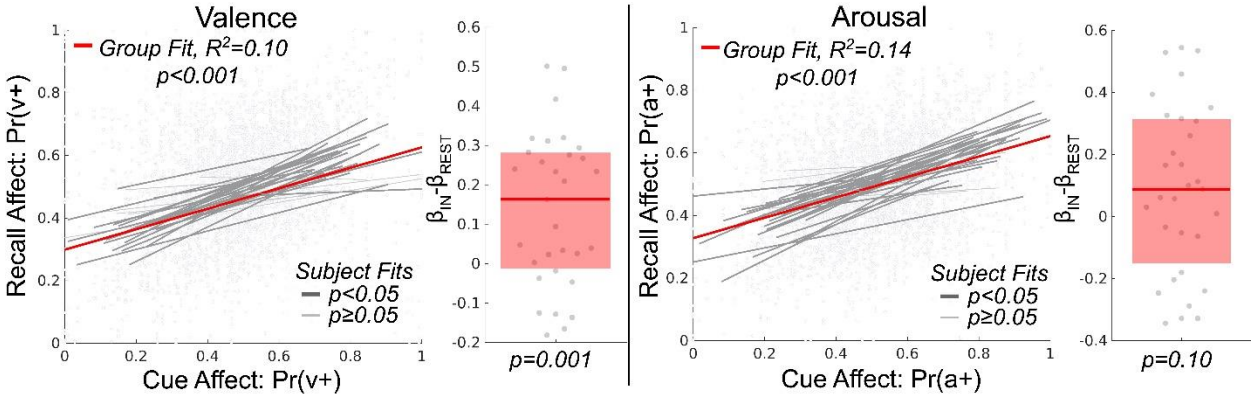

Figure 5: *Estimation and validation of explicit intrinsic affect regulation effects within the cued-recall task. The figure depicts the effect size of cue affect processing in explaining affect processing occurring during recall (controlling for time lag in the 4 repeated measures of recall per each measure of cue). Here affect processing measurements are Platt-scaled hyperplane distance predictions of our fitted support vector machine models. Valence and arousal dimensions of affect are predicted by separate models. The figure's scatterplots depict the group-level effects computed using linear mixed-effects models which model random effects subject-wise. Bold red lines depict group-level fixed-effects of the cue affect. Bold gray lines depict significant subject-level effects whereas light gray lines depict subject-level effects that were not significant. The figure's boxplots depict the group-level difference between each subject's affect regulation*

measured during the cued-recall trials in comparison to surrogate affect regulation constructed from the resting state task. The bold red line depicts the group median difference in effect size between task and surrogate. The red box depicts the 25-75th percentiles of effect size difference.

Explicit Affect Regulation Performance as a Predictor of Real-time fMRI-Guided Self-Induction

Finally, we tested whether unguided explicit affect regulation performance explained the level of rtfMRI-guided self-induced valence (measured immediately prior to presentation of the Mod-FS cue image). We modeled the decoded valence of the final volume of the self-induce step of Mod-FS trials as a function of the individual subjects' explicit affect regulation performance parameters (slope and intercept, respectively, for the valence and arousal properties of affect processing) controlling for the subjects' age and sex. We found no significant group-level effects, however, all four measures of interest affected the measure of interest in the proposed direction: self-induced valence slope (β=.085, p=.46); self-induced arousal slope (β=.083; p=.43); self-induced valence intercept (β=.161; p=.41); and self-induced arousal intercept (β=.25; p=.17).

Discussion

This work made two important contributions to our current and future understanding of emotion processing and regulation. First, we found significant support for the utility of volitional positively valent affect processing as a mechanism for positively biasing the perceived affective valence of environmental stimuli. This finding causally and mechanistically supports the common notion of "positive thinking" and may provide deeper understanding of how and why attentional re-deployment strategies used in CBT benefit those suffering from deficits of emotion regulation and negatively biased affect. Second, we demonstrated a novel application of real-time brain state decoding in which we guided subjects' explicit emotion regulation toward a pre-defined affective goal state (positive valence) and then triggered experimental stimuli when the subjects' affective states fell within designed criteria representing that goal state. This new technology, while still in its infancy, may provide scientists with a much needed tool for causal exploration of intrinsic emotion processing mechanisms and their relationships with other cognitive processes and environmental factors.

A secondary consideration of this work was an attempt to explain individual differences observed in real-time fMRI guided explicit emotion regulation toward a known goal. Explicit affect regulation can be achieved volitionally, without the use of neurofeedback technology. Therefore, our use of real-time fMRI-based predictions of affect to guide (or focus) this innate process enabled us to test (using unguided explicit affect regulation ability as a baseline) the association between innate affect regulation performance and the performance achievable using our real-time fMRI feedback approach. Our null finding for this association suggests that existing explanations of performance[15–17] may apply to explicit affect regulation as well, as innate regulation ability, which varied widely over our sample, was not a contributing factor.

The application of neural decodings derived from IAPS image stimulus induction of affect processing as markers of perceived affect has well-known limitations, which we have noted in earlier reports [28,29,43]. Indeed, our validation process detected a significant negative effect of decoded arousal associated with decoded valence, suggesting that our cohort of subjects perceived the affective content of Mod-PS image stimuli differently than that which was captured

by the IAPS normative scores. However, the nature of our investigation – real-time moment-to-moment affect processing, regulation, and stimulus-triggering – does not, unfortunately, permit the use of subject self-reported measures of affect, thereby preventing us from achieving full concordance of our findings across cognitive, physiological, and behavioral domains. We also acknowledge technical limitations in our real-time fMRI approach. Despite significant findings of an overall effect, we believe that our implementation was suboptimal due both to latency as well as insufficient optimization of parameters within our real-time pipeline. A limitation of real-time approaches is that parametric choices in the processing pipeline (e.g., trigger threshold) interact with experimental outcomes; therefore, it is difficult to use batch-wise optimization to inform the design criteria *a priori*. Our small study sample did not permit sufficient piloting of parameters prior to fixing the processing design and testing. Further, our analysis included all rtfMRI-guided self-induction trials, even those that required emergency triggering due to a failure to meet the design criteria of the goal state. This was intentional in order to put forth the most conservative, and therefore reproducible, estimate of the valence self-induction effect sizes possible using this new technology. Therefore, we believe the performance of the system, and its effect sizes, are underreported, which suggests the potential to refine this technology for larger-scaled deployment of brain-state driven experiment designs to causally test interactions between internal cognitions and external stimuli.

Conclusion

We combined established neural decoding methods with real-time fMRI to construct a dynamic experiment in which the subject's self-induced positive affect state triggered the randomized presentation of affectively congruent or incongruent image stimuli. We first validated the experiment's ability to induce affect processing with independent measures of psychophysiology as well as the decoding models' ability to predict affect processing in novel task domains. We then demonstrated that self-induced positive affect states positively bias the perceived affect of subsequent image stimuli.

Acknowledgements

This study was funded by Brain and Behavioral Research Foundation NARSAD Young Investigator Award #26079 sponsored by the Families for Borderline Personality Disorder Research (K.A.B). Elements of the real-time fMRI infrastructure deployed in this work were supported by National Science Foundation grant BCS-1735820 (K.A.B). Additional personnel support was provided by National Institute on Drug Abuse grant 1T32DA022981 (C.D.K). Subject recruitment for the project was supported by the UAMS Translational Research Institute (TRI) through the National Center for Advancing Translational Sciences (1U54TR001629-01A1). The authors thank Kevin Fialkowski and Ivan Messias for their help in curating project data and Maegan Calvert for her thoughtful comments on the manuscript. The authors would also like to thank Kayla A. Wilson, Anthony A. Privratsky, Bradford S. Martins, Jennifer Payne, Emily Hahn, Natalie Morris, Nathan Jones, and Laura Spell for their assistance in recruiting and assessing research subjects and acquiring subject data as well as Stephen LaConte and Jonathan Lisinski for their assistance in developing our real-time fMRI capability. Finally, the authors thank Favrin Smith for her efforts in gaining the study's IRB protocol approval and maintaining human subject research compliance throughout the study's duration.

Authorship Contributions

Conception: K.A.B. Design, implementation, and testing: K.A.B.; Analysis: K.A.B; Interpretation of results, manuscript preparation, and revisions: K.A.B, C.D.K.

Competing Interests

The authors declare no competing interests.

Source Code and Data Availability

The authors have made the full source code used in this analysis publicly available: https://github.com/kabush/CTER. The authors have also made a Brain Imaging Data Structure[44] (BIDS) formatted variant of the full study dataset publicly available (as well as raw real-time log files and training materials) via the Open Science Framework: https://osf.io/yn4vq/. The source code used to convert raw data files to BIDS format has also been made publicly available: https://github.com/kabush/CTER2bids.

References

1. Boden, M. T., Thompson, R. J., Dizén, M., Berenbaum, H. & Baker, J. P. Are emotional clarity and emotion differentiation related? *Cognition & Emotion* **27**, 961–978 (2013).
2. Berking, M. & Wupperman, P. Emotion regulation and mental health: recent findings, current challenges, and future directions. *Current Opinion in Psychiatry* **25**, 128–134 (2012).
3. Kessler, R. C., Chiu, W. T., Demler, O. & Walters, E. E. Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**, 617 (2005).
4. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). (1994).
5. Butler, A., Chapman, J., Forman, E. & Beck, A. The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review* **26**, 17–31 (2006).
6. McRae, K. *et al.* The Neural Bases of Distraction and Reappraisal. *Journal of Cognitive Neuroscience* **22**, 248–262 (2010).
7. Flaisch, T., Junghöfer, M., Bradley, M. M., Schupp, H. T. & Lang, P. J. Rapid picture processing: Affective primes and targets. *Psychophysiology* **0**, 071003012229006-??? (2007).
8. MacNamara, A., Foti, D. & Hajcak, G. Tell me about it: Neural activity elicited by emotional pictures and preceding descriptions. *Emotion* **9**, 531–543 (2009).
9. Wu, L., Winkler, M. H., Andreatta, M., Hajcak, G. & Pauli, P. Appraisal frames of pleasant and unpleasant pictures alter emotional responses as reflected in self-report and facial electromyographic activity. *International Journal of Psychophysiology* **85**, 224–229 (2012).
10. Fujimura, T., Katahira, K. & Okanoya, K. Contextual Modulation of Physiological and Psychological Responses Triggered by Emotional Stimuli. *Front. Psychol.* **4**, (2013).
11. Czekóová, K., Shaw, D. J., Janoušová, E. & Urbánek, T. It's all in the past: temporal-context effects modulate subjective evaluations of emotional visual stimuli, regardless of presentation sequence. *Frontiers in Psychology* **6**, (2015).

12. Gross, J. J. The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology* **2**, 271–299 (1998).
13. Weiskopf, N. *et al.* Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. *NeuroImage* **19**, 577–586 (2003).
14. Blankertz, B. *et al.* Neurophysiological predictor of SMR-based BCI performance. *NeuroImage* **51**, 1303–1309 (2010).
15. Kober, S. E., Witte, M., Ninaus, M., Neuper, C. & Wood, G. Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. *Frontiers in Human Neuroscience* **7**, (2013).
16. Halder, S. *et al.* Prediction of Auditory and Visual P300 Brain-Computer Interface Aptitude. *PLoS ONE* **8**, e53513 (2013).
17. Witte, M., Kober, S. E., Ninaus, M., Neuper, C. & Wood, G. Control beliefs can predict the ability to up-regulate sensorimotor rhythm during neurofeedback training. *Frontiers in Human Neuroscience* **7**, (2013).
18. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (2008).
19. Oldfield, R. The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia* **9**, 97–113 (1971).
20. Brownell, R. Receptive One-Word Picture Vocabulary Test-2000. (2000).
21. Baddeley, A. D. & Hitch, G. Working Memory. *Psychology of Learning and Motivation* **8**, 47–89 (1974).
22. Beck, A., Steer, R. & Brown, G. Manual for the Beck Depression Invetory-II. (1996).
23. Bernstein, D. P. *et al.* Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect* **27**, 169–190 (2003).
24. Gratz, K. L. & Roemer, L. Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of psychopathology and behavioral assessment* **26**, 41–54 (2004).
25. Costa, P. & McCrae, R. Revised NEO Personality Inventory (NEO-P-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. (1992).
26. Spielberger, C. Manual for the state-trait anxiety inventory (form Y). (1983).
27. Zielinski, M. J. & Veilleux, J. C. The Perceived Invalidation of Emotion Scale (PIES): Development and psychometric properties of a novel measure of current emotion invalidation. *Psychological Assessment* ((in press)).
28. Bush, K. A. *et al.* Brain States That Encode Perceived Emotion Are Reproducible but Their Classification Accuracy Is Stimulus-Dependent. *Frontiers in Human Neuroscience* **12**, (2018).
29. Bush, K. A., Privratsky, A., Gardner, J., Zielinski, M. J. & Kilts, C. D. Common Functional Brain States Encode both Perceived Emotion and the Psychophysiological Response to Affective Stimuli. *Scientific Reports* **8**, (2018).
30. Bush, K. A., James, G. A., Privratsky, A. A., Fialkowski, K. P. & Kilts, C. D. An action-value model explains the role of the dorsal anterior cingulate cortex in performance monitoring during affect regulation. 23.
31. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research* **29**, 162–173 (1996).

32. Bernhard E. Boser, Isabelle M. Guyon & Vladamir N. Vapnik. A Training Algorithm for Optimal Margin Classifiers. in *Proceedings of the fifth annual workshop on Computational Learning* 144–152 (1992).
33. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111–116 (2019).
34. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* **59**, 2142–2154 (2012).
35. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, 320–341 (2014).
36. Rissman, J., Gazzaley, A. & D'Esposito, M. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* **23**, 752–763 (2004).
37. Platt, J. C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. in *Advances in Large Margin Classifiers.* (MIT Press, 1999).
38. Bush, K. A., Privratsky, A. A. & Kilts, C. D. Predicting Affective Cognitions in the Resting Adult Brain. in *Proceedings of the Conference on Cognitive Computational Neuroscience* (2018). doi:10.32470/CCN.2018.1010-0.
39. Bradley, M. M., Codispoti, M., Cuthbert, B. N. & Lang, P. J. Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion* **1**, 276–298 (2001).
40. Lang, P. J., Greenwald, M. K., Bradley, M. M. & Hamm, A. O. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* **30**, 261–273 (1993).
41. Heller, A. S., Greischar, L. L., Honor, A., Anderle, M. J. & Davidson, R. J. Simultaneous acquisition of corrugator electromyography and functional magnetic resonance imaging: A new method for objectively measuring affect and neural activity concurrently. *NeuroImage* **58**, 930–934 (2011).
42. Baucom, L. B., Wedell, D. H., Wang, J., Blitzer, D. N. & Shinkareva, S. V. Decoding the neural representation of affective states. *NeuroImage* **59**, 718–727 (2012).
43. Wilson, K. A., James, G. A., Kilts, C. D. & Bush, K. A. Combining Physiological and Neuroimaging Measures to Predict Affect Processing Induced by Affectively Valent Image Stimuli. *Sci Rep* **10**, 9298 (2020).
44. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* **3**, 160044 (2016).

Appendix

Table 2: *International Affective Picture Set Image Identification Numbers*

Trial Type	Identification Numbers
Mod-PS	8510, 9421, 3350, 7502, 9908, 3266, 3061, 6821, 9910, 3140, 2799, 2717, 8420, 7230, 3168, 2800, 8503, 4520, 9253, 5460, 9250, 5830, 4608, 8380, 9901, 2208, 2160, 9400, 7260, 5825, 8300, 4660, 4640, 8210, 9500, 9040, 7405, 9412, 2075, 3131
Mod-FS	8034, 9419, 2071, 8490, 9570, 9300, 5480, 3301, 9830, 8170, 4680, 3215, 9183, 8370, 3225, 9921, 3064, 4599, 7350, 5450