An action-value model explains the role of the dorsal anterior cingulate cortex in performance monitoring during affect regulation

Authors
Keith A. Bush*, G. Andrew James, Anthony A. Privratsky, Kevin P. Fialkowski, Clinton D. Kilts

Affiliations
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 engineering control theory, machine learning, and human neuroimaging to critically test the putative role of the dorsal anterior cingulate cortex (dACC) in performance monitoring during an emotion regulation task. Healthy adult participants (n=75) underwent cued-recall of affective image stimuli with concurrent functional magnetic resonance imaging and psychophysiological response recording. During cued-recall, participants engaged in explicit self-regulation of their affective state toward defined affective goals. Established decoding methods measured affect processing from fMRI BOLD signals across the orthogonal affective dimensions of valence and arousal. We independently validated participants' affective state representations via stimulus-dependent facial electromyography (valence) and electrodermal activity (arousal) responses. We then used the decoded affective signatures to test and compare four computational models of performance monitoring (i.e., error, predicted response outcome, action-value, and conflict) by their relative abilities to explain task-related dACC activation. We found that the dACC most plausibly encodes action-value for both valence and arousal processing. We confirmed that the dACC directly encodes affective arousal and also likely encodes recruitment of attention and regulation resources. Beyond its contribution to parsing the roles of the dACC in emotion regulation, this study introduced a novel analytical framework through which affect processing and regulation may be functionally dissociated, thereby permitting mechanistic analysis of real-world emotion regulation strategies, e.g., distraction and reappraisal, which are widely employed in cognitive behavioral therapy to address clinical deficits in emotion regulation.

Keywords
Affect, regulation, performance monitoring, control theory, dACC, MVPA, psychophysiology, emotion

Introduction

The expression and perception of emotions are valuable social cognitive resources that allow us to focus our attention to salient environmental features[1], orchestrate social exchanges[2], prioritize our decisions[3], and engage in appetitive and aversive behaviors[4]. Conversely, disruption of the multiple roles that emotions perform in directing our cognitions and behaviors is associated with mental health problems. Emotion misprocessing (e.g., emotional hypo/hyperreactivity and lability) engenders prolonged maladaptive emotional states implicated in psychiatric illnesses[5]. Thus, we routinely deploy emotion regulation strategies to mitigate such processing errors and to mold our emotional experiences in ways that benefit our well-being[6].

The process model of emotion regulation[7,8] furnishes us with a conceptual bridge between the temporal steps of characteristic emotion formation (i.e., the modal model[9]) and the cognitive processes serving to extinguish, alter, or promote such formation. For example, two widely deployed emotion regulation strategies, distraction and reappraisal, respectively target attentional deployment and cognitive appraisal in order to disrupt the characteristic (and potentially dysfunctional) trajectory of emotional response. Extant empirical studies of emotion regulation in healthy subjects describe the comparative efficacy of these strategies[10], individual differences in strategy efficacy[11], as well as their attendant functional neurocircuitry[12–14]. The process model (and its extension[15]) provides a framework by which to organize and map these findings onto reported patterns of emotion dysregulation pervading psychopathology[16–20].

New models of emotion regulation must build upon this framework and establish mechanistic descriptions that merge empirical observations with formal analytical rigor to describe the underlying regulation processes[21]. A heuristic framework for such model development is provided by the superordinate process of cognitive control, which has long drawn mechanistic inference from engineering control theory[22–24] to empirically test computational models[25,26]. As a step in this direction, leading scholars recently proposed a neuroanatomically-constrained unified emotion regulation framework[27] rooted in reinforcement learning[28], a value-based multi-step decision strategy having important mathematical and conceptual connections to dynamic programming[29], optimal control theory[30], and machine learning[31].

This study sought to lay the groundwork for applying control theoretic analysis to the empirical study of emotion regulation mechanisms as a means of directly comparing the major extant models of performance monitoring posited by both the cognitive control and emotion regulation research communities. For this comparison we proposed a novel experimental task performed concurrently with functional magnetic resonance imaging (fMRI) and multimodal psychophysiological measurements that tested and validated a restricted variant of emotion regulation. The salient finding of the study is that it identified, from multiple major computational models, the most likely mathematical form of performance monitoring engaged during effortful emotion regulation.

Unfortunately, the emotion and emotion regulation literatures do not operate with a consistent, defined terminology[32]. For this work, we framed emotion regulation within a narrow, but important, subset of the field – explicit (effortful) regulation of the orthogonal dimensions of affect processing (i.e., valence and arousal)[33] reflected by moment-to-moment processes during the cued-recall of complex affective image stimuli (see Figure 1 panel A). We restricted our hypotheses and analyses to these core affective cognitions which encompass aspects of the broader fields of emotion processing (e.g., the discrete emotion model) and emotion regulation strategies (e.g., implicit and extrinsic regulation)[34].

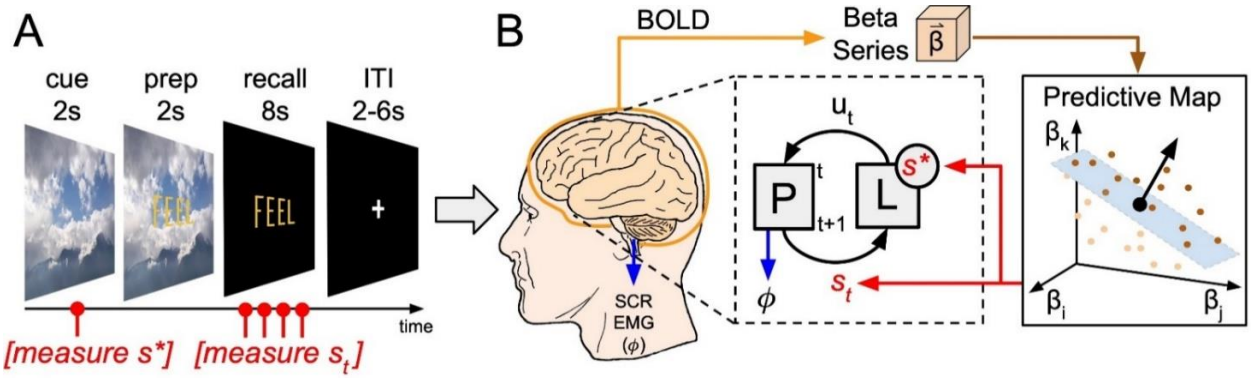

Figure 1: *Experimental task design and conceptual model.* ***(A)*** *Schematic of a single trial of the cued-recall task. The trial presents a cue image stimulus for 2 s. The trial then overlays the word "FEEL" for 2 s in yellow font. The image disappears, replaced by a black screen, leaving only the word "FEEL" for 8 s. The trial ends with a fixation cross for an inter-trial interval sampled uniformly randomly on the range of 2-6 s. Prior to performing the task 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."* ***(B)*** *A closed-loop conceptual model of the brain processing related to explicit affect regulation induced by the cued-recall task. The proposed closed-loop control system's box diagram falls within the dashed area of the figure. The system of interest, aka the plant, P, is assumed to be observed according to measurements of its state, s. In this study, the plant is composed of the independent dimensions of affect processing, valence and arousal. The control law, L, evaluates the plant's state and emits a control action, u, intended to perturb P such that the system state moves closer to the desired state, i.e., the goal, s*. Each cued-recall task trial first captures the goal state and then captures four separate measurements of the system state as the subject regulates their affective state to match the goal state. Goal and system states are measured by passing beta maps, i.e., whole-brain gray matter patterns of neural activation extracted from the subject's fMRI BOLD signal, through machine learning models to predict valence and arousal. For validation purposes, psychophysiological measures of autonomic arousal (skin conductance response, SCR) and hedonic valence (facial electromyography, EMG) are captured concurrently with fMRI BOLD signal.*

Explicit affect regulation toward a known affective goal (cued-recall) may be recast as top-down cognitive control[21,35] of affect processing. Moreover, the shared conceptual relationship between explicit affect regulation and cognitive control can be formalized through the lens of engineering control theory. In doing so, our conceptual model of affect regulation conforms to a closed-loop control system[36] (see Figure 1 panel B) employing a feedback loop composed of three sequential steps: 1) measurement of the system of interest, P (i.e. the plant), as the momentary system state, s_t; 2) evaluation of this system state with respect to a goal state, s* (i.e. goal-directed performance monitoring), and 3) action (i.e., system perturbation), u_t, to align a future measurement of the system's state more closely with the goal state. Within the emotion regulation literature, the perception-valuation-action (PVA) model[37] describes operational elements of this system. Within the cognitive control literature, this system is labeled the guided activation framework (GAF)[38], a generalization of the biased competition model[39,40].

Both explicit affect regulation and non-affective cognitive control exhibit robust similarities in their corollary neuroanatomical activations (localized to the lateral and medial prefrontal cortex) that support the proposed control theoretic connections between these cognitive processes. The critical gap in our understanding of these two distinct but related cognitive regulatory processes stems from mathematical ambiguity concerning how the three closed-loop subsystems (i.e., measurement, monitoring, and action, Figure 1 panel B) interact to constitute a neural processing network capable of accomplishing the regulatory outcome. The first of these subsystems (measurement of the system state) has been well-characterized by machine learning approaches that predict affective dimensions[41–45] and discrete emotions[46,47] from neural features (i.e., affect

decodings). Consistent activation localizations related to affective state measurements include medial prefrontal cortex (mPFC), bilateral amygdala, precuneus, and bilateral ventral temporal lobe [42,43,46]. Prior cognitive control studies have associated the third subsystem (action) with the lateral prefrontal cortex (lPFC) via its mechanistic roles in executive control[48], goal selection[49], adaptation to novelty[50], and task-general control[23]. Moreover, meta-studies have linked the lPFC (both ventral and dorsal) with explicit emotion regulation through a network that includes both parietal cortex and pre-supplementary motor area (pre-SMA)[13,14]. A similar network is also implicated in cognitive control[51,52].

In this study, we used engineering control theory to mathematically model the second subsystem, performance monitoring, and to identify its underlying neural processing mechanisms. Given the available functional neuroanatomical evidence, we proposed that the dorsal anterior cingulate cortex (dACC) represents the most plausible candidate involved in the neural information processing that encodes the performance monitoring subsystem of the closed-loop model. Prior cognitive control studies have linked the dACC to performance monitoring[53] via numerous process measures and mechanisms: error detection[54–56], confict[57–63], and value prediction[51,53,64,65]. Similarly, meta-analyses and multivariate prediction studies consistently implicate the dACC with emotional reactivity[42,66–69], and Buhle et al. (2012)[13] suggested the dACC may be linked to emotion regulation through appraisal.

To test the closed-loop conceptual model of affect regulation, we devised a novel experiment to identify the mathematical model of performance monitoring most likely employed by the dACC. First, we constructed predictive models of affective state processing (both valence and arousal) from whole-brain gray matter patterns of neural activation observed via fMRI in response to an information-optimized[70] set of image stimuli drawn from the International Affective Picture Set (IAPS)[71]. Concurrently, we recorded multimodal psychophysiological stimulus responses to verify the underlying induction of affect processing. These predictive models served as instruments by which to precisely measure the system under control conditions prompted by a cued-recall task (see Figure 1, panel A). Our models measured both the affect regulation goal (affect prediction coinciding with the cue image) as well as the control trajectory (the sequence of affect predictions coinciding with the subject's recall and re-experiencing of the cue's signaled affective properties). From these predictions we calculated three competing computational models of goal-directed performance monitoring (perceived error[72], predicted response outcome[26], and action-value[22]) and critically compared each model's ability to explain task-related dACC activation, and thus its role in affect regulation.

In our study of (n=75) healthy adult subjects, each of whom performed 30 cued-recall task trials, we found that the dACC monitors control performance according to an action-value model in which both recall error (with respect to the cue's perceived affect) and the cost of action (e.g., cognitive effort) contribute. However, we also identified dACC activation – independent of the performance measurement – that is likely related to task performance in the form of a conflict signal reflecting the task demand for control against the prepotent response of inaction. Our findings also confirmed previously observed activations of bilateral ventral lPFC (vlPFC) and bilateral parietal cortex that are broadly associated with emotion regulation. However, in contrast to the extant literature, we observed bilateral deactivation of the dorsal lPFC (dlPFC) and bilateral activation of the dorsal medial PFC (dmPFC) as well as bilateral deactivation of the ventral visual stream, amygdala, and ventral striatum. These findings suggest the existence of task-specific variation in the neural mechanism underlying emotion regulation, particularly dlPFC activation during cognitive control of affect processing, which may reflect the task's goal structure – the element that is unique to our study compared to explicit emotion regulation writ large.

In sum, using affect processing decodings of fMRI signals, we observed the moment-to-moment affect processing of subjects during cued-recall of affective image stimuli and calculated measures of performance monitoring that we then functionally related to neural activations in the dACC. In this manner, we identified the likely processing model of performance monitoring critical

to explicit affect regulation. These findings establish a framework by which the complete closed-loop model of explicit affect regulation, including action selection, may be both mathematically described and identified by its functional neurocircuitry.

Methods

Study Overview

This study explored data acquired from two separate experiments, the Intrinsic Neuromodulation of Core Affect (INCA) experiment and the Cognitive Control Theoretic Mechanisms of Real-time fMRI-Guided Neuromodulation (CTM) experiment (National Science Foundation, BCS-1735820). Both experiments were control theoretic functional neuroimaging explorations of affect processing and regulation, which incorporated both unguided and real-time fMRI guided affect regulation tasks. Importantly, both experiments shared identical affect processing, affect regulation, and resting state task designs as well as identical task ordering; and, we have previously reported the relevant details of these studies' shared affect processing task demands[70]. Here, we elaborate the experiment design details that informed the present affect regulation analysis.

We conducted both the INCA and CTM studies over two separate sessions, each occurring on separate days. During Session 1, participants provided written informed consent, received screening for clinically relevant exclusionary criteria via structured clinical interview, and completed behavioral surveys and questionnaires. We acquired neuroimaging and concurrent psychophysiological measurements during Session 2. The analyses reported here include only data acquired during the first three functional image acquisitions (i.e., scans) of Session 2, which are composed of the two System Identification Tasks and one Resting State Task. The relevant task descriptions are elaborated below.

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 data 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 sought study participants who lived within a one hour drive of the UAMS campus in Little Rock, Arkansas for recruitment into this study via publicly posted flyers, direct emails drawn from the ARresearch registry (arresearch.org), and social media advertisements. From our total participant sample (n=78), we excluded from analysis two subjects enrolled in the CTM study who did not complete the resting state portion of the study (due to early exit from the scanner) as well as one subject enrolled in the INCA study who was inadvertently included in the study despite meeting exclusionary criteria. The final participant sample (n=75; n_{CTM}=56 and n_{INCA}=19) possessed the following demographic characteristics: age [mean(s.d.)]: 35.5(13.1), range 18–63; sex: 46(61%) female, race/ethnicity: 65(86.7%) self-reporting as White or Caucasian, 9(12%) as Black or African-American, 1(1.3%) as Asian; education [mean(s.d.)]: 16.7(2.7) years, range 12–23; WAIS-IV IQ [mean(s.d.)]: 106.6(13.7), range 74–145. All of the study's participants were right-handed 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[73]. All participants reported no current use of psychotropic medication and produced a negative urine screen for drugs of abuse (cocaine, amphetamines, methamphetamines, marijuana, opiates, and benzodiazepines) immediately prior to the MRI scan. CTM participants also produced a negative urine screen prior to SCID-IV clinical interview. When necessary, we corrected participants' vision to 20/20 using an MRI compatible lens system (MediGoggles™, Oxfordshire, United Kingdom), and we excluded all participants endorsing color blindness.

System Identification Task Design

The System Identification task had two independent roles in our study. First, this task induced, via image stimuli, affect processing in our participants that was recorded via concurrent fMRI and psychophysiology in order to construct and validate model-based predictions of affect processing using features derived from neural activity. Second, this task employed cued-recall trials, again based upon image stimuli, to predict (post-hoc via our constructed models) both the affect regulation goal (induced by the cue) and the moment-to-moment affect processing of explicit affect regulation (volitionally induced during recall).

We achieved the first of these roles using 90 image stimuli that were computationally sampled from the International Affective Picture Set (IAPS) in order to maximize the subspace span of valence-arousal combinations in the resultant set of image stimuli. We have extensively reported our methodology and motivation for this sampling process[42,70]. We presented each of these image stimuli for 2 s followed by an inter-trial interval (ITI) uniformly randomly sampled from the range 2–6 s during which we presented a fixation cross. We labeled these presentations as extrinsic trials. We achieved the second of these roles using 30 image stimuli (independently but identically sampled as described above from the remaining IAPS images, see Supplemental Figure S1) as part of cued-recall trials (see Figure 1, panel A) in which we presented a cue image stimulus for 2 s followed by a visual preparation instruction (the word “FEEL” superimposed over the image) for 2 s followed by a recall instruction in which the image disappears (leaving only the word “FEEL”) for 8 s. During recall the participant explicitly regulated their affect processing to match the affect processing induced by the cue image stimulus. Finally, a fixation cross replaced the word “FEEL” for an ITI sampled uniformly randomly from the range 2–6 s. During training for the cued-recall trials, we instructed participants 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.” To promote replication, we provide the complete list of IAPS images used as cue image stimuli in Supplemental Table S1.

We temporally arranged both the extrinsic trials and cued-recall trials across two 9.25 min System Identification Task scans by iterative search of uniformly randomly sampled trial onset times such that: 1) the maximum of the bivariate correlations (R) between hemodynamic response function derived regressors (two for valence and two for arousal dichotomized according to the image stimuli’s normative scores in relation to Likert score of 5) was less than 0.25; 2) the scan began with an extrinsic format; and, 3) all scans started and ended with positively valent image stimuli. We fixed the order of the trials for all subjects.

MR Image Acquisition and Preprocessing

We acquired all imaging data for both the INCA and CTM studies using the same Philips 3T Achieva X-series MRI scanner (Philips Healthcare, Eindhoven, The Netherlands) with a 32-channel head coil. We previously reported the MR acquisition parameters for both anatomic and functional images that we used[70], which are summarized as follows. 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. We performed all MRI preprocessing using AFNI[74] (Version AFNI_19.1.04) unless otherwise noted. We performed the following sequence of steps to process our anatomical data: skull stripping, spatial normalization to the icbm452 brain atlas, and segmentation (via FSL[75]) into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). From the individual participant GM segmentations we also constructed a group-level GM mask composed of voxels in which ≥ 50% of the individual participant GM masks were present. To preprocess functional images, we performed the following sequence of steps:

despiking, slice-time correction, deobliquing, motion correction, transformation to the spatially normalized anatomic image, regression of the mean timecourses and temporal derivatives of the WM and CSF masks as well as a 24-parameter motion model[76,77], spatial smoothing (8 mm FWHM Gaussian kernel), and scaling to percent signal change. For resting state functional images we took the additional step of global mean signal subtraction prior to smoothing and scaling.

Psychophysiology Data Acquisition and Preprocessing

We acquired all psychophysiological measures using the BIOPAC MP150 Data Acquisition System (BIOPAC Systems, Inc., Goleta, CA) in conjunction with AcqKnowledge software. We augmented the MP150 with acquisition modules in order capture multiple physiological modalities including galvanic skin response (EDA 100C-MRI module), pulse plethysmography (TSD200-MRI module), respiration transduction (TSD221-MRI module), and facial electromyography (EMG100C-MRI module). We acquired all physiological signals at 2000 Hz. Specifically related to this study, we acquired galvanic skin response from the left hand; see Bush et al. (2018)[70] for electrode placement details. We captured two separate measurements of facial electromyography (EMG): zygomaticus major (zEMG) and corrugator supercilii (cEMG). Placement of electrodes conformed to guidelines reported in Fridlund and Cacciopo (1986)[78].

We have previously reported in detail our methods for preprocessing and modeling galvanic skin response measures in past work[70]. In sum, we used the canonical skin conductance response function in conjunction with the beta-series method[79] to capture the temporally succinct physiological corollary of autonomic arousal associated with stimulus-induced affect processing[80]. We repurposed that methodology here for application to the cued-recall task trials in which we extracted arousal responses for the cue condition as well as each volume of the recall condition.

In this work, we used facial electromyography to provide an independent physiological measure of hedonic valence. We first bandpass filtered the raw EMG signal on the range 10–500 Hz to remove artifacts[81] and then rectified the filtered signal. We then extracted the sum of the rectified signal for each relevant task period as our feature set, e.g., over the 2 s cue stimulus as well as for each of the fMRI acquisition volumes associated with the recall task (4 volumes, each 2 s in duration).

Multivariate Pattern Classification and Platt-Scaling

We have previously reported in detail our methods for extracting whole-brain patterns of neural activation as input features to predictive models of affect processing[70]. To summarize, we combined the canonical hemodynamic response function and the beta-series method[79] to construct whole-brain patterns of neural activation in response to individual extrinsic stimuli presented within the System Identification Task. For each subject, we then fit a linear support vector machine model to classify the features of the affective stimuli using labels that were dichotomized according to the stimuli's normative affect scores relative to the middle Likert score, 5, respectively, for the independent affective dimensions of valence and arousal. We also re-validated (first reported in Bush et al., (2018)[70]) that the SVM model architecture was robust to sparse, high-dimensional feature-spaces (as is typical of whole-brain analysis of 3 Tesla fMRI data) by comparing SVM predictions based upon whole-brain GM features against SVM predictions based upon Gram-Schmidt dimensionally reduced variants of these features. As in prior work, we characterized all SVM models based upon intra-subject leave-one-out-cross-validation (LOOCV) predictions in which each cross-validation prediction was the mean of 30 randomly balanced training datasets that excluded a single feature-label test pair. Finally, we reported prediction performance for both individual participants (using the binomial distribution as a null hypothesis, assuming probability .5 for the class label) or by a group-level ensemble of intra-subject predictions (using the Wilcoxon signed-rank test assuming probability .5 for the null classification accuracy).

Here we extended this prediction methodology by applying these SVM models to cued-recall task trials. Recent work supported novel, out-of-sample application of predictive models of affect processing[82–84]. To this end, we extracted two separate beta-series for each cued-recall trial: 1) the cue condition (see Figure 1, panel A), indexed by i, which yielded series $\beta(i)$; and 2) the four individually acquired volumes related to the recall condition following the cue condition of each trial (see also Figure 1, panel A), each volume indexed by t given i, which yielded series $\beta(t|i)$, where t|i referred to the tth recall volume succeeding the ith cue. We then input the resultant beta-series into the SVM classifiers (trained on separate stimuli) to predict hyperplane classification distances, $h^{P}_{\beta(i)}(i)$ and $h^{P}_{\beta(t)}(t)$, for all i and t where: 1) P denoted the plant of interest, $P \in \{valence, arousal\}$, 2) subscripted $\beta(\cdot)$ denoted the index of the beta-series feature input to the model, and 3) $(\cdot)$ denoted the index of the classification. In the base case, described here, these indices matched, but they differed, e.g., when we applied the model to forecast future affect. As a final step, we applied Platt scaling[85] to convert our models' predictions of distance from a classification decision boundary (i.e., hyperplane) to continuous predictions of the affective state's probability of conforming to the positive class (i.e., the probability of positive valence or the probability of high arousal) such that

$$p^{P(+)}_{\beta(\cdot)}(\cdot) = 1/(1+\exp(-h^{P}_{\beta(\cdot)}(\cdot))).$$

We validated this transformation of our model predictions according to two tests. First, we applied Platt scaling to all of the cross-validated predictions (and normatively scored labels) of our extrinsic stimulus trials and tested, via a linear mixed-effects model, the group-level significance of our models' predictive fidelity in the transformed probability space. Second, we repeated this validation test for the cue condition stimuli of the cued-recall task trials using their IAPS normative scores as labels.

Control Condition Cued-Recall Task Modeling via Resting State

In previous work we demonstrated the application of our predictive models to characterize the moment-to-moment affect processing dynamics entrained in resting-state fMRI BOLD signal[82]. In brief, for each subject this approach averaged the predictions of the subject's fit models to numerous beta-series constructed from sets of independently, uniformly randomly sampled onset times. We validated this approach out-of-sample and also showed temporal agreement of the model predictions with concurrent measures of psychophysiology. Here, we leveraged this methodology to construct cued-recall tasks within resting state task data which provided controls against which to measure the significance of affect processing dynamics entrained during explicit affect regulation. From each subject's resting state task fMRI data we constructed affect processing estimates, respectively, for valence and arousal. We then uniformly randomly sampled (n=30) onset times of cue stimuli and extracted the attendant cue and recall affect processing predictions according to the trial's structure (see Figure 1, panel A).

Computational Models of Performance Monitoring

We designed the cued-recall task as a means of exploring the mathematical structure of performance monitoring to best explain the functional attributes of dACC activation during explicit affect regulation. From the cognitive control research literature, we initially identified five influential explanatory models of dACC activation (see Table 1); however, upon inspection, we restricted our exploration to just three models which most closely align with performance monitoring in the engineering control literature and would be best evaluated by our experiment design. We included error processing[72], which directly represents the plant's current state with respect to a goal (closed-loop control), predicted response outcome[26], which estimates the plant's likely future state with respect to a goal (model-based control), and expected value of control[22], which

estimates the expected value (with respect to a goal, operationalized as a reward function) of making a particular control decision in the current state (reinforcement learning).

The remaining two performance monitoring-related models, cognitive conflict[59] and prediction error likelihood (PEL)[25], evaluate important theoretical aspects of cognitive control that are related to our included models and may be subsumed by, or derived from, them. However, PEL research has shown that the dACC activates in response to contexts in which there is a learned probability of failure, even when acting rationally. It is unclear if the neural mechanisms underlying explicit affect regulation conform to such a highly stochastic transition function. Moreover, our analysis relied on the subject's awareness of the affective goal throughout the control trial, thereby conceptually excluding this model from our analysis. Similarly, our task design did not accommodate a canonical test of cognitive conflict. We had two reasons for concluding this. First, our previous work showed that affect processing was encoded orthogonally across the independent affective dimensions of valence and arousal[42]; therefore, affect processing should not constitute a conflict signal in the dACC. Second, our most plausible assumption of prepotent response to explicit affect regulation was that of "no control." Baseline dACC activation during the recall step of the cued-recall task would suggest cognitive control recruitment to overcome this prepotent response and, therefore, conflict did not need to be modeled explicitly.

Error formed the foundation of our modeling approach. As indicated in Table 1 (and conceptualized by Figure 1, panel B), we computed error, δ, as the difference in predicted affect processing between each recall volume and the target cue (i.e., the affective goal) separately for both valence and arousal. From these basic calculations, we constructed the remaining models.

Model Class	**dACC Activation Measure**	**Computational Model**
Performance	Error	$\delta(t) = \lVert p^{P(+)}_{\beta(t)}(t) - p^{P(+)}_{\beta(i)}(i) \rVert$
	Predicted Response Outcome (PRO)	$\delta_{PRO}(t) = \delta_{\psi(\beta(t-1))}(t)$
	Expected Value of Control (EVC)	$Q(s(t),a(t))$ $s(t) \approx \psi(\beta(t))$ $a(t) \approx p^{P(+)}_{\beta(t+1)}(t+1) - p^{P(+)}_{\beta(t)}(t)$ $r(t) = - f_a \cdot \lvert a(t) \rvert - (1-f_a) \cdot \delta(t+1)$
Cognitive Demand	Conflict	1*
Control Outcome Uncertainty	Prediction Error-likelihood (PEL)	n/a

Table 1: *Summary of cognitive control models of dACC function. *Theorized model of conflict assuming that "no control" is a valid prepotent response to the cued-recall task.*

Both the PRO and EVC models relied on predictions of derived quantities based upon a state space of neural activations. This posed a significant challenge in that activations of the dACC are likely to be correlated with other discrete neural activations, potentially resulting in false positive model outcomes. To control for this possibility, we constructed an elaborate pipeline to isolate our modeling process in both time and neuroanatomical space from the dACC activations that we were attempting to functionally characterize.

First, we generated a mask of the entire medial frontal cortex (mFC) by inflating an existing mask[86] by two voxels. We then used this mask to exclude mFC voxels from an established 20-component (18 usable components) partition of the brain derived from independent components analysis of the BrainMap database[87]. We then mapped all beta-series into this mFC-excluded 18-dimensional space. We denoted this transformation of the beta-series as ψ(β) and refer to it as the restricted beta-series, which was used (as described below) to model task-related neural activity for neural processing networks that excluded the mFC.

We next fit linear support vector machine regression models of the computed error trajectories, using the restricted beta-series as input features, where the input feature is drawn from the previous time index as the error to be predicted (i.e., error forecasting, see Table 1, PRO). To predict each subject's future error, we fit intra-subject models for all other subjects, using the external subjects' restricted beta-series and error trajectories. We then estimated the target subject's PROs by ensemble averaging of the out-of-sample subjects' model predictions, using the target subject's restricted beta-series as input features to the models. To validate this modeling approach we measured the effect size of the out-of-sample ensemble average in approximating the target subject's intra-subject predictions. We measured this effect size using a linear mixed-effects model where random slope and intercept effects were modeled subject-wise. We estimated effect sizes separately for valence and arousal.

We showed by direct proof (see Supplemental Figure S2) that EVC is mathematically equivalent to Q-learning[88] with a composite reward function that incorporates the cost of acting in addition to the transition reward. We computed an approximate batch solution to the action-value function, Q, via fitted Q-iteration[89] using the following constituent components: states, s, actions, a, and rewards, r (see Table 1, EVC). We modeled actions as the differences between successive affect processing predictions (i.e., forward Euler approximations of the first temporal derivative), respectively, for valence and arousal. We modeled states identically to the methods used for the PRO model (i.e., restricted beta-series). Finally, we modeled the reward function as the weighted combination of the action and the error (note, these equations assume the machine learning convention in which rewards can have either positive or negative sign).

We conducted all EVC experiments over a parameter space composed of the cross product of the discount factor, γ, sampled from the range [0,1] at intervals of 0.1 and fraction of action cost, f_a, sampled from the range [0,1] at intervals of 0.2. We also discretized the action space to five actions, $a_d \in [-2,-1,0,1,2]$ according to the following heuristic. We computed actions for each subject and standardized these values subject-wise. We then pooled all standardized actions and computed the group standard deviation, σ. We then assigned discrete actions according to a standard deviation driven partition of the standardized action space of each subject such that

$$a_d = \begin{cases} 2 & : a \geq 2\sigma \\ 1 & : a \geq 1\sigma \text{ and } a < 2\sigma \\ 0 & : a < 1\sigma \\ -1 & : a \leq -1\sigma \text{ and } a > -2\sigma \\ -2 & : a < -2\sigma. \end{cases}$$

For each subject, we standardized, separately, the actions and errors for all cued-recall task trials and then calculated the reward according to the appropriate meta-parameters for the given experiment. We then fit the action-value function according to a random forest based implementation of fitted Q-iteration[89,90]. However, similar to our scheme for estimating PROs, we employed an out-of-sample ensemble averaging for our EVC estimate for each action at each state. We stored for additional analysis: the ensemble average prediction of Q-value for the on-policy actions, the expected Q-values of random action at each state (where probabilities of each discrete action were estimated from the subject's distribution of actions), and the errors between the on-policy action and the optimal action for each state.

For each set of parameters, we tested for group-level significant differences between the Q-values of on-policy actions and the expected Q-values of random actions using the Wilcoxon rank-sum test. Within significant parameter sets, we calculated the group-level mean error between on-policy actions and optimal actions. We then selected the meta-parameter set such that group-level mean error was minimized. We broke ties as follows: maximum γ (to decorrelate

EVC from PRO) and minimum f_a (we assumed cognitive effort was subordinate to control performance).

Comparing Performance Monitoring Models of Medial Prefrontal Cortical Activation

We critically compared our constructed models of performance monitoring according to whole-brain gray matter linear mixed-effects models, implemented via AFNI's 3dlme function. To control for our use of neural activations outside the dACC to predict model values within the dACC, we constructed two separate mixed-effects models. The first model (see Model 1) incorporated the performance monitoring fixed effects. We used this model to characterize only neural activations within the dACC. The notation ($Fixed_i$ x Sex x Age) represents an expansion of fixed effects representing all possible interactions of age and sex with each of the explicitly listed primary fixed effects, $Fixed_i$, which by default also included pair-wise interactions, e.g. $Fixed_i$ x Sex and $Fixed_i$ x Age as well as Sex and Age alone.

Model 1:

β = 1 + Duration + Affect + Error + PRO + nEVC + (1+Affect|Subj) + ($Fixed_i$ x Sex x Age)

We used the second model (see Model 2), which omitted the measurements of performance monitoring as fixed effects, to characterize neural encodings falling outside of the dACC.

Model 2:

β = 1 + Duration + Affect + (1+Affect|Subj) + ($Fixed_i$ x Sex x Age)

The resulting solutions provided F-statistics for each fixed-effect (and interaction) for each voxel included in the model. We also constructed general linear tests for all primary fixed effects of interest to characterize the sign of the effect sizes for subsequent analysis. An important detail, we negated our EVC predictions (denoted nEVC) to enforce semantic symmetry among all performance monitoring regressors such that positive values indicated poor quality states and negative values indicated high quality states across all three models (Error, PRO, and nEVC).

We computed cluster-size thresholds according to previously reported methods[91]. First, we used the mixed-effects model residuals to estimate the shape parameters of the spatial autocorrelation function (ACF) via AFNI's 3dFWHMx function. We then simulated, via AFNI's 3dClustSim function (-acf option), the cluster-size threshold for our voxel-wise statistical maps (thresholded to $p<0.001$, uncorrected) that survive family-wise error (FWE) corrected thresholds of $p < 0.05$, assuming clusters were formed from contiguous (face-touching, NN=1) voxels. We performed this thresholding process separately for F- and Z-statistical maps.

Finally, we restricted reporting of our first model to only those voxels within a dACC mask (see Figure 3) constructed from the dACC-engaged component of an established 70-component independent component analysis-based partition of the BrainMap database[87]. Note, this mask was thresholded to remove several small activations or voxel clusters not belonging to the dACC proper. We restricted our reporting of the second model to all surviving voxels falling outside of our inflated mFC mask.

Results

In our first experiment, we constructed and validated predictive models of affect processing according to whole-brain neural activation responses to image stimuli in order to confirm prior reports of prediction performance in this class of model. In our second experiment, we applied these models to measure trajectories of affect processing from whole-brain neural activation responses to the cued-recall of image stimuli in order to test the effect size of explicit affect regulation against affect regulation occurring at rest. In our third experiment, we converted the cued-recall task's affect processing trajectories into estimates of affect regulation performance

monitoring relative to the task goal and critically tested their association with activations in the dACC. In our fourth experiment, we characterized the neural activations associated with the cued-recall task itself to contextualize our findings within the extant emotion regulation literature.

Affect Processing Prediction

We constructed linear support vector machine (SVM) classifiers for 74 subjects (excluding 1 subject due to excessive head motion) based upon responses to 90 image stimuli computationally sampled from the IAPS and labeled according to previously published methods[42,70]. We observed classification accuracies (see Supplemental Table S2) in agreement with previous findings reported for a smaller subset of subjects[70] that were both group-level significant ($p<0.001$; Wilcoxon signed rank; h_0: $\mu=0$) and consistent with the highest classification accuracies reported in the literature. Moreover, controlling for the affect processing induction reliability of stimuli[70], we found that 65 of 74 subjects (87.8%) exhibited significant intra-subject classification of affective valence labels and 62 of 74 subjects (83.8%) exhibited significant intra-subject classification of arousal labels ($p<0.05$; binomial distribution, h_0: acc=0.5). As in previous work, we controlled for the dimensionality of the feature space[70] by relating the hyperplane distances predicted by classifiers built from whole-brain gray matter voxel-wise features (30,000–40,000 dimensions) with Gram-Schmidt dimensionally-reduced features (90 dimensions, see Supplemental Figure S3). This experimental control demonstrated that our machine learning methodology scaled well to the whole-brain gray matter feature space.

In this work, we further used SVM models to measure moment-to-moment affect processing states during affect regulation. Our control theoretic analysis presumed that these measurements conform to a continuum of real-values. Therefore, we converted our predictions of hyperplane distance (i.e., classifications) into probabilities via Platt scaling[85], using the probability of the positive class (i.e., positive valence or high arousal) as our measurement convention. To validate this approach, we predicted Platt-scaled normative scores of SVM training stimuli from Platt-scaled intra-subject cross-validated model predictions and demonstrated significant predictive effect size (see Supplemental Figure S4).

We also extended our validation of the predictive models to include those image stimuli reserved for the cued-recall task (a distinct set of 30 IAPS image stimuli sampled independently, but identically, to the SVM training image set). We predicted the Platt-scaled normative scores of these task cue image stimuli using Platt-scaled intra-subject model predictions (based on the cue-related neural activations but according to the SVM hyperplane fitted to the training image set) and, again, we demonstrated significant predictive effect size (see Supplemental Figure S5).

Finally, we extended our validation to the affect processing induction trials on which we trained our predictive models. We predicted the normative scores of the stimulus set from their induced facial electromyography and electrodermal responses, respectively, for the independent affective dimensions of valence and arousal. We found that corrugator supercilii electromyography is a weak ($R^2=.0007$), but significant ($p=.037$; F-test; h_0: $\beta=0$), index of normative valence scores (see Supplemental Figure S6 and Supplemental Figure S7, panel A) when applied to detect valence in polar-extreme extrinsic stimuli (thresholded to remove approximately the middle third of stimuli as measured by normative score). We also found electrodermal activity to be a weak ($R^2=.002$), but significant ($p<0.0002$; F-test; h_0: $\beta=0$), index of normative arousal score (see Supplemental Figure S7, panel B), which was consistent with previous work using similar methods[42].

Explicit Affect Regulation Effects

We applied our predictive models to decode moment-to-moment trajectories of affect processing during the cued-recall task for (n=67) subjects (we excluded 7 additional subjects due to head motion censoring that precluded extraction of cued-recall neural activations). We measured explicit affect regulation according to a linear mixed-effects model in which the fixed-effects were

the Platt-scaled predictions of affect processing in response to the cue stimuli as well as the times elapsed since presentation of the cue. The measurements of interest were standardized Platt-scaled predictions of affect processing for the recall-related volumes. We modeled random slope and intercept effects subject-wise. For each subject we also constructed a cued-recall task control condition composed of neural activations extracted from 30 uniformly randomly sampled time-points within the subject's resting state fMRI scan. We found a significant explicit affect regulation effect related to the affective properties of the cue stimulus (see Figure 2) across both affective dimensions. Further; we demonstrated that this effect was significantly greater than that of the resting state task control effect for both valence and arousal.

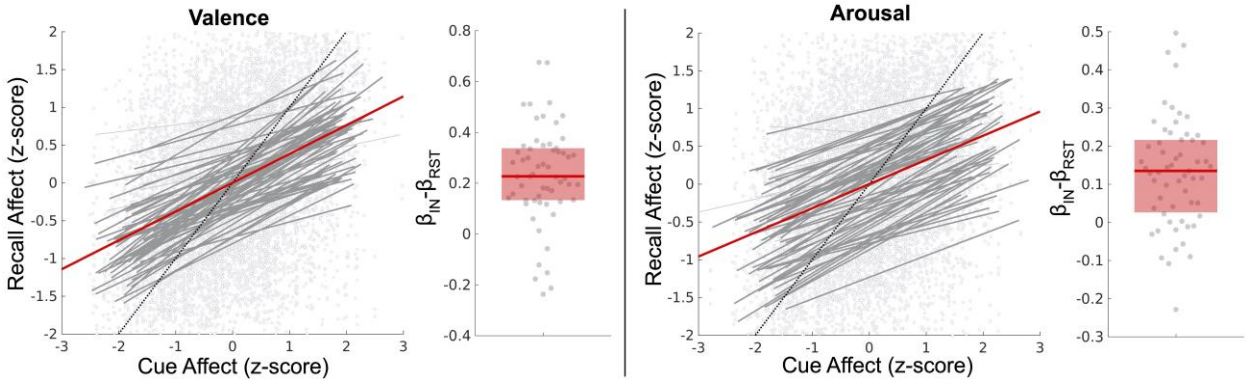

Figure 2: *Estimation and validation of explicit affect regulation effects within the cued-recall task. The figure depicts the effect size of cue affect processing in explaining affect processing occurring during recall (controlling for time lag in the 4 repeated measures of recall per each cue condition). Here affect processing measurements are standardized Platt-scaled hyperplane distance predictions of our fitted SVM 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 cued 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 task in comparison to affect regulation constructed from the resting state task. The bold red line depicts the group median difference in effect size between cued-recall and resting state. The red box depicts the 25–75th percentiles of effect size difference.* ***Valence****. The fixed effect (R^2=.15) is significant ($p<0.001$; F-test; h_0: $\beta = 0$). Random effects significantly improve effect-size ($p<0.05$; likelihood ratio test; h_0: observed responses generated by fixed-effects only). Cued-recall affect regulation effects are significantly greater than that of resting state effects ($p<0.001$; Wilcoxon signed rank; h_0: β_{IN}-β_{RST}=0). The fixed-effect of control duration is significant (β =.02; p=.031; F-test; h_0: $\beta = 0$).* ***Arousal.*** *The fixed effect (R^2=.11) is significant ($p<0.001$, F-test; h_0: $\beta = 0$). Random effects significantly improve effect-size ($p<0.05$; likelihood ratio test; h_0: observed responses generated by fixed-effects only). Cued-recall affect regulation effects are significantly greater than that of resting state effects ($p<0.001$; Wilcoxon signed rank; h_0: β_{IN}-β_{RST}=0). The fixed-effect of control duration is not significant (β=.02; p=.106; F-test; h_0: $\beta = 0$).*

In support of these findings, we took the additional step of verifying, through independent psychophysiological measurements[92], affect processing induction in the context of explicit regulation. To this end, we constructed both cued-recall and comparable control condition (i.e. resting state) measurements of facial electromyography (cEMG and zEMG) associated with hedonic valence[93] as well as electrodermal activity (SCR) associated with autonomic arousal[93]. We then calculated the effects of explicit affect regulation according to linear mixed-effects modeling (see Supplemental Figure S8) and we verified the existence of explicit affect regulation-related psychophysiological responses that were positively related to psychophysiological responses induced by the cue stimuli across all three psychophysiological measurements: zEMG, cEMG, and SCR. Further, we confirmed that the true affect regulation effects were significantly

greater than the control condition regulation effects extracted from resting state measurements, again, across all three physiological measurements.

Functional Neuroanatomical Correlates of Affect Regulation-related Performance Monitoring

We next interrogated the most plausible mathematical form of performance monitoring encoded within the dACC, our candidate functional region-of-interest implicated in this role. To this end, we first computed the moment-to-moment error trajectory for each cued-recall trial for each subject. We strictly partitioned our neural activations into two sets: activations falling within our liberal mask of medial frontal cortex (mFC; including dACC), and those falling outside the mFC. We then constructed and validated out-of-sample inter-subject ensemble moment-to-moment estimates (temporally aligned with the error measurements) of both the predicted response outcome (see Supplemental Figure S9) and expected value of control (see Supplemental Table S3). We based these estimates entirely on machine learning models of task-related neural activations falling outside the mFC. As an additional test, we computed bivariate correlation coefficients, R, between our models of performance monitoring and observed very small ($|R|<0.10$), but significant, positive correlations between estimated expected value of control (EVC) and estimated predicted response outcome (PRO) as well as between error and PRO. We also observed very small ($|R|<0.10$), but significant, negative correlations between EVC and error. The strength, direction, and significance of these correlations were similar across both the orthogonal affective dimensions of valence and arousal (see Supplemental Table S4).

We then estimated, via linear mixed-effects model, mFC neural activations as functions of error, PRO, and negated EVC (i.e., nEVC) measurements as well as affect processing measurements, time elapsed since cue onset (i.e., control duration), and the base task (i.e., model intercept). Note, negating the action-value fixed-effect (nEVC) enforced semantic symmetry among the three competing models of dACC activation such that poor-quality control states were associated with positive values and high-quality control states were associated with negative values. As part of this modeling process, we controlled for age and sex and included interaction terms between age and sex and all other fixed effects. Finally, we examined the specific role of the dACC in performance monitoring by restricting the models' outputs to our dACC anatomic mask. Our findings are summarized in Figure 3.

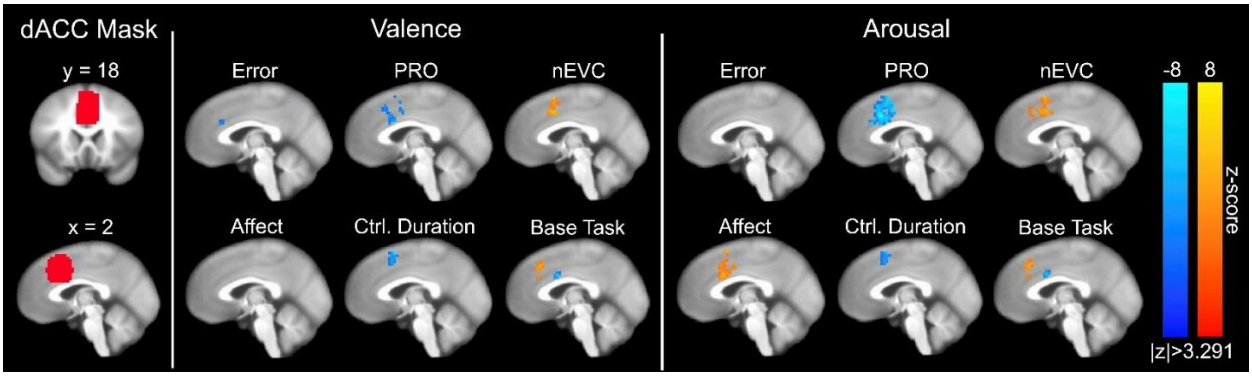

Figure 3: *Group-level linear mixed-effect model distributions for the main fixed-effects of error, predicted response outcome (PRO), negated expected value of control (nEVC), affect processing, control duration, and base task (i.e., intercept) constrained to a mask of the dorsal anterior cingulate cortex (dACC). The figure depicts slices in Talairach coordinate space and neurological convention (image left equals participant left). The figure depicts voxel intensities as colors having a minimum (thresholded) voxel intensity of $|z| > 3.291$ and a maximum voxel intensity of $|z| = 8.0$, i.e., color saturates for z-scores above the maximum intensity. The figure depicts only valence derived clusters having ≥ 16.7 contiguous voxels (measured as face wise nearest neighbors, i.e., NN=1) or arousal derived clusters having ≥ 16.5 contiguous voxels.*

We found that the dACC activation positively encoded performance according to the negated action-value (nEVC) as well as the base task (within the superior dACC bordering on the pre-supplementary motor area, pre-SMA). These findings were duplicated across the independent affective dimensions of valence and arousal. Concurrently, the dACC positively encoded arousal processing (see Fig 3 right, Affect). In contrast, we observed that dACC negatively encoded error forecasting (PRO) across both valence and arousal as well as the perceived error of valence. Further, we observed that the superior dACC (bordering the pre-SMA) negatively encoded the duration of affect regulation (see Fig 3, Ctrl. Duration) whereas the base task encoding was isolated to a small inferior and posterior region of the dACC, bordering the mid-cingulate cortex. We also detected significant dACC clusters implicating the interactions of age and sex effects with fixed effects in our model (see Supplemental Table S5 for a summary of these clusters).

We then modeled the second partition set of the neural response data, those activations falling outside the mFC, using affect processing and control duration as well as age and sex effects and their interactions with the prescribed fixed effects. We restricted our analysis to these fixed effects in order to avoid false positive associations between these data and the performance measurements that we constructed from these data using machine learning models. We summarize our findings in Figure 4.

The base task activated bilateral dmPFC, vlPFC, parietal cortex, and cerebellum as well as the left temporal parietal junction. The base task was also associated with bilateral deactivation in the striatum, dorsolateral prefrontal cortex, temporal poles, and posterior temporal cortex as well as the medial occipital cortex. We observed control duration to be associated with broad bilateral deactivation of the striatum, visual cortex, and ventral visual stream as well as left-lateralized frontoparietal network deactivation. Finally, we localized affect processing encodings that agreed well with prior multivariate representations drawn from similar subjects in a similar task environment[42,43,46]. Again, we detected significant clusters implicating age and sex effects and their interactions with fixed effects in our model. Due to the large number of surviving clusters, we provide direct access to these cluster maps via our Open Science Framework repository (see Source Code and Data Availability).

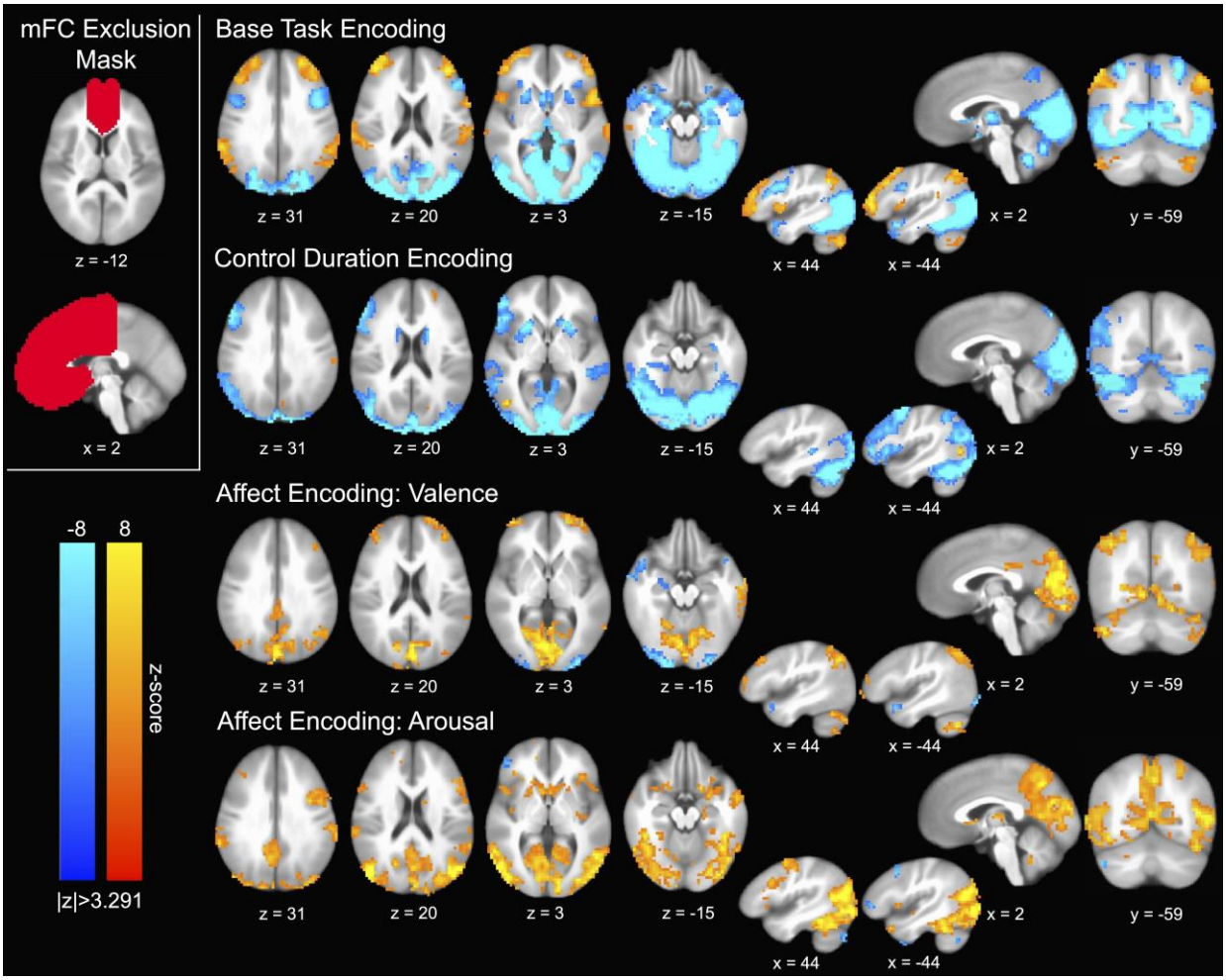

Figure 4: *Group-level linear mixed-effect model distributions of the main fixed-effects of base task (i.e., model intercept), control duration, and affect processing for all gray matter excluding the medial frontal cortex (upper left). Note, we report only the distributions for valence as they were highly similar to those for arousal (see OSF repository full maps). The figure depicts slices in Talairach coordinate space and neurological convention (image left equals participant left). The figure depicts voxel intensities as colors having a minimum (thresholded) voxel intensity of |z| = 3.291 and a maximum voxel intensity of |z| = 8.0 i.e., color saturates for z-scores above the maximum intensity. The figure depicts only valence derived clusters having ≥ 16.7 contiguous voxels (measured as face wise nearest neighbors, i.e., NN=1) or arousal derived clusters having ≥ 15.9 contiguous voxels.*

Discussion

The findings of this study constitute two important contributions to the affect regulation literature. For our first contribution, we constructed affect processing measurement instruments, derived from machine learning predictive models, to detect moment-to-moment fluctuations in affect processing during regulation. We then applied these instruments to the cued-recall task and demonstrated significant explicit affect regulation effects concordant across two independent, objective response measurement modalities (fMRI and psychophysiology) as well as across the orthogonal affect processing dimensions of valence and arousal. These measurement instruments provide a novel framework to observe, analyze, and parse the functional neuroanatomical mechanisms underlying a control theoretic description of affect regulation. The second contribution of this work was our application of these novel measurement tools to test

competing computational models of performance monitoring and critically compare each model's ability to explain the putative role of the dACC in affect regulation.

We created functional neuroanatomical maps associating dACC neural activation with performance monitoring (Fig 3). These maps most plausibly indicate that dACC activation encodes the negated action-value of the control decision (nEVC model) which was conserved across both the valence and arousal dimensions of affect regulation. We interpret these findings as suggesting that dACC activation signals the minimum expected sum of discounted future costs of the control decision, which, according to our parameter selection process, integrates both estimated future errors and future effort in its cost expectation.

In addition to supporting the nEVC model, our findings provide evidence to reject the dACC PRO model. We observed negatively encoded prediction response outcomes for dACC, which cannot plausibly be expected under the canonical definition of the PRO model. Moreover, we demonstrated that the PRO and nEVC regressors were weakly but positively correlated (see Supplemental Table S4); therefore, the dACC's differential encoding of these performance models cannot be attributed to our mixed-effects modeling. In contrast, anti-correlation between the error and nEVC regressors, as reported in Supplemental Table S4, could plausibly explain the small, but significant, region of dACC negative encoding of perceived valence error.

These interpretations of the study's findings drew upon two observations. First, one could view our implementation of PRO as a special case of nEVC in which there was no cost of acting (f_a=0) and future action-value was completely discounted (γ=0). During a parameter search, see Supplemental Table S3, we found that on-policy control decisions input to a Q-function built using these parameters yielded action-values significantly greater than that of a random control policy. We concluded from this finding that, in general, control actions target decreased future error (as the average random action would induce zero change relative to the goal). Second, one could interpret the dACC's negative encoding of PRO as either deactivation associated with relatively higher error or activation associated with relatively lower error. Combined, these observations suggest that the dACC positively encodes the engagement of control actions, which we indirectly detected as a forecast of decreased future error.

We also observed that the dACC positively encoded the base task via the mixed-effects model's intercept (see Figure 3). We hypothesized two complementary sources of this encoding[21]. Cognitive control theory would suggest that the observed dACC encoding represented conflict detection and processing driven by the base task's cognitive demand, assuming that inaction was the prepotent response (see Table 1). A process model perspective of the base task would suggest that the observed dACC encoding signaled attentional deployment preceding (and in concert with) the implementation of self-control processes. Incorporating evidence from the broader view of base task-related cortical activation (Figure 4) we observed that bilateral parietal cortex, associated with attentional shifting[12,94], positively encoded the base task whereas the dorsolateral prefrontal cortex, strongly associated with cognitive emotional control[35], negatively encoded the base task. These observations suggested that attentional deployment, rather than cognitive control, was the most plausible interpretation of the role of the dACC in the base task. Alternatively, the dACC response may reflect the role of emotion reactivity (i.e., arousal), also linked to attentional deployment[95]; however, we explicitly modeled this confound and observed a strong dACC encoding of affective arousal that is spatially distinct (inferior and posterior).

Limitations

Beyond the difficulty in disambiguating and interpreting the dACC's encoding of prediction response outcome, our analysis excluded one important model of dACC activation (prediction-error likelihood); it also relied on a strong *a priori* assumption concerning the prepotent response to task demand in order to interpret our findings within the context of cognitive conflict. We acknowledge these limitations and admit that, due to our multivariate approach to modeling dACC

activation as a function of multiple performance monitoring theories, our inferences are conditional based upon the absence of explicit and convincing representations of these models.

We also acknowledge that the presence of significant interactions of sex and age with several of our primary performance monitoring fixed effects suggests that the inferences we have drawn may be conditional on the composition of our study's sample. Indeed, similar analysis conducted on less diverse populations could, potentially, make different inferences. The age and sex diversity of the sample we report in this study, however, does provide us with some confidence concerning the inferences draw from our main fixed effects.

As with all machine learning studies, our model predictions (on which we built the primary findings of this work), relied extensively on the quality of the fMRI-derived features and their labels. We have reported extensively on the limitations of exploiting IAPS normative scores as affective labels for training predictive models[42,70,96], which also likely contributed to our observed interactions of age and sex with the neural encodings of both performance monitoring and the base task. We have also reported limitations of the non-critical use of galvanic skin response as a surrogate measure of autonomic arousal[42] and as an independent validation of our machine learning models. We acknowledge similar limitations in our application of facial electromyography to signal stimulus-related induction of hedonic valence processing in the cued-recall task.

Broader Contributions

Beyond our exploration of the computational models underlying dACC-based performance monitoring, both the conceptual model we introduced and the methods we developed in this study may potentially make broad contributions to future studies of emotion regulation. When viewed through the lens of engineering control theory (Figure 1 panel B), our cued-recall task, which elicits a quantifiable affect regulation goal and facilitates moment-to-moment self-assessment of emotion regulation dynamics, have enabled us to:

1) clarify the differences between affect processing and affect regulation[15,32,97] and, for the specific form of emotion regulation we studied, dissociate two of three control components possibly involved in this process;
2) formalize and study affective self-awareness, previously posited as a determining factor for successful emotion regulation[32], enabling novel discovery in this area; and,
3) envision future control theoretic explorations of clinically relevant emotion regulation strategies, e.g. distraction and reappraisal, through formal mapping to emotional goals and goal-directed actions of relevant cognitive processes.

Conclusion

We combined established machine learning methodology for the prediction of affect processing states from fMRI BOLD signal with independent psychophysiological measures of affect processing to validate the cued-recall task as a model of explicit affect regulation. We then exploited the cued-recall task's design to separate the affective goal from moment-to-moment affect regulation to computationally model multiple measures of performance monitoring and critically test the specific association of dACC recruitment with these model variants. We found that the dACC most plausibly computes the estimated sum of discounted future costs, including both control process error and effort. This role was conserved across the orthogonal affective dimensions of valence and arousal. Concurrently with performance monitoring, we demonstrated that the dACC directly encodes affective arousal and also likely encodes recruitment of attention and regulation resources.

Acknowledgements

This study was funded 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 would like to thank 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. The authors also 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., G.A.J; Analysis: K.A.B, G.A.J., A.A.P, K.P.F.; Interpretation of results, manuscript preparation, and revisions: K.A.B, G.A.J., A.A.P, K.P.F, 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, as well as logfiles and raw plots generated by this code, publicly available: https://github.com/kabush/IN. De-identified derivative data created and used in this study's analysis (e.g., masks, beta-series activation maps, support vector machine hyperplanes, computational models, and functional neuroanatomical activation maps) are also publicly available via the study's Open Science Framework repository: https://osf.io/jwv6c/. The authors also intend to make a Brain Imaging Data Structure (BIDS)[98] formatted variant of the CTM study's dataset publicly available at the conclusion of enrollment (circa 2021). Upon public release, the BIDS dataset will be made available via the study's Open Science Framework repository. Note, this public data release will not include (n=19) subjects enrolled in the INCA study as they did not consent to public repositing of their raw data. For replication purposes only, de-identified raw data from the INCA study may be made privately available upon request.

References

1. Lang, P. J. & Davis, M. Emotion, motivation, and the brain: Reflex foundations in animal and human research. in *Progress in Brain Research* vol. 156 3–29 (Elsevier, 2006).
2. *Handbook of social psychology*. (Springer, 2006).
3. Lerner, J. S., Li, Y., Valdesolo, P. & Kassam, K. S. Emotion and Decision Making. *Annu. Rev. Psychol.* **66**, 799–823 (2015).
4. Plutchik, R. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist* vol. 89 344–350 (2001).
5. Berking, M. & Wupperman, P. Emotion regulation and mental health: recent findings, current challenges, and future directions. *Current Opinion in Psychiatry* **25**, 128–134 (2012).
6. Gross, J. J. Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry* **26**, 1–26 (2015).
7. Gross, J. J. Antecedent- and Response-Focused Emotion Regulation: Divergent Consequences for Experience, Expression, and Physiology. 14.
8. Gross, J. J. & John, O. P. Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology* **85**, 348–362 (2003).
9. Gross, J. J. The Emerging Field of Emotion Regulation: An Integrative Review. *EMOTION REGULATION* 29.

10. Webb, T. L., Miles, E. & Sheeran, P. Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin* **138**, 775–808 (2012).
11. Vaughan-Johnston, T. I. *et al.* The role of individual differences in emotion regulation efficacy. *Journal of Research in Personality* **84**, 103904 (2020).
12. McRae, K. *et al.* The Neural Bases of Distraction and Reappraisal. *Journal of Cognitive Neuroscience* **22**, 248–262 (2010).
13. Buhle, J. T. *et al.* Cognitive Reappraisal of Emotion: A Meta-Analysis of Human Neuroimaging Studies. *Cerebral Cortex* **24**, 2981–2990 (2014).
14. Kohn, N. *et al.* Neural network of cognitive emotion regulation — An ALE meta-analysis and MACM analysis. *NeuroImage* **87**, 345–355 (2014).
15. Sheppes, G., Suri, G. & Gross, J. J. Emotion Regulation and Psychopathology. *Annu. Rev. Clin. Psychol.* **11**, 379–405 (2015).
16. Glenn, C. R. & Klonsky, E. D. Emotion Dysregulation as a Core Feature of Borderline Personality Disorder. *Journal of Personality Disorders* **23**, 20–28 (2009).
17. Aldao, A. & Nolen-Hoeksema, S. Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy* **48**, 974–983 (2010).
18. McLaughlin, K. A., Hatzenbuehler, M. L., Mennin, D. S. & Nolen-Hoeksema, S. Emotion dysregulation and adolescent psychopathology: A prospective study. *Behaviour Research and Therapy* **49**, 544–554 (2011).
19. Bradley, B. *et al.* Emotion Dysregulation and Negative Affect: Association With Psychiatric Symptoms. *The Journal of Clinical Psychiatry* **72**, 685–691 (2011).
20. Shaw, P., Ch, M. B. B., Stringaris, A., Nigg, J. & Leibenluft, E. Emotion Dysregulation in Attention Deficit Hyperactivity Disorder. *Am J Psychiatry* 18 (2014).
21. Ochsner, K. & Gross, J. The cognitive control of emotion. *Trends in Cognitive Sciences* **9**, 242–249 (2005).
22. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron* **79**, 217–240 (2013).
23. Cole, M. W., Repovš, G. & Anticevic, A. The Frontoparietal Control System A Central Role in Mental Health. *The Neuroscientist* **20**, 652–664 (2014).
24. Gu, S. *et al.* Optimal trajectories of brain state transitions. *NeuroImage* **148**, 305–317 (2017).
25. Brown, J. W. & Braver, T. S. Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. *Science* **307**, 1118–1121 (2005).
26. Alexander, W. H. & Brown, J. W. Computational Models of Performance Monitoring and Cognitive Control. *Topics in Cognitive Science* **2**, 658–677 (2010).
27. Etkin, A., Büchel, C. & Gross, J. J. The neural bases of emotion regulation. *Nature Reviews Neuroscience* **16**, 693–700 (2015).
28. Richard S. Sutton & Andrew G. Barto. *Reinforcement Learning: An Introduction*. (MIT Press, 1998).
29. Bellman, R. *Dynamic Programming*. (Princeton University Press, 1957).
30. Pontryagin, L. S. *Mathematical theory of optimal processes*. (Interscience Publishers John Wiley & Sons, 1962).
31. Barto, A. G., Sutton, R. S. & Anderson, C. W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEE Transactions on Systems, Man, and Cybernetics* **SMC-13**, 834–846 (1983).
32. Gross, J. J. *Handbook of emotion regulation*. (Guilford publications, 2013).
33. Posner, J., Russell, J. & Peterson, B. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev Psychopathology* **17**, 715–734 (2005).
34. Gyurak, A., Gross, J. J. & Etkin, A. Explicit and implicit emotion regulation: A dual-process framework. *Cognition & Emotion* **25**, 400–412 (2011).

35. Ochsner, K. N. & Gross, J. J. Cognitive Emotion Regulation: Insights From Social Cognitive and Affective Neuroscience. *Curr Dir Psychol Sci* **17**, 153–158 (2008).
36. William L Brogan. *Modern Control Theory*. (Prentice-Hall, 1991).
37. Ochsner, K. & Gross, J. The Neural Bases of Emotion and Emotion Regulation: A Valuation Perspective. in *Handbook of Emotion Regulation* (2014).
38. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* **24**, 167–202 (2001).
39. Desimone, R. & Duncan, J. Neural Mechanisms of Selective Visual Attention. *Annual review of neuroscience* **18**, 193–222 (1995).
40. Desimone, R. Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **353**, 1245–1255 (1998).
41. 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).
42. 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).
43. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLOS Biology* **13**, e1002180 (2015).
44. Skerry, A. E. & Saxe, R. A Common Neural Code for Perceived and Inferred Emotion. *Journal of Neuroscience* **34**, 15997–16008 (2014).
45. Jongwan, K. *et al.* A study in affect: Predicting valence from fMRI data. *Neuropsychologia* **143**, (2020).
46. Saarimäki, H. *et al.* Discrete Neural Signatures of Basic Emotions. *Cerebral Cortex* **26**, 2563–2573 (2016).
47. Kragel, P. A. & LaBar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience* **10**, 1437–1448 (2015).
48. Tanji, J. & Hoshi, E. Role of the Lateral Prefrontal Cortex in Executive Behavioral Control. *Physiological Reviews* **88**, 37–57 (2008).
49. Saito, N., Mushiake, H., Sakamoto, K., Itoyama, Y. & Tanji, J. Representation of Immediate and Final Behavioral Goals in the Monkey Prefrontal Cortex during an Instructed Delay Period. *Cerebral Cortex* **15**, 1536–1546 (2005).
50. Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L. & Petersen, S. E. A dual-networks architecture of top-down control. *Trends in Cognitive Sciences* **12**, 99–105 (2008).
51. Rushworth, M., Walton, M., Kennerley, S. & Bannerman, D. Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences* **8**, 410–417 (2004).
52. Lau, H., Rogers, R. D. & Passingham, R. E. Dissociating response selection and conflict in the medial frontal surface. *NeuroImage* **29**, 446–451 (2006).
53. Kolling, N. *et al.* Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience* **19**, 1280–1285 (2016).
54. Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E. & Donchin, E. A NEURAL SYSTEM FOR ERROR DETECTION AND COMPENSATION. *Psychological Science* **4**, 385–390 (1993).
55. Bush, G., Luu, P. & Posner, M. I. Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences* **4**, 215–222 (2000).
56. Holroyd, C. B. & Coles, M. G. H. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review* **109**, 679–709 (2002).

57. Pardo, J. V., Pardo, P. J., Janer, K. W. & Raichle, M. E. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences* **87**, 256–259 (1990).
58. Angus W. MacDonald, Jonathan D. Cohen, V. Andrew Stenger & Cameron S. Carter. Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science* **288**, 1835–1838 (2000).
59. Botvinick, M. M., Bravier, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict Monitoring and Cognitive Control. *Psychological Review* **108**, 624–652 (2001).
60. Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L. & Snyder, A. Anterior Cingulate Cortex and Response Conflict: Effects of Frequency, Inhibition, and Errors. *Cerebral Cortex* **11**, 825–836 (2001).
61. Milham, M. P. *et al.* The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research* **12**, 467–473 (2001).
62. Van Veen, V. & Carter, C. S. The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & behavior* **77**, 477–482 (2002).
63. Ridderinkhof, K. R. The Role of the Medial Frontal Cortex in Cognitive Control. *Science* **306**, 443–447 (2004).
64. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *The Journal of Neuroscience* **27**, 8366–8377 (2007).
65. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural Mechanisms of Foraging. *Science* **336**, 95–98 (2012).
66. Murphy, F. C., Nimmo-Smith, I. & Lawrence, A. D. Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience* **3**, 207–233 (2003).
67. Mechias, M.-L., Etkin, A. & Kalisch, R. A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage* **49**, 1760–1768 (2010).
68. Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A. & Ochsner, K. N. Prefrontal-Subcortical Pathways Mediating Successful Emotion Regulation. *Neuron* **59**, 1037–1050 (2008).
69. Kragel, P. A. & LaBar, K. S. Decoding the Nature of Emotion in the Brain. *Trends in Cognitive Sciences* **20**, 444–455 (2016).
70. 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).
71. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (2008).
72. Scheffers, M. K. & Coles, M. G. H. Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance* **26**, 141–151 (2000).
73. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). (1994).
74. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research* **29**, 162–173 (1996).
75. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
76. 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).
77. Power, J. D., Schlaggar, B. L. & Petersen, S. E. Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage* **105**, 536–551 (2015).
78. Fridlund, A. J. & Cacioppo, J. T. Guidelines for Human Electromyographic Research. *Psychophysiology* **23**, 567–589 (1986).

79. Rissman, J., Gazzaley, A. & D'Esposito, M. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* **23**, 752–763 (2004).
80. Privratsky, A. A., Bush, K. A., Bach, D. R., Hahn, E. M. & Cisler, J. M. Filtering improves skin-conductance response measures in the fMRI environment: validation in a sample of women with PTSD. *International Journal of Pscyhophysiology* **(in press)**, (2020).
81. Application Note 241: EMG Signal Processing During fMRI. (2016).
82. 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.
83. Kragel, P. A., Knodt, A. R., Hariri, A. R. & LaBar, K. S. Decoding Spontaneous Emotional States in the Human Brain. *PLOS Biology* **14**, e2000106 (2016).
84. Chan, H.-Y., Smidts, A., Schoots, V. C., Sanfey, A. G. & Boksem, M. A. S. Decoding dynamic affective responses to naturalistic videos with shared neural patterns. *NeuroImage* **216**, 116618 (2020).
85. Platt, J. C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. in *Advances in Large Margin Classifiers.* (MIT Press, 1999).
86. de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *Journal of Neuroscience* **36**, 6553–6562 (2016).
87. Ray, K. L. *et al.* ICA model order selection of task co-activation networks. *Frontiers in Neuroscience* **7**, (2013).
88. Watkins, C. Learning from Delayed Rewards. (King's College, 1989).
89. Ernst, D., Geurts, P. & Wehenkel, L. Tree-based batch mode reinforcement learning. *Journal of Machine Learning Research* **6**, 503–556 (2005).
90. Busoniu, L., Babuska, R., De Schutter, B. & Ernst, D. *Reinforcement Learning and Dynamic Programming Using Function Approximators*. (CRC Press, 2010).
91. Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C. & Taylor, P. A. FMRI Clustering in AFNI: False-Positive Rates Redux. *Brain Connectivity* **7**, 152–171 (2017).
92. 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).
93. 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).
94. Ferri, J., Schmidt, J., Hajcak, G. & Canli, T. Neural correlates of attentional deployment within unpleasant pictures. *NeuroImage* **70**, 268–277 (2013).
95. Sharot, T. & Phelps, E. A. How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience* **4**, 294–306 (2004).
96. 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).
97. Gross, J. J. & Jazaieri, H. Emotion, Emotion Regulation, and Psychopathology: An Affective Science Perspective. *Clinical Psychological Science* **2**, 387–401 (2014).
98. 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).