

**Neural predictors of decisions to cognitively control emotion**

Bruce Pierre Doré, Jochen Weber, and Kevin Nicholas Ochsner

Department of Psychology, Columbia University

Contact:

Bruce Doré  
Kevin Ochsner  
Department of Psychology  
Columbia University  
Schermerhorn Hall  
New York, NY 10027, USA  
[brucedore@gmail.com](mailto:brucedore@gmail.com)  
[ochsner@psych.columbia.edu](mailto:ochsner@psych.columbia.edu)

The authors declare no competing financial interests.

Acknowledgments: We thank Alexa Hubbard, Chelsea Boccagno, Daisy Burr, and Kan Long for assistance with data collection. Completion of this work was supported by NIA grant R01AG043463 and Conte grant PAR-11-126 awarded to K.N.O.

**Abstract**

Deciding to control emotional responses is a fundamental means of responding to environmental challenges, but little is known about the neural mechanisms that predict the outcome of such decisions. We used fMRI to test whether human brain responses during initial viewing of negative images could be used to predict decisions to regulate affective responses to those images. We found that 1) decisions to regulate were more frequent in individuals exhibiting higher average levels of activity within the amygdala and regions of prefrontal cortex known *a priori* to be involved in the cognitive control of emotion and 2) within-person expression of a distributed brain pattern associated with regulating emotion predicted choosing to regulate responses to particular stimuli, beyond the predictive value of stimulus intensity or self-reports of emotion. These results demonstrate the behavioral relevance of variability in brain responses to aversive stimuli and provide a model that leverages this variability to predict behavior.

**Significance statement**

Everyone experiences stressors, but how we respond to them can range from protracted disability to resilience and growth. One key process underlying this variability is the agentic decision to exert control over emotional responses. We present an fMRI-based model predicting decisions to control emotion, finding that activity in brain regions associated with the generation and regulation of emotion were predictive of which people choose to regulate frequently, and a distributed brain pattern associated with regulating emotion was predictive of for which stimuli regulation was chosen. These brain variables predicted future decisions to regulate emotion beyond what could be predicted from stimulus and self-report variables.

Distressing events are unavoidable, but how we respond to them can be a matter of deliberate choice. Converging evidence suggests that one such choice – the choice to effortfully regulate one's negative emotions, as opposed to letting them unfold naturally – serves a critical protective function for well-being (Major et al., 1998; Russo et al., 2012; Sheppes et al., 2014). Although dozens of imaging studies have focused on the brain systems supporting the regulation of emotion (reviewed in Buhle et al., 2014; Ochsner, et al., 2012), none have given participants the choice as to whether or not they will regulate their emotions, instead instructing participants when to regulate vs. respond naturally. As such, the neural processes supporting agentic decisions to regulate emotional responses are unknown.

We sought to build a predictive model of these decisions as a step toward a neuroscientific understanding of the different ways individuals respond to aversive life experiences (see Russo et al., 2012; Chang et al., 2015). We began with the idea that specific brain processes measured by neuroimaging – here, those associated with the generation and regulation of emotion – could be used to predict behavioral outcomes that depend on engagement of the same or similar brain processes (see Berkman and Falk, 2013). This led us to ask whether brain activity measured during initial uninstructed encounters with affectively charged events – when people are reacting and/or engaging regulatory processes in an uninstructed manner – could predict subsequent

choices to regulate one's emotional responses to those events, when the choice to regulate is presented explicitly.

We focused on *reappraisal*, a regulation strategy that entails thinking differently about a negative stimulus in order to change how one feels about it (e.g., looking for a potential bright side or otherwise taking a new perspective) (see Ochsner et al., 2012). Instructed implementation of reappraisal reliably increases activity within a network of regions implicated in cognitive control, including ventrolateral prefrontal (vIPFC), dorsolateral prefrontal (dIPFC), and dorsomedial prefrontal (dmPFC) cortices, and can influence (i.e., both up- and down-regulate) activity in the amygdala, a subcortical brain region involved in triggering affective responses (see Buhle et al., 2014; Phelps and Ledoux, 2005). Building on these prior neuroimaging studies, we derived predictor variables reflecting brain activity triggered by an initial encounter with a stimulus, asking whether these variables could predict subsequent decisions to regulate one's emotional responses to that stimulus. Predictors included 1) activity in the amygdala, thought to reflect emotional reactivity to a stimulus 2) activity in vIPFC, dIPFC, and dmPFC regions, thought to reflect controlled processing and/or regulation of emotional responses to a stimulus and 3) a whole-brain pattern reflecting the global network of activity associated with implementing reappraisal to either up- or down-regulate emotion (see van Ast et al., 2016; Woo et al., 2015).

To address these questions, we first trained participants in what reappraisal is and how to use it – ensuring they were knowledgeable about what

a choice to regulate emotion would entail. Next, we used functional magnetic imaging (fMRI) to measure brain responses during a negative image viewing task, when participants were free to think about the images in any way they chose. After scanning, we presented these negative images – as well as new ones – and asked participants to decide whether to regulate their emotions or simply view the images. This design allowed us to test the hypothesis that variability in brain responses associated with emotional reactivity and/or emotion regulation evoked during the negative image viewing task would be predictive of these subsequent emotion regulation decisions, above and beyond stimulus and self-report variables. This hypothesis was tested at both (1) the level of the person, asking whether we could predict the individuals for whom decisions to regulate would be most likely (i.e., their brain activity would predict more frequent decisions to regulate not just for previously seen stimuli, but for new stimuli as well), and (2) at the level of the stimulus, asking whether we could predict the events for which decisions to regulate would be most likely.

## **Materials and Methods**

### **Participants**

Participants were 20 adults (12F, 8M) recruited from the New York City area (mean age = 24.6, SD=4.5), and screened to confirm that they were right-handed, could read and speak fluently in English, had normal or corrected-to-normal vision, had never been diagnosed with a psychiatric disorder, did not report current depressive symptoms (i.e., scored below 16 on the Center for Epidemiologic Studies Depression scale), and had no conditions that

contraindicated magnetic resonance imaging. Informed consent was obtained according to procedures approved by the Columbia University Institutional Review Board.

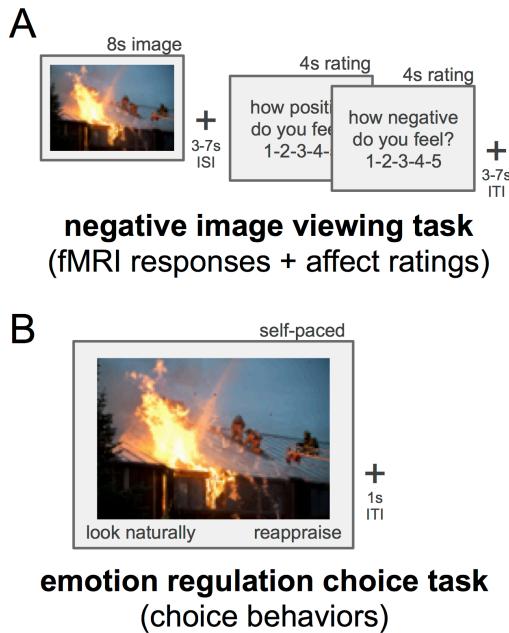
### **Image acquisition**

Data were acquired on a 3T GE MR750 whole-body scanner with a 32-channel RF head coil. Structural volumes were acquired using a high-resolution T1-weighted sagittal 3D BRAVO sequence yielding 1mm<sup>3</sup> isotropic voxel size. Functional volumes were acquired using a T2\*-sensitive echo planar imaging EPI sequence with a repetition time TR of 2000ms, an echo time TE of 25ms, a 77° flip angle, and a 19.2cm FOV consisting of 45 interleaved 3mm slices acquired parallel to the AC-PC axis.

### **Design**

*Emotion regulation training/practice.* Immediately before scanning, all participants completed experimenter-guided emotion regulation training modules, which included training in positive reappraisal (i.e., focusing on potential positive aspects or outcomes of a negative situation) and minimizing reappraisal (i.e., focusing on potential neutral aspects or outcomes of a negative situation) strategies (see McRae et al., 2012). In the scanner, they applied these strategies within an instructed reappraisal task (not of direct interest here). Such training and practice in reappraisal helped ensure that participants would be knowledgeable about what a choice to regulate emotion (or not to regulate) would entail.

*Scanner negative image viewing task.* After the instructed task, participants were informed that they would be viewing images and were asked to attend to and rate their responses to these images, but were not instructed to think about the images in a particular way (i.e., they were not instructed to regulate their emotional responses to the images). They then completed the negative image



**Figure 1. A)** Negative image viewing task, completed in scanner, and **B)** emotion regulation choice task, completed post in a post-scan session. Predictive models use brain responses in the negative image viewing task to predict participant choice behaviors in the emotion regulation choice task.

viewing task, which consisted of two runs of 10 trials each. The fact that participants had received prior training in and experience with reappraisal, ensured that they knew what reappraisal was and how to do it – and critically – were free to choose to engage with stimuli in this task in a way that could reflect

agentic decisions to reappraise. Figure 1A shows the trial sequence for this task, consisting of image viewing period, inter-stimulus interval (ISI), affect rating period (positive and negative affect ratings appeared in a randomized order), and inter-trial interval (ITI). Images (mean normative valence=2.49; mean normative arousal=5.71) were selected from the International Affective Picture System (Lang et al., 2008), and were counterbalanced to block and randomly assigned to trial number. Images depicted instances of illness and injury, human and animal waste, acts of aggression, members of hate groups, and transportation accidents. For the affect ratings, we asked participants to base their ratings on how negative and positive they felt at the end of the image viewing period. Stimuli were presented with E-Prime 1.2 (Psychology Software Tools), and participants made behavioral responses on a five-button response pad.

*Emotion regulation choice task.* Immediately after leaving the scanner, participants completed a surprise final task in which they viewed in a random order the 20 negative images presented in the task plus 20 novel negative images matched on content, arousal, and valence and were asked to choose whether they would prefer to 1) regulate their emotional response to the image with reappraisal, or 2) simply look at the image without reappraising (see Figure 1). Stimuli were presented with E-Prime 1.2 (Psychology Software Tools), and participants made behavioral responses on a keyboard.

### **fMRI Analyses**

*Preprocessing/GLM.* Data were preprocessed with SPM8 (Wellcome Department of Cognitive Neurology, UCL), and consisted of slice-time correction,

realignment, coregistration of functional and structural images, and normalization to the standard Montreal Neurological Institute (MNI) brain by segmentation of the structural image and applying the parameters from this step during warping. Normalized images were interpolated to 3mm<sup>3</sup> voxels and smoothed with a 6mm Gaussian kernel.

First-level (individual participant) GLM analyses were implemented in NeuroElf v1.0 ([neuroelf.net](http://neuroelf.net)). Stimulus and response periods of each trial were modeled as boxcar functions convolved with the canonical hemodynamic response function. Motion parameters and a high-pass filter for 128 seconds were included as regressors of no interest. All analyses focused on brain signal estimated during the stimulus (image viewing) period of each trial. Average brain activity during this stimulus (image viewing) period was defined as the parameter estimate for this stimulus (image viewing) regressor.

Second-level (group) random-effects analyses were implemented in NeuroElf v1.0. All brain coordinates are reported in standard MNI space. For our follow-up whole-brain analysis significant voxels were identified using a joint height ( $p=.0025$ ) and extent ( $k=103$ ) threshold determined by AlphaSim, using smoothness parameters estimated from the residuals of the statistical map (11.7mm).

*Regions of interest.* Regions of interest (ROIs) were constructed for bilateral amygdala, and bilateral vIPFC , dIPFC, and dmPFC – all regions known to be important for reappraisal (Ochsner et al., 2012). The amygdala ROI was defined anatomically from the Harvard-Oxford anatomical atlas for 25%

probability (L -23, -5, -18; R 23, -4, -18) ( $5324 \text{ mm}^3$ ), and three ROIs for bilateral vIPFC (L -51, 21, 9; R 60, 24, 3) ( $14121 \text{ mm}^3$ ), bilateral dIPFC (L 51, 15, 48; R -33, 3, 54) ( $8235 \text{ mm}^3$ ), and a region spanning bilateral dmPFC (9, 30, 39) ( $8343 \text{ mm}^3$ ) were constructed from the results of a meta-analysis from our lab of 48 neuroimaging studies of emotion regulation (Buhle et al., 2014). We defined these prefrontal ROIs directly from the statistical map resulting from this meta-analysis (provided by the authors), by selecting the clusters of contiguous voxels in vIPFC, dIPFC, and dmPFC that achieved whole-brain significance in the meta-analytic contrast of reappraisal greater than natural response (reported in Table 1 of Buhle et al., 2014).

*Pattern expression analyses.* We conducted pattern expression analyses in order to test whether whole-brain responses to individual images could predict subsequent choices to reappraise those images. Previous studies using this approach have asked whether expression of a brain pattern associated with working memory is modulated by social threat (van Ast, et al., 2016) and whether a pattern predictive of physical pain is modulated by emotion regulation (Woo et al., 2015). Our analyses used the “single-trial” or “beta-series” approach (Koyama et al., 2003) to estimate brain responses for each trial of the negative image viewing task for each participant. We did this by building a GLM that included trial-specific regressors for each image presented in the viewing task in addition to single regressor for the response period, six regressors for motion parameters, and a high-pass filter for 128 seconds.

In order to calculate the extent to which trial-level beta images expressed the meta-analytic reappraisal pattern (from Buhle et al., 2014), we treated the unthresholded meta-analytic map as a pattern of weights, reflecting the degree to which each voxel is reliably associated in the extant literature with implementing reappraisal. We then calculated the dot-product of the activation image for each trial of the task for each person (*beta map*) with the unthresholded meta-analytic map (*weight map*), yielding a continuous scalar value (*beta map* · *weight map*), reflecting the extent to which each trial-level beta map expressed the weight map pattern. These values were mean-centered by participant to yield a measure of within-subject variation in reappraisal pattern expression relative to their average.

*Person- and trial-level prediction.* We used R ([cran.r-project.org](http://cran.r-project.org)) to implement person-level Poisson regression models (using `glm` from the ‘stats’ package) and multilevel logistic regression models (using `glmer` from the ‘lme4’ package) to test whether brain activity in our ROIs and expression of the whole-brain reappraisal pattern could predict counts of person-level choice behavior (i.e., the number of times each person chose to reappaise, from 0-40), and trial-level choices (coded as 0 – chose to look naturally without reappraising; 1 – chose to reappaise). Fitted multilevel models included parameters allowing model intercept and slopes to vary by participant when estimating effect sizes (see Barr et al., 2015), and, for model comparisons, varying slope parameters were included where supported by the data (see Bates et al., 2015). We implemented mediation analyses in R (using `mediate` from the ‘mediation’ package). Where noted, we adjusted for normative ratings of image intensity (i.e.,

both valence and arousal norms) (Lang et al., 2008), and self-report ratings of negative and positive affect by including these variables as covariates. All predictor variables were standardized, yielding as measures of effect size beta coefficients indicating the expected difference in the outcome variable across a difference of one standard deviation in the predictor. (Poisson coefficients are equal to the log of the incidence rate ratio, and logistic coefficients the log of the odds ratio, across a one-unit change in the predictor.)

An important consideration in these analyses is that trial-level estimates can be strongly affected by acquisition artifacts that occur during that trial (e.g., sudden motion, scanner pulse artifacts, etc.). For this reason, trials with an estimated Mahalanobis distance (across pattern expression and ROI variables)  $\pm 3$  standard deviations from the grand mean were excluded in multilevel models and when calculating participant averages (less than 2% of all observations).

*Estimating model predictive accuracy.* To estimate the out-of-sample predictive accuracy of our linear models, we approximated Bayesian leave-one-out cross validation using Pareto-smoothed importance sampling (LOO; Vehtari et al., 2016), fitting models with uniform priors using the Bayesian inference software platform Stan ([mc-stan.org](http://mc-stan.org); ver 2.10), and the ‘brms’ R package (Bayesian Regression Models using Stan ver 0.10.0). Instead of model re-fitting, as in exact cross-validation, the LOO procedure draws samples from posterior distributions of the model parameters in order to estimate expected log-likelihood for new data and thus adjust for over-optimism (i.e., bias) inherent to within-sample measures of model fit (e.g., the uncorrected log-likelihood). We used this

procedure to derive LOO-adjusted deviance values (LOOIC) that can be used to compare models in terms of their expected out-of-sample predictive accuracy. This is conceptually similar to comparing AIC scores (which also approximate a model's out-of-sample predictive accuracy), and, to a lesser extent, BIC scores (which approximate a model's marginal likelihood – the likelihood of observing the data given the model, marginalized across possible parameter values), which we examined as well (see Gelman et al., 2014). Lastly, we applied receiver operating characteristic (ROC) analysis (implemented with the 'ROCR' package in R) to assess predictive performance of our multilevel logistic regression models. In the ROC framework, model performance is expressed as the area under the curve (AUC) in a plot of the model's sensitivity (the proportion of reappraisal choices correctly predicted as such) against its specificity (the proportion of look naturally choices correctly predicted as such) across a range of prediction thresholds. The AUC can be interpreted as the probability that a randomly selected image/trial where reappraisal was actually chosen is predicted as more likely to be reappraised than a randomly chosen image/trial where reappraisal was not chosen. Thus, AUC represents a threshold-independent metric of model performance, with values from 0.5 (prediction at chance) to 1 (perfect prediction).

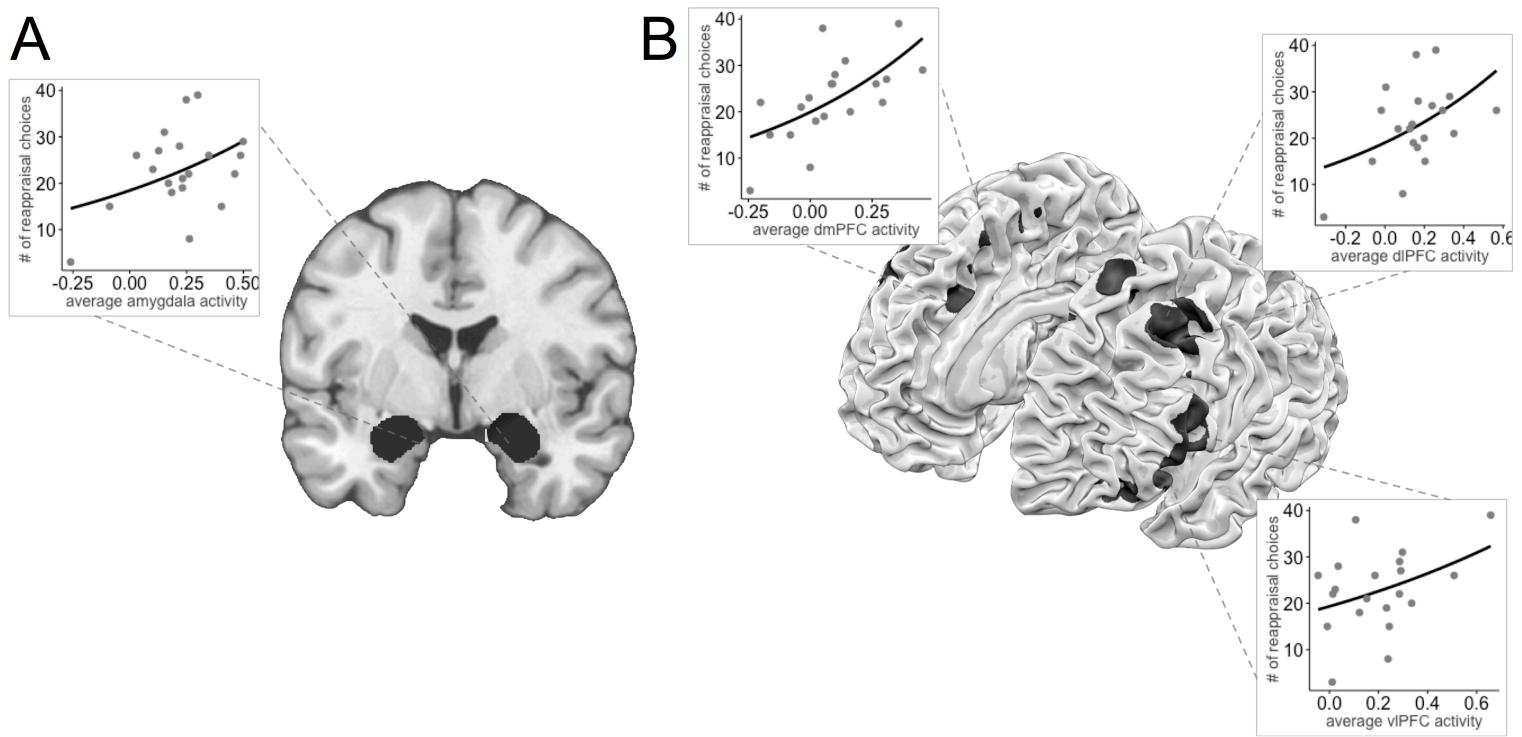
## Results

### **Brain responses in amygdala and prefrontal cortex predict person-to-person differences in emotion regulation choices**

*ROI-based prediction of person-level reappraisal choice frequencies.* In an initial analysis, we aggregated our data to the person level (i.e., computed choice frequencies and average brain activity estimates for each person) to run regression models asking whether person-to-person differences in activity within our a priori ROIs during viewing of negative images (when participants were reacting to the images in an uninstructed manner) could predict the number of times each participant subsequently explicitly chose to use reappraisal in the reappraisal choice task. As shown in Figure 2A and B, we found that more frequent reappraisal choices were predicted by greater activity in the amygdala,  $b=.17$ , 95%CI[.07, .27],  $p<.001$ , as well as vIPFC,  $b=.14$ , 95%CI[.05, .23],  $p=.002$ , dIPFC,  $b=.19$ , 95%CI[.09, .28],  $p<.001$ , and dmPFC,  $b=.24$ , 95%CI[.15, .33],  $p<.001$ . These relationships held when estimating them with robust generalized linear models (via iteratively reweighted least squares). Each of these relationships also remained nonzero when adjusting for self-report ratings of positive and negative affect, amygdala  $b=.20$ , 95%CI[.10, .32],  $p<.001$ , vIPFC,  $b=.10$ , 95%CI[.01, .20],  $p=.03$ , dIPFC,  $b=.16$ , 95%CI[.06, .27],  $p=.001$ , and dmPFC,  $b=.23$ , 95%CI[.13, .33],  $p<.001$ , indicating that greater responses in these regions during viewing of negative images predicted emotion regulation choice frequency independent of affective experience while viewing images in the scanner.

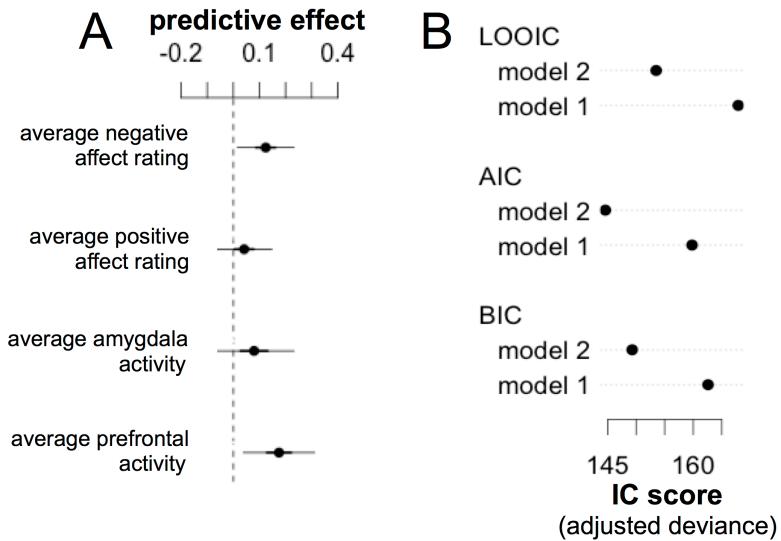
*Prediction of decisions for old versus novel images.* In the reappraisal choice task, participants made decisions about whether to regulate responses to 20 old images (that had been seen previously in the scanner) and 20 new

images. Reappraisal choice frequencies were highly correlated across old and new images,  $r = .85$ , 95%CI[.65, .94]. The ability of brain activity during the image viewing task to predict subsequent reappraisal choices was comparable across old and new images for the amygdala,  $b_{\text{old}} = .15$ , 95%CI[.01, .29],  $p=.04$ ,  $b_{\text{new}} = .19$ , 95%CI[.05, .33],  $p=.007$ , and for the prefrontal cortex,  $b_{\text{old}} = .22$ , 95%CI[.09, .35],  $p<.001$ ,  $b_{\text{new}} = .22$ , 95%CI[.09, .34],  $p<.001$ . These results suggest that the person-to-person predictive value of these ROIs was comparable for previously seen and never before seen images, which together indexed individual differences in the general tendency to choose to regulate.



**Figure 2.** Scatterplots and generalized linear model fits reflecting predictive relationships between person-to-person variability in brain activity in the negative image viewing task and subsequent reappraisal choice frequencies, for a priori **A**) anatomically-defined amygdala, and **B**) meta-analytically defined dmPFC, dIPFC, and vIPFC bilateral regions of interest.

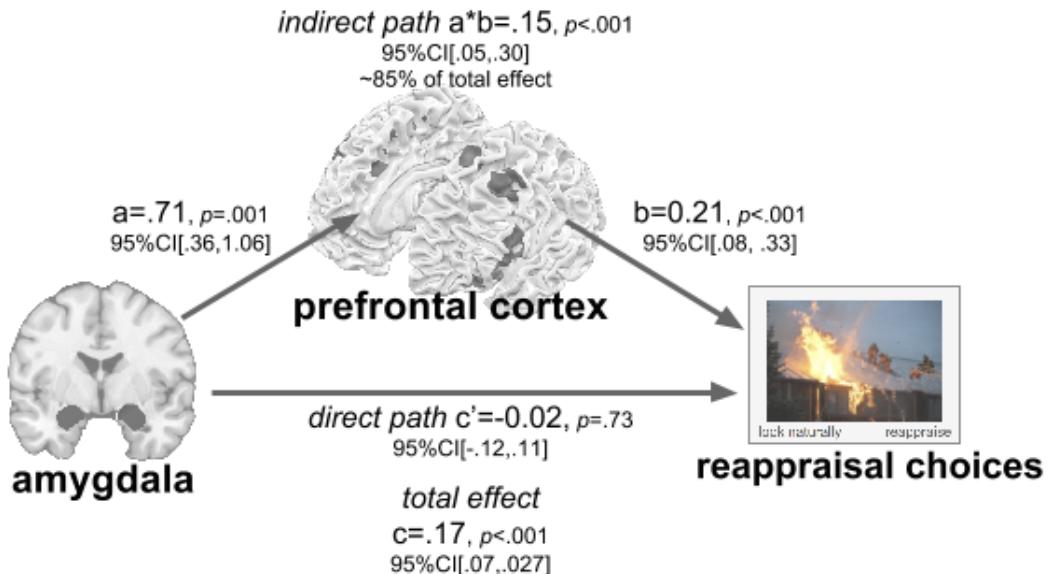
*Correlation structure of the regions of interest.* Next, we inspected the correlation structure of these ROIS, finding that activity in vIPFC, dIPFC, and dmPFC was highly correlated (mean  $r=0.83$ ) and also correlated with activity in the amygdala (mean  $r=0.67$ ). To reduce model complexity, and to reflect correspondence with a model of regulation in which different prefrontal regions are components of a coordinated system for cognitively controlling emotion (see Ochsner et al., 2012), we averaged our vIPFC, dIPFC, and dmPFC ROI variables into a compound prefrontal ROI variable for all subsequent analyses.



**Figure 3. A)** Plot of regression coefficients (with 95%CI) for the full model (model 2) predicting reappraisal choice frequencies from brain and self-report variables (entered as simultaneous predictors). **B)** Model comparison metrics for full model (model 2), and reduced model including only negative affect ratings and positive affect ratings (model 1). Model fit is summarized by leave-one-out cross-validated error (LOOIC), AIC, and BIC information criterion scores (i.e., adjusted model deviance, a measure of the relative quality of the models for these data). A lower number indicates better fit.

*Comparing predictive fit of models.* A crucial question is whether including our amygdala and prefrontal brain variables improves the prediction power of our models, relative to models with only self-report measures of emotion. To address this question, we compared models with and without brain predictors in terms of

predictive accuracy estimated by cross-validation (Vehtari et al., 2016; Gelman et al., 2014). As shown in Figure 3, a model that included the predictors for brain activity (i.e., both the compound prefrontal ROI and the amygdala ROI) showed substantially better predictive fit by leave-one-out (LOO) cross-validation ( $M_{full}=153.5$ ) than a reduced model including only self-report affect ratings ( $M_{reduced}=167.9$ ), and was also preferred by Akaike Information Criterion (AIC) ( $M_{full}=144.6$ ;  $M_{reduced}=159.8$ ) and Bayesian Information Criterion (BIC) ( $M_{full}=149.3$ ;  $M_{reduced}=162.6$ ) metrics (see Figure 3A). This indicates that the model with brain predictors was identified as higher in expected out-of-sample accuracy (by LOO and AIC), and a more plausible model of the data generation process (by BIC), when compared to the model including only self-reports of negative and positive affect. (Because all participants viewed the same images, stimulus qualities did not differ from person to person.)



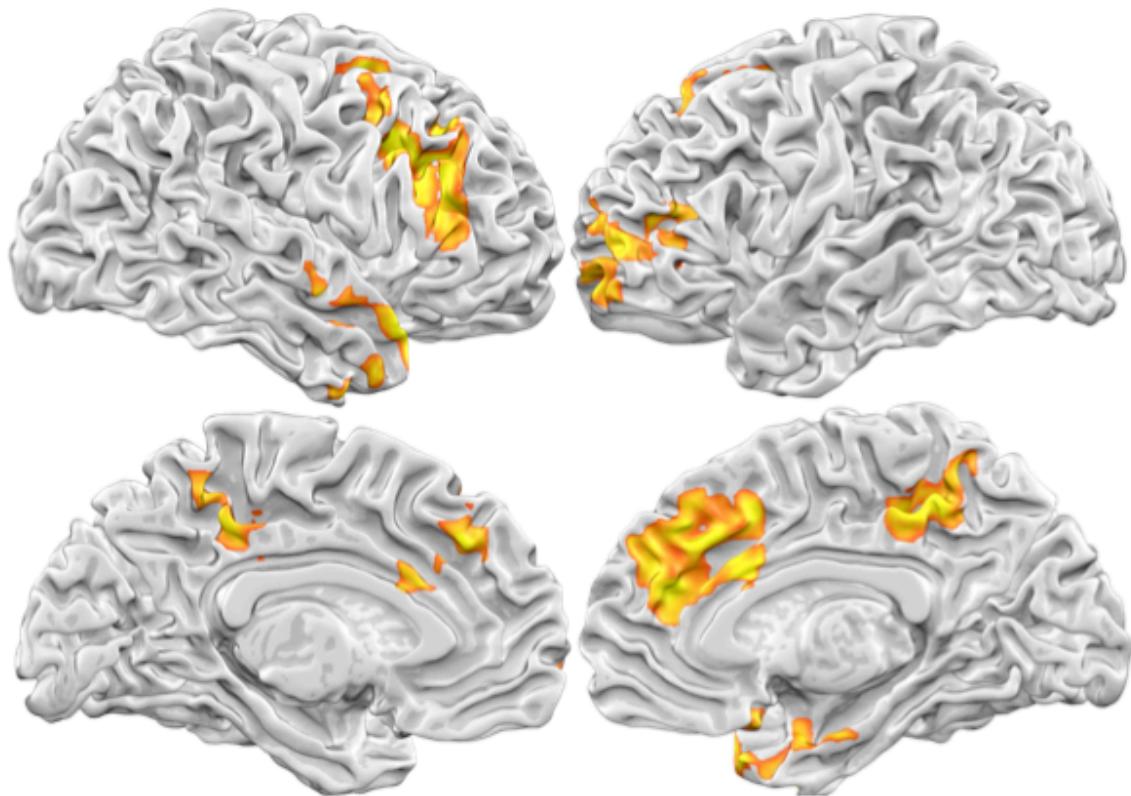
**Figure 4.** Greater average amygdala activity in the negative image viewing task predicts subsequently choosing to regulate emotion more frequently, and this relationship is mediated by greater prefrontal cortex activity during the image viewing task.

*Person-level mediation analysis.* We conducted a follow-up mediation analysis to test the evidence for a causal model whereby the effect of amygdala activity (the x variable) on subsequent emotion regulation choice frequencies (the y variable) is transmitted via changes in prefrontal cortex activity (the mediator variable). We found a significant total effect of amygdala activity on reappraisal choice frequencies,  $c = .17$ , 95%CI[.07, .27],  $p < .001$  that was fully mediated by prefrontal cortex activity,  $a^*b = .15$ , 95%CI[0.05, 0.30],  $p < .001$ , such that the direct effect of amygdala activity on reappraisal choice frequencies dropped to near zero,  $c' = 0.02$ , 95%CI[-.12, 0.11],  $p = .73$  when adjusting for prefrontal activity (see Figure 4). Although these variables were observed (not manipulated), the results of this mediation analysis are consistent with a causal model whereby greater amygdala reactivity elicits greater recruitment of prefrontal regions, which in turn leads to and/or reflects more frequent decisions to regulate emotion.

#### ***Follow-up whole-brain analysis for reappraisal choice frequencies***

To complement the primary ROI-based analyses described in the main text, we conducted a follow-up whole-brain analysis in order to identify regions of the brain that were most strongly correlated with future emotion regulation decisions at the person-to-person level. Unlike the ROI analyses, which were designed to estimate effect sizes for the predictive relationship between brain activity and reappraisal choices for given brain regions, this analysis was

designed to identify any regions across the whole brain that were highly correlated with future choice frequencies. Consistent with the ROI-based analyses, we found clusters that positively correlated with reappraisal choice frequency within bilateral vIPFC [L -39, 36,3; R 57,36,12], bilateral dIPFC [L -21,21,39; R 42,18,33], and bilateral dmPFC [3, 33, 36], as well as two additional regions: bilateral precuneus [9, -33, 39] and right anterior temporal lobe [R 33,21,-33] (see Figure 5).



**Figure 5.** Regions identified in a whole brain analysis for which person-to-person variability in activity during the negative image viewing task was most correlated with subsequent reappraisal choices (whole-brain FWE  $p < .05$ ).

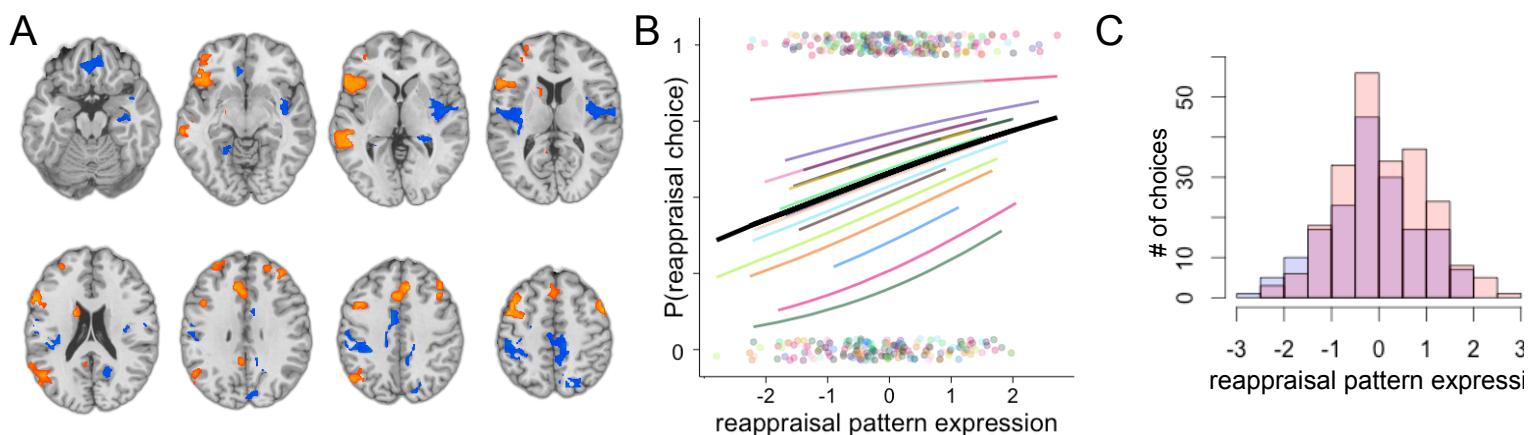
**Expression of a distributed brain pattern associated with reappraisal predicts stimulus-to-stimulus emotion regulation choices**

Our initial analyses showed that average levels of brain activity in specific ROIs could predict person-to-person variability in reappraisal choices (i.e., whether people will reappraise frequently or infrequently), but they did not model stimulus-level variability in reappraisal choices (i.e., whether patterns of brain activity can be used to predict for which events people are more likely to reappraise). To address this question we ran multilevel logistic regression models including brain activity measured during the image viewing task as predictors, and trial-level decisions to reappraise for specific stimuli in the subsequent choice task as the outcome.

*ROI-based prediction of stimulus-level reappraisal choices.* We also used multilevel logistic regression to evaluate the predictive value of the amygdala and prefrontal ROI variables. Consistent with the single-level models reported above, considering only the between-subject component of the ROI variables (i.e., average activity within the ROI across all trials for a given person) we found that average activity within the amygdala,  $b=.56$ , 95%CI[.15, .97],  $p=.007$ , and within prefrontal control regions,  $b=.75$ , 95%CI[.30, 1.19],  $p=.001$ , was predictive of future choices. However, considering only the within-subject component of these variables (i.e., trial-to-trial deflections from these overall person averages), there was no trial-by-trial relationship for the amygdala,  $b=-.06$ , 95%CI[-.49, .34],  $p=.63$ , but there was a trend-level positive relationship for the prefrontal regions,  $b=0.20$ , 95%CI[-.01, .42],  $p=.07$ . This suggests that, in contrast to the pattern expression variable, these ROI variables may be lower in predictive value for trial-to-trial decisions.

*Pattern-based prediction of stimulus-level reappraisal choices.* As a more global representation of reappraisal-related brain activity occurring while viewing each image, we used trial-by-trial expression of the meta-analytically derived whole-brain pattern associated with implementing reappraisal (i.e., the whole-brain map resulting from the Buhle et al., 2014 meta-analysis).

Using the data from the image-viewing task, we asked whether expression of this whole-brain pattern of interest could prospectively predict decisions to reappraise. We found that the trial-to-trial differences in pattern expression predicted greater probability of choosing to regulate emotion,  $b=.31$ ,  $95\%CI[.07,$

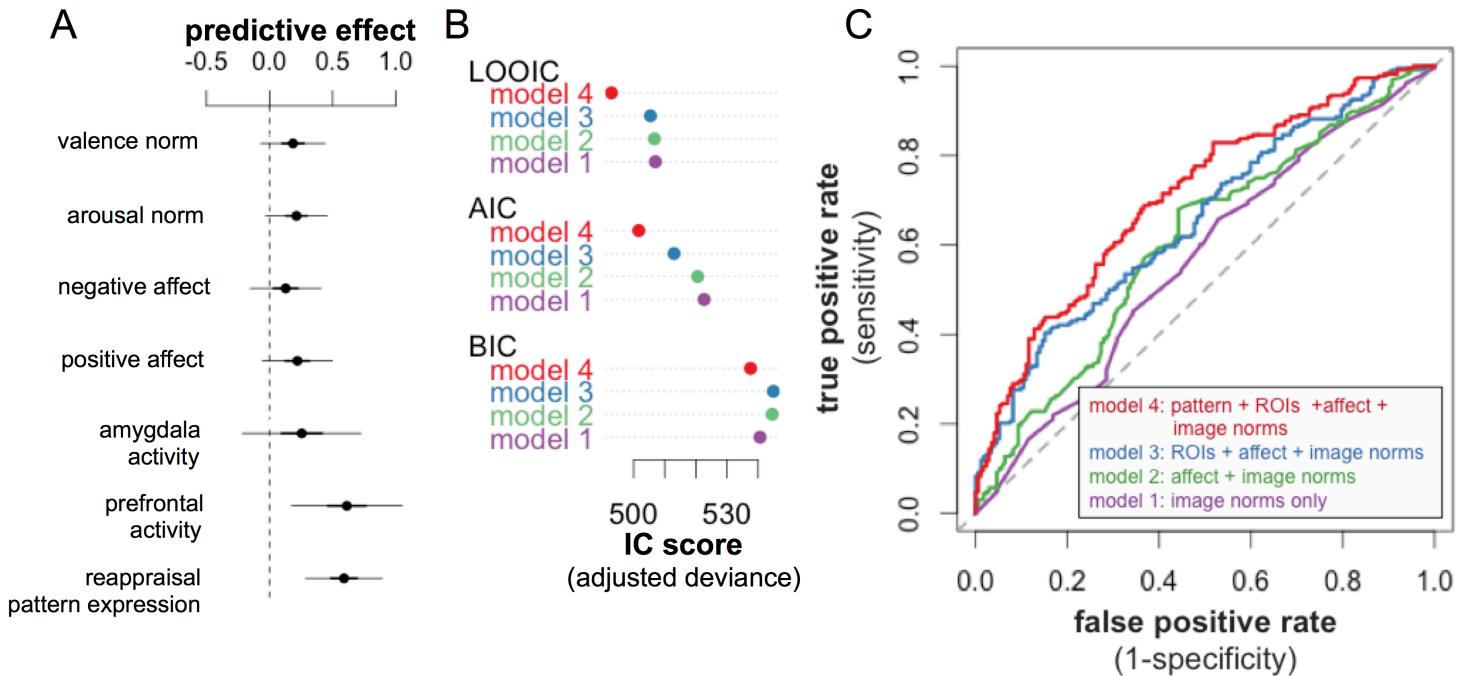


**Figure 6.** **A)** Meta-analytically derived whole-brain pattern associated with regulating emotion via reappraisal (display is thresholded at  $z>2.6$ ,  $k=20$ , but all voxels were used in analyses). **B)** Trial-to-trial variability in expression of this pattern ( $z$ -transformed) is predictive of subsequently choosing to use reappraisal for particular images. **C)** Histograms of all reappraisal choices (in pink) and all natural response choices (in blue), with choices from all participants pooled into a single distribution (per choice outcome).

.55],  $p=.009$  (see Figure 6). This predictive relationship held when adjusting for self-report ratings of positive and negative affect and normative ratings of image valence and arousal,  $b=.31$ ,  $95\%CI[.07, .42]$ ,  $p=.01$ , and also held (actually increasing in magnitude),  $b=.59$ ,  $95\%CI[.29, .89]$ ,  $p<.001$ , when additionally

adjusting for activity within our individual prefrontal cortex and amygdala ROIs. In this overall model, only the prefrontal cortex,  $b=.61$ ,  $95\%CI[.18, 1.04]$ ,  $p=.006$ , independently predicted choosing to reappraise, beyond the predictive capacity of the pattern expression variable (see Figure 7A).

*Comparing predictive fit of models.* A full model, including the pattern expression variable, stimulus-level estimates of activity in the prefrontal and amygdala ROIs, self-report ratings, and image intensity norms, showed substantially better predictive fit by LOO cross-validation ( $M_{full} = 492.9$ ) than a model predicting choices from only self-report affect ratings and normative ratings of image valence and arousal ( $M_{reduced} = 505.4$ ) and was also preferred by AIC ( $M_{full} = 501.7$ ,  $M_{reduced} = 513.0$ ) and BIC ( $M_{full} = 537.6$ ,  $M_{reduced} = 544.9$ ) metrics. This indicates that the model including brain predictors was higher in expected out-of-sample accuracy (by LOOIC and AIC) and a more plausible model of the data generation process (by BIC), than the reduced model including only image characteristics and self-report (see Figure 7B).



**Figure 7.** **A)** Plot of regression coefficients (with standard error and 95%CI) for the full model (model 4) predicting reappraisal choices from brain, self-reports of affect, and normative ratings of the images (entered as simultaneous predictors) **B)** Model comparison metrics for full model (model 4), and reduced models with only ROIs, affect, and image norms (model 3), affect and image norms (model 2), and image norms only (model 1) **C)** ROC curves depicting prediction accuracy for models 1,2,3, and 4 in terms of sensitivity (correct prediction of reappraisal choices) and specificity (correct prediction of natural response choices), across a range of possible prediction thresholds.

Next, we used a receiver operating characteristic (ROC) analysis to quantify the absolute predictive performance of our models (i.e., the models' ability to correctly predict reappraisal choices versus natural response choices). Shown in Figure 7C, the model with only normative ratings of image arousal and valence as predictors had an AUC of 0.56, corresponding to poor prediction (shown in purple). The model with normative ratings and participant positive and negative affect ratings collected in the scanner had an AUC of 0.61 (shown in green). The model with prefrontal and amygdala ROI predictors in addition to

image norms and affect ratings had an AUC of 0.65 (shown in blue). Finally, the model that included expression of the meta-analytically defined whole-brain reappraisal pattern in addition to these variables had an AUC of 0.71. At a prediction threshold of 0.5 (i.e., a predicted probability  $>0.5$  is considered predicted reappraisal, and  $<0.5$  is considered predicted natural responding), this model showed 70% correct prediction of participant choice behaviors.

## Discussion

In order to make contact with translational applications, neuroimaging studies must go beyond mapping correlates of experimentally-cued regulation to begin constructing neuroscience-informed predictive models that can forecast which people will choose to regulate their emotions and for which events they will choose to do so (see Doré et al., 2016). Here, we provide the first example of such a model, leveraging variability in brain responses to negative images to predict agentic decisions to regulate emotion.

Two key findings were obtained. First, at the level of the individual, we found that greater activity in the amygdala (a region involved in generating emotion), and in vIPFC, dIPFC and dmPFC (regions involved in controlling emotion) predicted more frequently choosing to regulate responses to emotional events, in general, including novel ones. Notably, the predictive relationship between amygdala activity and more frequent reappraisal choices was mediated by increased prefrontal activity, consistent with a model whereby greater amygdala reactivity prompts greater prefrontal activity, which in turn generates and/or reflects more frequent decisions to regulate emotion. Notably, we did not

see evidence that negative affect was down-regulated within the scanner task, in that higher amygdala responses and higher negative affect were predictive of regulating emotion more frequently. Second, at the level of the emotion-eliciting stimulus, we found that expression of a meta-analytically defined brain pattern associated with implementing reappraisal (Buhle et al., 2014) predicted choosing to regulate emotional responses for that stimulus. Overall, a predictive model that included amygdala activity, prefrontal cortex activity and expression of this distributed brain pattern showed substantially better performance than a model using only emotion self-reports and data on the normative affective potency of image stimuli, reaching 70% accuracy in predicting participant choice behaviors.

### **Implications for neural models of emotion regulation**

Neural models of emotion regulation have previously highlighted the importance of interacting brain systems for the top-down control and bottom-up generation of emotion (e.g., Ochsner et al., 2012; Kelley et al., 2015). The results of this study extend these models in several ways. First, these data indicate that people who show stronger bottom-up reactivity to affective stimuli (as measured by amygdala responses) are also more likely to engage prefrontally-mediated control processes, and, ultimately, to choose to regulate emotional responses to those stimuli. This suggests greater amygdala reactivity and prefrontal recruitment as neurobiological mechanisms underpinning the phenomenon of being motivated to regulate one's negative emotional responses. Overall, this pattern of data is consistent with a model whereby amygdala responses to a negative stimulus may reflect the need to regulate, whereas prefrontal responses

may reflect uninstructed recruitment of emotion regulation processes and/or controlled processing of stimulus meaning. For example, people who are unusually unreactive to aversive stimuli may show lower amygdala activity, lower prefrontal activity, and less frequent regulation choices.

Second, we found that expressing a distributed brain pattern associated with implementing reappraisal was independently predictive of choosing to regulate responses to particular stimuli, beyond activity estimates from the prefrontal regions alone. This indicates that greater activity in prefrontal regions per se – *and* expression of this distributed pattern – were independently predictive of choosing to regulate emotional responses to particular events. In other words, emotion regulation choices were most probable when participants showed high absolute levels of prefrontal cortex activity in addition to expressing this distributed pattern (characterized by relatively more activity in prefrontal regions compared to other parts of the brain).

Third, in a whole-brain analysis we also observed brain-behavior correlations with emotion regulation choice frequencies within two regions not of primary *a priori* interest – the precuneus and anterior temporal lobe. Both of these regions have been implicated in episodic memory retrieval and social cognition, among other functions (Bonner and Price, 2013; Wagner et al., 2005; Doré et al., 2014). Future studies could test the role of these regions in regulation contexts by asking whether precuneus and anterior temporal activity differ in predictive value for choosing to implement emotion regulation strategies that differ in demands on social cognition or long-term memory.

Finally, this study used a brain-as-predictor approach (see Berkman and Falk, 2013; Demos et al., 2012) to integrate neural models of emotion regulation with emerging theories of cognitive control that distinguish between signaling the need for (or expected value of) controlled processing (thought to be subserved by the dorsal anterior cingulate and adjacent regions of dmPFC) versus directly implementing this control (thought to be subserved by lateral PFC) (see Shenhav et al., 2013; Botvinick, 2007; Braver, 2012). Critically, we show that activity in the amygdala, lateral PFC, and dmPFC during an initial uninstructed encounter with an evocative stimulus can predict later choices to regulate, when the choice to do so was presented explicitly. This suggests that fMRI can detect variability in psychological and neural processes – like affective reactivity, signalling a need for top-down control, and engaging top-down control – that can be used to predict subsequent emotion regulation behaviors.

### **Implications for clinical disorders involving emotion dysregulation**

A common observation in the clinical literature is that patients with emotion-related clinical disorders do not show dramatic behavioral deficits on laboratory tests of emotion regulation capacity (see Joorman and Vanderlind, 2014). This suggests that core mechanisms of clinical dysfunction may not be well indexed by tasks that directly instruct participants how and when to regulate emotion (see Ochsner et al., 2012; Sheppes et al., 2015). Emotional dysfunction could be caused by abnormalities in 1) the bottom-up generation of emotion, 2) the ability to use top-down strategies for emotion regulation when instructed to, and/or 3) the tendency to self-identify emotion regulation opportunities and self-

initiate use of a regulation strategy (i.e., decisions to regulate or not, including selection of a context-appropriate strategy). Although behavioral and brain correlates of bottom-up generation and top-down control capacity (1 and 2) are typically assessed with existing lab tasks, tendencies to make regulatory choices of particular kinds (3) typically are not. Future work could ask whether particular clinical disorders are associated with disproportionate disruption in the capacity to deploy processes for emotion regulation when instructed versus the tendency to use them without being instructed to do so (see Doré et al., 2016). For example, patients with depression, bipolar disorder, or social anxiety may show atypical initial brain responses (i.e., in the image viewing task), atypical regulatory preferences (i.e., in the emotion regulation choice task), or atypical relationships between patterns of brain activity and subsequent regulation choices.

### **Limitations and future directions**

In this study we used fMRI measurements of brain responses at one time point to predict behaviors observed in a relatively controlled lab-based decision-making task at a later time point. Future studies could extend these findings by attempting to relate variability in brain responses to emotional behaviors in everyday contexts in which people are typically not prompted to enact regulation and are free to select any strategy they know (see Brans et al., 2013). In addition, we asked participants to view negative images in the scanner, when they were reacting to and/or regulating responses in an uninstructed manner, but did not ask them to indicate decisions to regulate until the subsequent choice

task. Future work could study all of these phases within the scanner – separating reactivity, decision, and implementation phases of choosing to regulate – in order to contrast the brain mechanisms of explicit choices with those reflecting uninstructed recruitment of brain processes predicting later choices identified here. Related, future work could ask whether situational variables (like how recently regulation strategies were acquired, trained, or practiced) can impact whether regulation decisions are predicted more by brain responses reflecting affective reactivity versus recruitment of top-down control. Moreover, studies with large sample sizes could ask whether patterns of brain activity can predict variability unique to novel images (i.e., above and beyond what can be predicted from behavioral responses to previously seen images).

Finally, it's possible that brain mechanisms of regulation decisions meaningfully differ across participant populations, like older adults (see Winecoff et al., 2011; Urry and Gross, 2010), or children and adolescents (Martin & Ochsner, 2016). Regulation decisions may also relate to subclinical variability in brain structure variables, like integrity of white matter tracts connecting brain systems associated with emotion and valuation (Chavez and Heatherton, 2014), or psychological variables like the motivation to experience particular emotional states (Tamir et al., 2015).

## **Conclusion**

When faced with emotional challenges, what determines whether we let our emotions unfold or attempt to rein them in? Here we suggest that, when confronted with distressing stimuli, 1) greater responses in brain regions

associated with emotional reactivity and cognitive control can be used to identify people who are more likely to regulate their emotional responses, and 2) expression of a brain pattern associated with cognitively regulating emotion can be used to predict whether regulation is chosen for a given stimulus. We hope that future work will build on the findings we describe here to work toward a mechanistic and prospectively predictive science of variable behavioral responses to distressing life circumstances.

## References

- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Berkman, E. T., and Falk, E. B. (2013). Beyond Brain Mapping Using Neural Measures to Predict Real-World Outcomes. *Current Directions in Psychological Science*, 22(1), 45–50.
- Bonner, M. F., and Price, A. R. (2013). Where Is the Anterior Temporal Lobe and What Does It Do? *The Journal of Neuroscience*, 33(10), 4213–4215.
- Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, 7(4), 356-366.
- Brans, K., Koval, P., Verduyn, P., Lim, Y. L., and Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion*, 13(5), 926.
- Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106-113.
- Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., ... Ochsner, K. N. (2014). Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cerebral Cortex*, 24(11), 2981–2990.
- Chang, L. J., Reddan, M., Ashar, Y. K., Eisenbarth, H., and Wager, T. D. (2015).

- The challenges of forecasting resilience. *Behavioral and Brain Sciences*, 38.
- Chavez, R. S., and Heatherton, T. F. (2014). Multimodal frontostriatal connectivity underlies individual differences in self-esteem. *Social Cognitive and Affective Neuroscience*, 10(3), 364-370.
- Demos, K.E., Heatherton, T.F., & Kelley, W.M. (2012). Individual differences in nucleus accumbens activity to food and sexual images predict weight gain and sexual behavior. *Journal of Neuroscience*, 32, 5549-5552.
- Denny, B., Silvers, J. and Ochsner, K. N. (2009). How we heal what we don't want to feel: The functional neural architecture of emotion regulation. In A. M. Kring and D. M. Sloan (Eds.), *Emotion regulation and psychopathology: A transdiagnostic approach to etiology and treatment* (pp. 59-87). New York: Guilford Press.
- Doré, B. P., Zerubavel, N., and Ochsner, K. N. (2014). Social cognitive neuroscience: A review of core systems. *APA Handbook of Personality and Social Psychology*, 693–720.
- Doré, B. P. & Silvers, J. A. & Ochsner, K. N. (2016). Toward a personalized science of emotion regulation. *Social and Personality Psychology Compass*, 10(4), 171-187.
- Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016.

- Joormann, J., and Vanderlind, W. M. (2014). Emotion Regulation in Depression  
The Role of Biased Cognition and Reduced Cognitive Control. *Clinical  
Psychological Science*, 2(4), 402-421.
- Kelley, W. M., Wagner, D. D., and Heatherton, T. F. (2015). In search of a human  
self-regulation system. *Annual Review of Neuroscience*, 38, 389-411.
- Koyama, T., McHaffie, J. G., Laurienti, P. J., and Coghill, R. C. (2003). The  
single-epoch fMRI design: validation of a simplified paradigm for the  
collection of subjective ratings. *Neuroimage*, 19(3), 976–987.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). International affective  
picture system (IAPS): Affective ratings of pictures and instruction manual.  
Technical Report A-8.
- Martin, R. E., & Ochsner, K. N. (2016). The neuroscience of emotion regulation  
development: implications for education. *Current Opinion in Behavioral  
Sciences*, 10, 142-148.
- Major, B., Richards, C., Cooper, M. L., Cozzarelli, C., and Zubek, J. (1998).  
Personal resilience, cognitive appraisals, and coping: an integrative model  
of adjustment to abortion. *Journal of Personality and Social Psychology*,  
74(3), 735.
- McRae, K., Ciesielski, B., and Gross, J. J. (2012). Unpacking cognitive  
reappraisal: goals, tactics, and outcomes. *Emotion*, 12(2), 250.
- Ochsner, K. N., Silvers, J. A., and Buhle, J. T. (2012). Functional imaging studies  
of emotion regulation: a synthetic review and evolving model of the

- cognitive control of emotion. *Annals of the New York Academy of Sciences*, 1251(1), E1–E24.
- Phelps, E. A., and LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, 48(2), 175–187.
- Russo, S. J., Murrough, J. W., Han, M.-H., Charney, D. S., and Nestler, E. J. (2012). Neurobiology of resilience. *Nature Neuroscience*, 15(11), 1475–1484.
- Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, 79(2), 217-240.
- Sheppes, G., Scheibe, S., Suri, G., Radu, P., Blechert, J., and Gross, J. J. (2014). Emotion regulation choice: a conceptual framework and supporting evidence. *Journal of Experimental Psychology: General*, 143(1), 163-181.
- Sheppes, G., Suri, G., and Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual Review of Clinical Psychology*, 11, 379–405.
- Tamir, M., Bigman, Y. E., Rhodes, E., Salerno, J., and Schreier, J. (2015). An expectancy-value model of emotion regulation: Implications for motivation, emotional experience, and decision making. *Emotion*, 15(1), 90-103.
- Urry, H. L., and Gross, J. J. (2010). Emotion regulation in older age. *Current Directions in Psychological Science*, 19(6), 352–357.

- van Ast, V. A., Spicer, J., Smith, E. E., Schmer-Galunder, S., Liberzon, I., Abelson, J. L., and Wager, T. D. (2016). Brain mechanisms of social threat effects on working memory. *Cerebral Cortex*, 26(2), 544-56.
- Vehtari, A., Gelman, A., and Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *arXiv Preprint arXiv:1507.04544*.
- Wagner, A. D., Shannon, B. J., Kahn, I., and Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, 9(9), 445–453.
- Winecoff, A., LaBar, K. S., Madden, D. J., Cabeza, R., and Huettel, S. A. (2011). Cognitive and neural contributors to emotion regulation in aging. *Social Cognitive and Affective Neuroscience*, 6(2), 165–176.
- Woo, C.-W., Roy, M., Buhle, J. T., and Wager, T. D. (2015). Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biology*, 13(1), e1002036.

### Figure Legends

**Figure 1.** **A)** Negative image viewing task, and **B)** emotion regulation choice task. Predictive models use brain responses in the negative image viewing task to predict participant choice behaviors in the emotion regulation choice task.

**Figure 2.** Scatterplots and generalized linear model fits reflecting predictive relationships between person-to-person variability in brain activity in the negative image viewing task and subsequent reappraisal choice frequencies, for *a priori* **A)** anatomically-defined amygdala, and **B)** meta-analytically defined dmPFC, dIPFC, and vIPFC bilateral regions of interest.

**Figure 3.** **A)** Plot of regression coefficients (with 95%CI) for the full model (model 2) predicting reappraisal choice frequencies from brain and self-report variables (entered as simultaneous predictors). **B)** Model comparison metrics for full model (model 2), and reduced model including only negative affect ratings and positive affect ratings (model 1). Model fit is summarized by leave-one-out cross-validated error (LOOIC), AIC, and BIC information criterion scores (i.e., adjusted model deviance, a measure of the relative quality of the models for these data). A lower number indicates better fit.

**Figure 4.** Greater average amygdala activity in the negative image viewing task predicts subsequently choosing to regulate emotion more frequently, and this relationship is mediated by greater prefrontal cortex activity during the image viewing task.

**Figure 5.** Regions of the brain for which person-to-person variability in activity during the negative image viewing task is most correlated with subsequent reappraisal choices (whole-brain FWE  $p < .05$ ).

**Figure 6.** **A)** Meta-analytically derived whole-brain pattern associated with regulating emotion via reappraisal (display is thresholded at  $z > 2.6$ ,  $k = 20$ , but all voxels were used in analyses). **B)** Trial-to-trial variability in expression of this pattern ( $z$ -transformed) is predictive of subsequently choosing to use reappraisal for particular images.

**Figure 7.** **A)** Plot of regression coefficients (with standard error and 95%CI) for the full model (model 4) predicting reappraisal choices from brain, self-reports of affect, and normative ratings of the images (entered as simultaneous predictors) **B)** Model comparison metrics for full model (model 4), and reduced models with only ROIs, affect, and image norms (model 3), affect and image norms (model 2), and image norms only (model 1) **C)** ROC curves depicting prediction accuracy for models 1,2,3, and 4 in terms of sensitivity (correct prediction of reappraisal choices) and specificity (correct prediction of natural response choices), across a range of possible prediction thresholds.