Advancing understanding of affect labeling with dynamic causal modeling

Salvatore J. Torrisi [a,*], Matthew D. Lieberman [b], Susan Y. Bookheimer [c], Lori L. Altshuler [a]

[a] Semel Institute for Neuroscience & Human Behavior, Dept. of Psychiatry, UCLA, USA
[b] Semel Institute for Neuroscience & Human Behavior, Department of Psychology, UCLA, USA
[c] Semel Institute for Neuroscience & Human Behavior, Center for Cognitive Neuroscience, UCLA, USA

ARTICLE INFO

ABSTRACT

Mechanistic understandings of forms of incidental emotion regulation have implications for basic and translational research in the affective sciences. In this study we applied Dynamic Causal Modeling (DCM) for fMRI to a common paradigm of labeling facial affect to elucidate prefrontal to subcortical influences. Four brain regions were used to model affect labeling, including right ventrolateral prefrontal cortex (vlPFC), amygdala and Broca's area. 64 models were compared, for each of 45 healthy subjects. Family level inference split the model space to a likely driving input and Bayesian Model Selection within the winning family of 32 models revealed a strong pattern of endogenous network connectivity. Modulatory effects of labeling were most prominently observed following Bayesian Model Averaging, with the dampening influence on amygdala originating from Broca's area but much more strongly from right vlPFC. These results solidify and extend previous correlation and regression-based estimations of negative corticolimbic coupling.

1. Introduction

The unique human ability to consciously control one's emotional experience also counts among the more difficult to execute. Nonetheless the psychology and neuroscience of explicit emotion regulation have been fruitfully studied for over two decades, yielding much understanding of the neural mechanisms of emotions and behavioral control (Gross, 2007). Neurobiologically, we are now aware of major prefrontal and emotional regions involved (Ochsner and Gross, 2005; Ochsner et al., 2012) and are beginning to understand the important connections between emotion regulation and health (DeSteno et al., 2013). However, research has more recently suggested that there is a disconnect between self-reported use of explicit emotion regulation strategies and their spontaneous use in daily life (Gruber et al., 2012; Volokhov and Demaree, 2010). One reason for this may be due to the substantial cognitive resources and time required to enact such strategies (Mauss et al., 2006). As a result, research on emotion regulation at a non-conscious level has emerged as an equally interesting and alternative avenue of investigation into how we commonly control our emotional experiences (Berkman and Lieberman, 2009; Koole and Rothermund, 2011).

By explicit emotion regulation we refer to strategies such as reappraisal of an emotional stimuli or suppression of an emotional response, while in contrast non-conscious emotion regulation (also called implicit or incidental emotion regulation) refers to those cognitive processes that result in the lessening of emotional reactivity and

where this effect was not consciously intended by the person engaging in it (Lieberman et al., 2011). Although incidental emotion regulation at a non-conscious level cannot easily be self-reported, functional magnetic resonance imaging (fMRI) offers a window into the process. This technology can reveal the presence of incidental emotion regulation via lessoned emotional reactivity and the extent to which prefrontal regions are recruited that overlap those used to explicitly control emotion (Burklund et al., under review; Payer et al., 2012).

There are a variety of psychological paradigms currently being used to probe non-conscious forms of emotion regulation using fMRI (Berkman et al., 2009; Egner et al., 2008; Meyer et al., 2011). One frequently studied cognitive process is affect labeling (Hariri et al., 2000; Lieberman et al., 2007) which is gaining increasing evidence as a form of incidental emotion regulation (Kircanski et al., 2012; Lieberman et al., 2011). This paradigm lies at the confluence of emotion, control, and language systems. Labeling emotional faces results in decreased amygdala response and the increased recruitment of prefrontal control and language regions, particularly the ventrolateral prefrontal cortex (vlPFC) in the right hemisphere and Broca's area (BA 44/45) in the left. To this end, a task-induced negative coupling between the activity in the amygdala and vlPFC has been shown using regression-based techniques (Foland-Ross et al., 2008; Hariri et al., 2000) and has been interpreted as a dampening of amygdala reactivity by the vlPFC. However, causal inferences using such methods remain circumspect. To advance beyond the simple identification of activation patterns or the changes in coupling between only two regions one must use more sophisticated analyses (Friston, 2011).

In this study, we examined the nature of the vlPFC–amygdala coupling (specifically, the directed influences between these regions)

* Corresponding author at: UCLA Mood Disorders Research Program, 300 Medical Plaza, Suite 1544, USA. Fax: +1 310 794 9915.
E-mail address: storrisi@ucla.edu (S.J. Torrisi).

during affect labeling, and additionally assessed the understudied contribution of Broca's area to amygdala activity. To do so we used Dynamic Causal Modeling (DCM), a validated and reliable Bayesian statistical framework for effective connectivity analysis which encourages the comparison of multiple user-defined models of causal interactions between a set of brain regions (Friston et al., 2003; Rowe et al., 2010; Schuyler et al., 2010). Ours is, to the best of our knowledge, the first application of this method to this common paradigm.

2. Materials and methods

2.1. Subjects

Fifty-two healthy subjects were recruited as part of a larger study of emotion regulation in bipolar disorder by advertisement in local newspapers and campus flyers. They provided informed consent in accordance with the Institutional Review Boards at the University of California, Los Angeles (UCLA). All participants completed the Structured Clinical Interview for DSM-IV Structured Clinical Interview for DSM-IV Axis I Disorders (SCIDI/P; First, 2002). Exclusion criteria included any concurrent or past psychiatric diagnosis (including history of substance abuse), neurological illness, left-handedness, metal implants, a history of skull fracture or head trauma with loss of consciousness of more than 5 min, or taking any medications with psychotropic effects.

2.2. Experimental design

The affect labeling paradigm consisted of three experimental conditions: 'match facial affect,' 'label facial affect,' and 'match forms' (Fig. 1) (Hariri et al., 2000). They were presented as nine experimental 30-second blocks: four displayed emotional faces and were interleaved with five control blocks displaying geometric forms. Of the four displaying faces, two required the subject to match a facial expression with one of two other facial expressions (*match faces* condition). Faces were shown with neutral or negative affect such as fear, surprise or anger. The other two blocks required subjects to choose one of two presented words (e.g., 'ANGRY', 'AFRAID') that best matched an emotional face (*label faces*). For each affect condition, 12 different faces were used, taken from a standard set of photographic stimuli (Ekman and Friesen, 1976). Each emotion was randomized across blocks and the order of task presentation was counterbalanced among subjects. Subjects responded with one of two buttons with their right hand and were told to answer "…as quickly as possible without making too many mistakes". Response times were collected and accuracy was calculated for each condition.

2.3. Image acquisition

All subjects were scanned on a 3 T Siemens Trio scanner. A high resolution structural T1 MPRAGE was acquired with parameters of TR = 1.9 s, TE = 2.26 ms, Flip-Angle = 9°, Matrix = 256 × 256, FOV =

250 mm, voxel size 1 mm isotropic, and total sequence time 6 min and 50 s. The fMRI scan was acquired using a T2*-weighted EPI gradient-echo pulse sequence with IPAT, with TR = 2.5 s, TE = 25 ms, Flip-Angle = 78°, Matrix = 64 × 64, FOV = 192 mm, in-plane voxel size 3 × 3 mm, slice thickness 3 mm, 0.75 mm gap, and 30 total interleaved slices. To allow for scanner equilibration, 2 TRs at the beginning of the scan were discarded. The total sequence time was 5 min and 45 s, with 135 volumes acquired. For co-registration we additionally acquired a matched-bandwidth structural scan with parameters TR = 5 s, TE = 34 ms, Flip-Angle = 90°, Matrix = 128 × 128, FOV = 192 mm, in-plane voxel size 1.5 × 1.5 mm, slice thickness 3 mm, and a total sequence time of 1.5 min. We were not able to acquire MPRAGE scans for four subjects, so their lower resolution matched-bandwidth images were used instead for registration. Foam padding was placed around the heads of participants to suppress motion, responses were recorded by button box, and stimuli was presented by LCD goggles.

2.4. Image preprocessing

All preprocessing and analyses were performed within SPM8/DCM10 (www.fil.ion.ucl.ac.uk/spm/). Subjects' functional volumes were slice-timing corrected (Descamps et al., 2007; Kiebel et al., 2007), then motion realigned, coregistered to the MPRAGE, normalized to a T1-weighted standard brain in MNI space, resliced 3 mm isotropically, and smoothed with a 6 mm FWHM Gaussian kernel. All subjects had maximum translational head movement of less than 2.5 mm, with means and standard deviations across subjects for three translation (x, y, z; in mm) parameters: 0.16(0.2), 0.10(0.11), and 0.29(0.31) and three rotation (pitch, roll, yaw; in radians) parameters: 0.005(0.005), 0.003(0.004), and 0.003(0.005).

2.5. First level (within-subject) analysis

First-level general linear modeling (GLM) of the preprocessed functional images included convolving task design blocks with a canonical hemodynamic response function, high-pass filtering at 128 s to remove low frequency drifts, adding six motion realignment parameters as covariates of no interest and specifying an F-statistical contrast for subsequent VOI extraction (i.e. when adjusting for effects of interest). The first level statistical maps were run twice; the second time with an explicit whole-brain mask derived from an optimal thresholding of the initial masks to ensure coverage of vlPFC (Ridgway et al., 2009).

2.6. Second level group GLM analysis

The standard mass univariate summary statistics approach was used to bring single-subject contrast images into a group random effects analysis. The contrast *label emotion vs match forms* was of interest to elucidate the incidental emotion regulation network while the
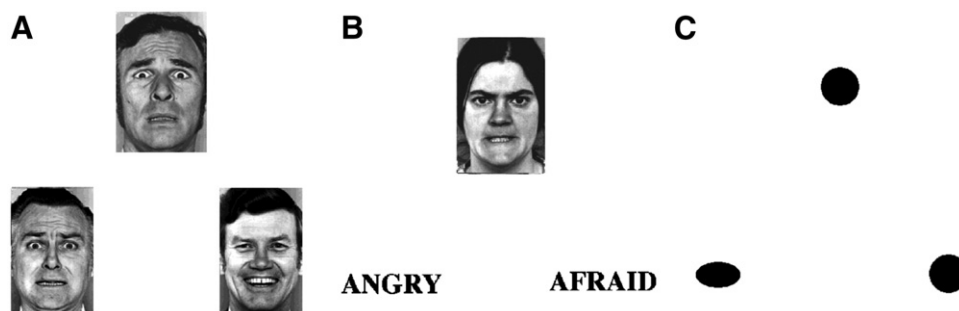

Fig. 1. Affect Labeling paradigm. (A) Match emotion condition; (B) label emotion; (C) match geometric forms. Hariri et al. (2000).

contrast *label emotion vs match emotion* more specifically revealed those regions involved in labeling (thresholds: p < 0.0005, cluster size 10 voxels; p < 0.05 AlphaSim-corrected (http://www.restfmri.net/)). Furthermore, the contrast *match emotion vs match forms* robustly activates the amygdala and was used to guide functionally defined node extraction (see below).

2.7. Region of interest analyses

To demonstrate and confirm with our data the expected labeling-induced increase of right vlPFC activity with the decrease of right amygdala activity we performed region of interest (ROI) analyses on both these regions with independently defined anatomical masks. We brought a histologically-defined probabilistic right amygdala (Amunts et al., 2005) and a Tailarach Daemon right vlPFC (BA 47) mask into the MarsBaR toolbox (Brett et al., n.d.) across the 45 subjects that survived a time series extraction procedure (see below). For the amygdala the contrast *match emotion vs label emotion* was used to verify that its activity is significantly greater during matching emotional faces than during labeling emotional faces, while the reverse contrast was used to verify greater activity of vlPFC.

2.8. Dynamic causal modeling and model space construction

To perform a dynamic causal modeling analysis of a network one first selects nodes for which one has a priori knowledge and hypotheses, but bases the selection on GLM maps of significant task-induced activation. We chose four regions that were shown to be activated in the group *match emotion vs match forms* or *label emotion vs match forms* contrasts which we hypothesized either launched the cognitive process (through visual input) or potentially caused the dampening of amygdala activity previously noted (Foland-Ross et al., 2008; Lieberman et al., 2005, 2007). These regions were the right inferior occipital gyrus (IOG), the right amygdala (both coordinates chosen at the individual subject level with the *match emotion vs match forms* contrast), the right ventral lateral prefrontal cortex (vlPFC; BA 47) and Broca's area (aka left posterior inferior IFG; pars triangularis; BA 44/45) (Fig. 2 top row; both coordinates chosen at the individual subject level with the *label emotion vs match forms* contrast). Two of these regions, Broca's area and vlPFC, also appeared in the targeted *label emotion vs match emotion* contrast (Fig. 2 bottom row), supporting their specific involvement in affect labeling over and
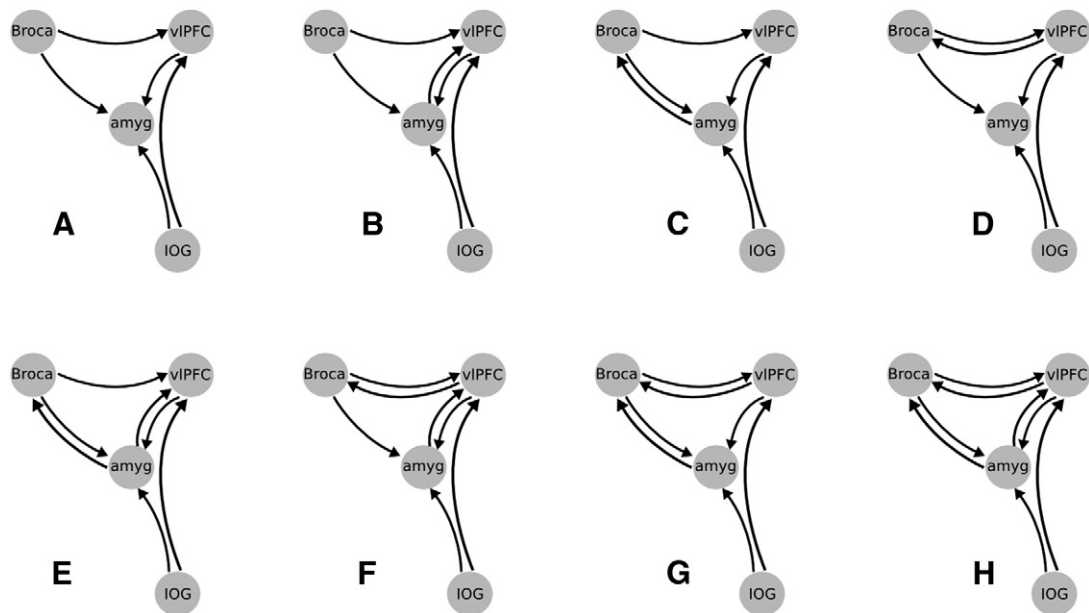
above matching or perceiving affect. However, coordinates for the latter two nodes were not chosen at the individual subject level within this latter contrast because of a decrease in sensitivity. After choosing these four regions based on their response profiles to affect labeling, DCM then allowed us to advance our research question from *what* occurred regionally during labeling to *how* it occurred mechanistically.

Dynamic causal modeling is a Bayesian framework to infer effective connectivity between brain regions in a neural system of interest (Friston et al., 2003). The investigator specifies a given model (i.e. hypothesis) by assigning an endogenous architecture to the regions, the location(s) of a stimulus input that drives the system, and specifies which connections' couplings are modulated by a particular task condition. The evaluation of a model results in posterior parameter estimates on the model's uni- or bidirectional edges, as well as a 'score' of the model as a whole. The negative free energy approximation to the model evidence is the score used to compare multiple models (which constitute a model space) in a Bayesian Model Selection (BMS) scheme. The negative free energy value represents a balance between a model's goodness of fit to the data and the model's complexity which additionally takes into account interdependencies or covariances among parameters (Penny, 2012). Given a particular network, a model space is generally a subset of all theoretically testable models built by the investigator to probe specific questions in a computationally pragmatic manner. The inferences one makes with DCM are therefore relative to the model space one tests. Multiple models can further be grouped into families (Penny et al., 2010), which contain a common element of interest and can be compared in this manner if one asks specific questions about that common element. Finally, one can test not only model architectures as a whole but specific connection strengths and their 'sign' (positive or negative) with Bayesian Model Averaging (BMA) (Penny et al., 2010) within a model family.

In service of these analytical options, we systematically constructed a factorial model space that embodied our neuroscientific assumptions and the hypotheses we wished to test. The basic model was constructed to investigate both forward and backward information propagation centered on the right amygdala. The presence of some connections among regions was held constant throughout all models while others were systematically permuted, resulting in eight basic patterns of endogenous connectivity (Fig. 3). Based on recent research on 'multiple routes' in the processing of written words, emotion and emotional faces (Dima et al., 2011; Pessoa and Adolphs,

Fig. 2. (A) *Label emotion vs match forms*. The map (N = 52) used to determine node peak selection, (four peaks circled) before the loss of seven subjects due to node extraction. (B) *Label emotion vs match emotion*, the targeted affect labeling contrast. See Supplemental Table 1 for cluster-specific information. Axial slices (R = R) at indicated Z coordinates overlaid on standardized brain in MNI space. Both maps p < 0.0005, k = 10.

Fig. 3. Eight templates of endogenous connectivity in schematic axial slice orientation. These were crossed with 2 different driving input and 4 different modulatory hypotheses, creating a 64-element model space.

2010; Richardson et al., 2011), we chose to keep the information propagation from IOG to the right vlPFC constant across all models, along with the more obvious assumption of affective stimuli also entering the amygdala via IOG. The inclusion of IOG as the visual processing node, rather than, e.g. the fusiform face area, also follows previous theory and empirical research on the effective connectivity of face processing (Cohen Kadosh et al., 2011; Dima et al., 2011; Fairhall and Ishai, 2007; Haxby et al., 2000). By conflating the matching and labeling emotional faces' conditions into a single regressor we specified the driving input to our models, which as such contained mostly face stimuli. However, one third of what a participant saw during these conditions were affective words, and so each model was also crossed with two driving input hypotheses (families): entering only at the right IOG or at both the right IOG and the left Broca's area. For the latter family it was assumed that language information entered previously via left hemisphere regions upstream from Broca's not explicitly included in our models. Naturally the same assumption applies to our choice of IOG, a higher-order visual processing region, however the degree of synaptic distance from sensory input is different between these regions. Finally, to investigate the specific changes in coupling that labeling affect engenders, our primary interest in this study, we created four families of amygdala afferent modulation. These were where labeling modulated either the Broca's to amygdala connection, the right vlPFC to amygdala connection, the Broca's to amygdala via right vlPFC pathway (i.e. these two connections simultaneously) or all three of these directed edges simultaneously. This resulted in an 8(endogenous) × 2(driving) × 4(modulatory) = 64 element model space, where each model was separately estimated for all 45 subjects.

2.8.1. Time series extraction

The peak coordinates of these four volumes of interest (VOIs) as observed in the second level group maps were the right IOG [42 −82 −16], right amygdala [24 −4 −13], right vlPFC [42 23 −7], and left Broca's area [−57 20 23], located within the *label emotion vs match forms* contrast (Fig. 2A). Left Broca's area and right vlPFC were also shown in the language specific contrast *label emotion vs match emotion* (Fig. 2B). Within each subject we then manually chose peak coordinates around these group peaks within anatomical constraints and within a threshold of p < 0.05 uncorrected overlaid on the subjects'

own normalized structural images (Leff et al., 2008). Because the quality and extent of a given region's atlas representations are quite variable (Bohland et al., 2009), we chose those anatomical reference tools we reasoned would optimize our selection for a particular region in our tested network. Threshold-crossing subject-specific activations were thusly considered within the IOG if they conformed to the Harvard–Oxford probabilistic atlas, within the amygdala or Broca's area (BA 44/45) if they conformed to the Jülich histological/probabilistic atlas (Eickhoff et al., 2005) and within the vlPFC (BA 47) if they conformed to the Tailarach Daemon (www.talairach.org) after non-linear coordinate conversion from MNI to Talairach space (www.bioimagesuite.org). All subject-specific coordinates were also within 15 mm Euclidean distance of the group peak (Supplemental Fig. 1). If any region for a given subject did not pass any of these constraints they were eliminated from the study. Seven subjects were lost in this manner, which produced our final subject pool of 45 subjects. All DCM and ROI results are based on this group (Table 1 for demographics). Finally, the principal eigenvariate time series was extracted from each peak, adjusted for effects of interest, within a 5 mm radius sphere.

2.8.2. Model comparisons and Bayesian Model Averaging

Our first DCM model comparison step was to exploit the factorial nature of our space and compare two large model families that differed only in the anatomical locations of the driving inputs. We chose to run a random-effects (rfx) Bayesian Model Selection because we could not assume that the optimal model was uniformly used by each individual in our group (Stephan et al., 2010). We looked at a given family or model's exceedance probability, which is the probability of a particular model being more likely than any other model in the space. Exceedance probabilities across all models sum to 1, providing a relative measure of fit specific to one's model space. We then performed rfx BMS on the winning half of the model space and included the BMA procedure. BMA averaged the connectivity parameters within each subject's 32 models, weighted by their posterior probabilities, which yielded estimates for their strength and directions in the form of posterior distributions (means and standard deviations) at both the subject and group levels. Because we tested 11 parameters of interest (8 endogenous plus 3 modulatory) we applied Bonferroni correction ($\alpha = 0.05/11 = 0.0046$) to two-tailed, one-sample tests.

Demographics and task information	
N = 45	
Age	41.9 (12.1)
Gender	20F 25M
Years of education	15.5 (2.2)
Race	
Caucasian	34 (76%)
African American	9 (20%)
Asian	2 (4%)
Match emo accuracy	0.94 (0.09)
Match emo RT (s)	2.09 (0.51)
Label emo accuracy	0.84 (0.08)
Label emo RT (s)	2.01 (0.47)

2.9. Correlations with behavioral measures

To explore the potential connection between the strengths of subject-specific effective connectivity parameters with behavioral measures, we ran correlations with reaction time and percent accuracy with the parameter estimates from BMA. We applied Bonferroni correction for multiple comparisons to these tests as well.

3. Results

3.1. Behavioral data

All participants performed at a high level of accuracy for both the match emotion and label emotion conditions (Table 1).

3.2. GLM analysis

As predicted, second level group GLM maps for the contrast of interest *label emotion vs match forms* replicated previous work by demonstrating a number of known regions activated by the affect labeling paradigm: most prominently bilateral occipital and fusiform regions, amygdalae, vlPFC, dlPFC (BA 6), a medial region of the superior frontal gyrus (BA 8), Broca's area and posterior superior temporal sulcus (Wernicke's Area; BA 22) in the left hemisphere (Fig. 2A and Supplemental Table 1A). The more targeted contrast *label emotion vs match emotion* revealed that a subset of these regions responded specifically to the linguistic stimuli (Fig. 2B and Supplemental Table 1B). After elimination of 7 subjects due to node finding for DCM all GLM second level analyses were re-run and the results were marginally less robust but consistent with those of the primary analysis (Supplemental Figs. 2A, 2B).

For the region of interest analyses the *match emotion vs label emotion* contrast was used to confirm that activity within the amygdala during matching emotion was significantly greater than during labeling emotion (t = 2.0, p = 0.026 corrected). The reverse contrast *label emotion vs match emotion* was used to confirm that activity within the vlPFC during affect labeling was significantly greater than during affect matching (t = 2.37, p = 0.011 corrected).

3.3. DCM analysis

3.3.1. Comparing families of models

We first split our model space to determine the better of two general hypotheses regarding the driving input to affect labeling (Penny et al., 2010), i.e. whether the driving input was best modeled as entering the right IOG alone or in conjunction with the left Broca's area. Random-effects Bayesian Model Selection showed that the family with driving input entering only at the right IOG strongly won over the family where the driving input entered additionally at the left

Broca's area (respective posterior means: 0.9493, 0.0507; exceedance probabilities: 100%, 0%).

3.3.2. Comparing individual models (BMS) and Bayesian Model Averaging (BMA)

The family-level result allowed us to focus on the winning family of 32 models in a comparison using rfx Bayesian Model Selection. This would help answer further questions: (1) are any of the three permuted connections especially important to the basic model structure? (2) If so, which connections are? (3) Does affect labeling produce significant changes in regional coupling over endogenous coupling? (4) If so, which connections are significantly modulated?

We found model 1 to outperform the other models with an exceedance probability of 22% (Fig. 4). This model had a basic endogenous architecture without any of the three permuted connections (Fig. 3, template 'a') and with modulation under affect labeling targeting the Broca's area to amygdala connection. Model 17 was the second-best model with an exceedance probability of 17%. In this model the same endogenous architecture as model 1 is present, although labeling modulates the Broca's area to amygdala via vlPFC pathway. However, neither of these modulatory hypotheses emerged as very clear winners at the BMS level, i.e. if we calculate the ratio of model posterior means between these two models (their Bayes Factor) we return 0.0804/0.0685 = 1.1737, which constitutes weak evidence in favor of model 1 (Penny et al., 2004). However, what emerged from this 32-model comparison was that the third and fourth strongest winning models also possessed that basic endogenous architecture that lacked amygdala afferents and the vlPFC to Broca's projection (Fig. 4, asterisks). In other words, while the top four models were dissimilar in the location of modulation under labeling, they all shared an identical endogenous architecture (Fig. 3, template 'a'), seemingly providing strong support about the best-fit endogenous architecture of the labeling task. To quantify this observation we next 'post hoc' grouped our model space into eight families of those endogenous architectures. Of these, the family with the first model's architecture strongly won the Bayesian model comparison with a family posterior mean of 0.479 and exceedance probability of 88% (Supplemental Fig. 3). As a parallel, exploratory analysis we also performed BMA (see below) on this smaller group of models and the main results below did not change (Supplemental Table 2 and Fig. 5).
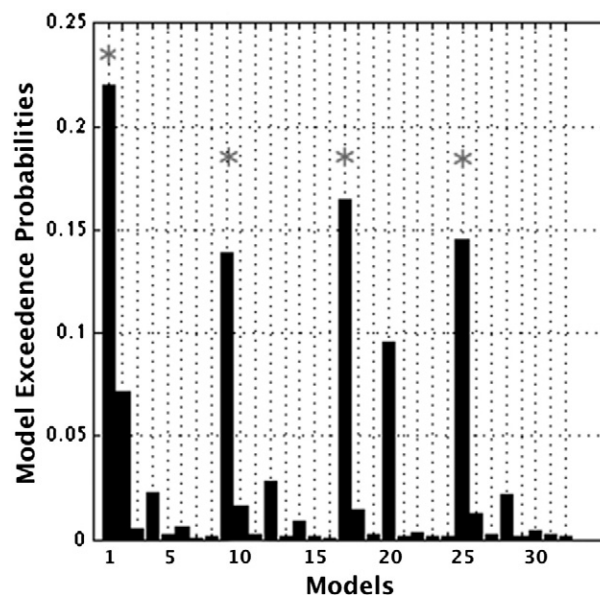

Fig. 4. Bayesian Model Selection for 32 models from the winning driving input family (input to IOG). Asterisks above the four models with the highest exceedance probabilities.

For greater insight into our main apriori interest of labeling-induced modulation of regional coupling we looked at the results of our Bayesian Model Averaging (BMA) of the 32 models that won the family-level driving input test. This provided subject and group-level posterior distributions for the endogenous and modulatory parameters. For a given parameter, i.e. a directed edge between two regions, we performed two-tailed, one-sample T-tests for significance on their mean values. Four endogenous connections survived Bonferroni correction: the IOG projections to amygdala and vlPFC as well as the bidirectional connections between vlPFC and Broca's area (Table 2 and Fig. 5). With the exception of the Broca's area to vlPFC connection, all of these values were positive, meaning that an increased rate of a source region's neural population-level activity results in an increase in the neural population-level activity of its target. Finally, the modulatory effect of affect labeling showed the vlPFC to amygdala influence to be very significantly negative, meaning we can infer that an increased rate of vlPFC activity under the labeling condition causes a decrease in the rate of amygdala activity ($p < 0.0001$). The modulatory effect of labeling showed Broca's area to also significantly decrease the rate of amygdala activity, albeit not as strongly as the vlPFC. Finally, the Broca's to vlPFC modulatory connection was found to be significantly negative, but did not survive correction. In sum, affect labeling appears to induce a dampening effect across the corticolimbic network where one of these connections becomes *more negative* under modulation than its endogenous coupling (Broca's to vlPFC) and the other two couplings (vlPFC to amygdala and Broca's to amygdala) become significantly negative where there was previously no significant endogenous coupling present.

3.3.3. DCM parameter correlations with behavior and external variables

There were no significant correlations between BMA parameter estimates and reaction time or percent accuracy.

4. Discussion

To our knowledge, the current study is the first application of an effective connectivity technique to investigate causal neural interactions during a task of affect labeling, a paradigm acquiring increasing evidence as a form of incidental emotion regulation. By including a strongly language-involved brain region in our network modeling we have advanced a mechanistic understanding of what occurs while 'putting feelings into words' and, namely, how the often-replicated observation of decreased amygdala activity may arise. Our findings reinforce common interpretations of previous regression and correlation-based

Table 2
Parameter-level results (as distributions) from Bayesian Model Averaging across 32 models and 45 subjects. SD = standard deviation. **Bonferroni corrected, *p < 0.05 uncorrected.
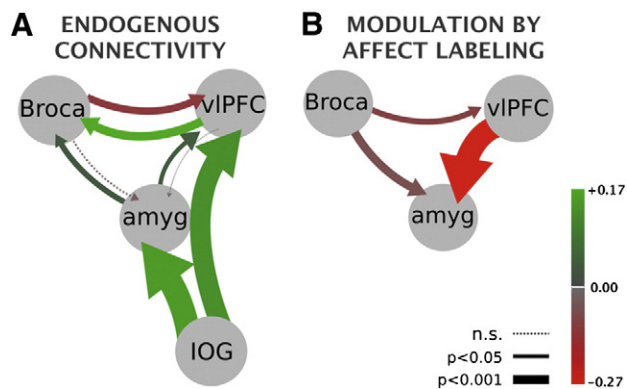
BMA results				
	Mean	SD	p value	
Endogenous				
IOG to amyg	0.17	0.11	0.0000000110	**
IOG to vlPFC	0.15	0.10	0.0000001550	**
Amyg to vlPFC	0.04	0.10	0.0336	*
Amyg to Broca's	0.03	0.07	0.0124	*
vlPFC to amyg	0.00	0.18	0.9549	
vlPFC to Broca's	0.20	0.09	0.0013	**
Broca's to amyg	−0.03	0.20	0.2228	
Broca's to vlPFC	−0.11	0.20	0.0038	**
Modulatory				
vlPFC to amyg	−0.27	0.12	0.0000000173	**
Broca's to amyg	−0.06	0.11	0.0013	**
Broca's to vlPFC	−0.09	0.10	0.0090	*

analyses of vlPFC–amygdala negative connectivity, however they both solidify and extend them, as dynamic causal modeling permits stronger claims about causal interactions and allows the testing of multiple connections simultaneously.

As predicted, we found the right vlPFC *to* right amygdala connection to be strongly modulated, negatively, by labeling facial affect. This means that during this task the vlPFC exhibits a dampening effect on the activity of the amygdala that is not endogenously present but rather manifests during the labeling of affect. The DCM model results additionally support a causal role for Broca's area during labeling affect, in that its involvement directly and negatively modulates the amygdala as well, and may additionally indirectly modulate the amygdala via the vlPFC. We also found strong evidence at the model level of the data supporting a most minimal endogenous architecture. Taken together, our results demonstrate that the most significant regulatory effect on amygdala originates from the right vlPFC, yet also present are weaker but nonetheless substantial labeling-induced contributions from Broca's area to both the amygdala and to the vlFPC. Finally, we found that the endogenous connection from IOG to vlPFC that we modeled was strongly supported by our data, corroborating recent work on parallel routes for the visual processing of facial affect (Dima et al., 2011; Pessoa and Adolphs, 2010).

This study has several limitations. It is only practical to test a small number of nodes with Dynamic Causal Modeling, and although this is common practice, we may have missed crucial players in the network that forms the basis of affect labeling. The dampened amygdala activation observed could additionally be caused or mediated by the left middle temporal gyrus (Wernicke's Area; BA 22) or the medial superior frontal gyrus (BA 8) which were both shown as active in the *label emotion vs match emotion* GLM contrast (Supplemental Table 1). Future DCM studies of affect labeling should systematically add these regions as other nodes to tested models, as well as the contralateral homologues of the regions investigated here. Additionally, other areas implicated by previous research may be important to this regulatory network. For example, in animal models ventromedial PFC activity has been demonstrated to dampen amygdala activity (Quirk et al., 2003) or mediate its dampening in humans (Lieberman et al., 2007). Other incidental emotion regulation paradigms have additionally shown increases in rostral anterior cingulate cortex (Berkman and Lieberman, 2009; Cohen et al., 2011), however, our task did not significantly recruit either the vmPFC or rACC, where significant recruitment is required for node inclusion in DCM analyses (Stephan et al., 2010).

As for brain regions whose effective connectivity we did investigate, the vlPFC and Broca's area were revealed in the targeted contrast *label emotion vs match emotion* and yet they are both regions that have been studied in cognitive contexts other than affect labeling.

Fig. 5. Graphic depiction of Bayesian Model Averaging results (Table 2). Color saturation values map to mean effective connectivity parameters (in Hertz) while arrow thickness maps to statistical significance proportional to $log(1/p\text{-}value)$. (A) Endogenous connectivity. (B) Modulation of connectivity by affect labeling representing the three tested connections. Due to the basic bilinear DCM equation the modulation values add to the endogenous values. Driving input (not shown graphically) enters the IOG in both states.

For example, and as previously mentioned, the right vlPFC has more broadly been considered a candidate mechanism for self-control of several types, as it additionally activates during many tasks of response inhibition, risk-taking (gambling) and temporal discounting (delaying gratification) (reviewed in Cohen et al., 2011). Additionally, the right vlPFC has been implicated in attentional switching (reflexive reorienting) (reviewed in Levy and Wagner, 2011), although the relation of this function to our affect labeling paradigm is unclear. The vlPFC has also been implicated in language studies, and in particular the processing of semantic information, however the evidence more strongly suggests hemispheric lateralization to the left (Dapretto and Bookheimer, 1999; Noesselt et al., 2003; Price, 2012). To this end we did observe bilateral vlPFC activation in the *label emotion vs match emotion* contrast (Supplemental Table 1) and so we may speculate that the left vlPFC could be processing semantic (affect-related) aspects of affect labeling while the right vlPFC acts as a regulating mechanism. Our paradigm, however, precludes more targeted contrasts or factorial interactions to better tease apart these relations. Likewise, Broca's area is a heterogeneous region (BAs 44/45) that has been studied in multiple contexts, most prominently as serving multiple sub-functions of affect-neutral language processing (extensively reviewed in Price, 2012). Therefore, given their known functional heterogeneity, our study offers greater understanding of how to interpret the changing roles of these regions given a "neural context" of affect labeling (Henson, 2005; McIntosh, 2004).

In sum, the present study investigated causal mechanisms ostensibly responsible for the effect of greater vlPFC and dampened amygdala during tasks of affect labeling, and in doing so helps situate affect labeling within a broader context of emotion regulation (Gyurak et al., 2011). While there are overlaps between the mechanisms and effects of explicit and incidental forms of emotion regulation, the latter offers advantages to evoking this cognitive process less directly. For instance, explicit emotion regulation paradigms, such as reappraisal, additionally test an individual's *capacity* to regulate their emotions following a set of directed instructions. As such, the subject's capability of quickly and creatively producing a scenario that achieves the desired goal is tested along with their regulatory abilities. In contrast, regulation of emotion that occurs incidental to intentionally instructed goals may reflect an individual's *tendency* to regulate more automatically in domains outside the laboratory (Berkman and Lieberman, 2009). Following this logic, if incidental emotion regulation lends insight into an individual's tendency to regulate, be it genetic or learned, it may also be used to probe 'latent' dysfunction caused by, or reflective of, disease states such as post traumatic stress disorder, anxiety, or mood disorders. In the case of affect labeling, for example, research has already been conducted with spider phobic subjects and suggests a pronounced effect in mitigating fear responses by labeling affect versus cognitive reappraisal (Kircanski et al., 2012; Tabibnia et al., 2008). Future behavioral and neuroimaging work leveraging an individual's tendency to spontaneously regulate emotion with incidental emotion regulation strategies and the ability for this to additionally inform disorders characterized by deficits in emotion regulation will be highly valued.

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.06.025.

References

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

Berkman, E.T., Lieberman, M.D., 2009. Using neuroscience to broaden emotion regulation: theoretical and methodological considerations. Soc. Pers. Psychol. Compass 3, 475–493.

Berkman, E.T., Burklund, L., Lieberman, M.D., 2009. Inhibitory spillover: intentional motor inhibition produces incidental limbic inhibition via right inferior frontal cortex. Neuroimage 47, 705–712.

Bohland, J.W., Bokil, H., Allen, C.B., Mitra, P.P., 2009. The brain atlas concordance problem: quantitative comparison of anatomical parcellations. PLoS One 4, e7200.

Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B. (Eds.), n.d. Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, vol. 16, No 2.

Burklund, L.J., Creswell, J.D., Irwin, M.R., Lieberman, M.D., 2013. The Common Neural Bases of Affect Labeling and Reappraisal (under review).

Cohen Kadosh, K., Cohen Kadosh, R., Dick, F., Johnson, M.H., 2011. Developmental changes in effective connectivity in the emerging core face network. Cereb. Cortex 21, 1389–1394.

Cohen, J., Berkman, E.T., Lieberman, M.D., 2011. Intentional and incidental self-control in ventrolateral PFC. In: Stuss, D., Knight, R. (Eds.), Principles of Frontal Lobe Function, pp. 417–440.

Dapretto, M., Bookheimer, S.Y., 1999. Form and content: dissociating syntax and semantics in sentence comprehension. Neuron 24, 427–432.

Descamps, B., Roggeman, C., Vandemaele, P., Achten, E., 2007. Image acquisition order and input order of BOLD-fMRI Data for slice time correction in fMRI data processing tools. Abstract Presented at First Benelux In Vivo MR Methods Symposium.

DeSteno, D., Kubzansky, L., Gross, J.J., 2013. Affective Science and Health: The Importance of Emotion and Emotion Regulation. Health Psychol. 32 (5), 474–486.

Dima, D., Stephan, K.E., Roiser, J.P., Friston, K.J., Frangou, S., 2011. Effective connectivity during processing of facial affect: evidence for multiple parallel pathways. J. Neurosci. 31, 14378–14385.

Egner, T., Etkin, A., Gale, S., Hirsch, J., 2008. Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. Cereb. Cortex 18, 1475–1484.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.

Ekman, P., Friesen, W.V., 1976. Pictures of Facial Affect. Consulting Psychologists Press, Palo Alto.

Fairhall, S.L., Ishai, A., 2007. Effective connectivity within the distributed cortical network for face perception. Cereb. Cortex 17, 2400–2406.

First, M.B., 2002. DSM_IV_TR Handbook of Differential Diagnosis. American Psychiatric Publishers.

Foland-Ross, L.C., Altshuler, L.L., Bookheimer, S.Y., Eisenberger, N.I., Townsend, J.D., Thompson, P.M., 2008. Evidence for deficient modulation of amygdala response by prefrontal cortex in bipolar mania. Psychiatry Res. Neuroimaging 162, 27–37.

Friston, K.J., 2011. Functional and effective connectivity: a review. Brain Connect. 1, 13–36.

Friston, K.J., Harrison, L., Penny, W.D., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.

Gross, J.J. (Ed.), 2007. Handbook of Emotion Regulation. The Guilford Press, New York.

Gruber, J., Harvey, A.G., Gross, J.J., 2012. When trying is not enough: emotion regulation and the effort–success gap in bipolar disorder. Emotion 12, 997–1003.

Gyurak, A., Gross, J.J., Etkin, A., 2011. Explicit and implicit emotion regulation: a dual-process framework. Cogn. Emot. 25, 400–412.

Hariri, A.R., Bookheimer, S.Y., Mazziotta, J.C., 2000. Modulating emotional responses: effects of a neocortical network on the limbic system. Neuroreport 11, 43–48.

Haxby, J., Hoffman, E., Gobbini, M., 2000. The distributed human neural system for face perception. Trends Cogn. Sci. 4, 223–233.

Henson, R., 2005. What can functional neuroimaging tell the experimental psychologist? Q. J. Exp. Psychol. A 58, 193–233.

Kiebel, S.J., Klöppel, S., Weiskopf, N., Friston, K.J., 2007. Dynamic causal modeling: a generative model of slice timing in fMRI. Neuroimage 34, 1487–1496.

Kircanski, K., Lieberman, M.D., Craske, M.G., 2012. Feelings into words: contributions of language to exposure therapy. Psychol. Sci. 23 (10), 1086–1091.

Koole, S.L., Rothermund, K., 2011. "I feel better but I don't know why": the psychology of implicit emotion regulation. Cogn. Emot. 25, 389–399.

Leff, A.P., Schofield, T.M., Stephan, K.E., Crinion, J.T., Friston, K.J., Price, C.J., 2008. The cortical dynamics of intelligible speech. J. Neurosci. 28, 13209–13215.

Levy, B.J., Wagner, A.D., 2011. Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. Ann. N. Y. Acad. Sci. 1224, 40–62.

Lieberman, M.D., Hariri, A.R., Jarcho, J.M., Eisenberger, N.I., Bookheimer, S.Y., 2005. An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. Nat. Neurosci. 8, 720–722.

Lieberman, M.D., Eisenberger, N.I., Crockett, M.J., Tom, S.M., Pfeifer, J.H., Way, B.M., 2007. Putting feelings into words: affect labeling disrupts amygdala activity in response to affective stimuli. Psychol. Sci. 18, 421–428.

Lieberman, M.D., Inagaki, T.K., Tabibnia, G., Crockett, M.J., 2011. Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. Emotion 11, 468–480.

Mauss, I.B., Evers, C., Wilhelm, F.H., Gross, J.J., 2006. How to bite your tongue without blowing your top: implicit evaluation of emotion regulation predicts affective responding to anger provocation. Pers. Soc. Psychol. Bull. 32, 589–602.

McIntosh, A.R., 2004. Contexts and catalysts a resolution of the localization and integration of function in the brain. Neuroinformatics 2, 175–182.

Meyer, M.L., Berkman, E.T., Karremans, J.C., Lieberman, M.D., 2011. Incidental regulation of attraction: the neural basis of the derogation of attractive alternatives in romantic relationships. Cogn. Emot. 25, 490–505.

Noesselt, T., Shah, N., Jäncke, L., 2003. Top-down and bottom-up modulation of language related areas — an fMRI study. BMC Neurosci. 4, 13.

Ochsner, K.N., Gross, J.J., 2005. The cognitive control of emotion. Trends Cogn. Sci. 9 (5), 242–249.

Ochsner, K.N., Silvers, J.A., Buhle, J.T., 2012. Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. Ann. N. Y. Acad. Sci. 1251, E1–E24.

Payer, D.E., Baicy, K., Lieberman, M.D., London, E.D., 2012. Overlapping neural substrates between intentional and incidental down-regulation of negative emotions. Emotion 12, 229–235.

Penny, W.D., 2012. Comparing dynamic causal models using AIC, BIC and free energy. Neuroimage 59, 319–330.

Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. Neuroimage 22, 1157–1172.

Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., Leff, A.P., 2010. Comparing families of dynamic causal models. PLoS Comput. Biol. 6, e1000709.

Pessoa, L., Adolphs, R., 2010. Emotion Processing and the Amygdala: From a 'Low Road' to "Many Roads" of Evaluating Biological Significance. Nat. Rev. Neurosci. 11 (11), 773–783.

Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62, 816–847.

Quirk, G.J., Likhtik, E., Pelletier, J.G., Pare, D., 2003. Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. J. Neurosci. 23, 8800–8807.

Richardson, F.M., Seghier, M.L., Leff, A.P., Thomas, M.S.C., Price, C.J., 2011. Multiple routes from occipital to temporal cortices during reading. J. Neurosci. 31, 8239–8247.

Ridgway, G.R., Omar, R., Ourselin, S., Hill, D.L.G., Warren, J.D., Fox, N.C., 2009. Issues with threshold masking in voxel-based morphometry of atrophied brains. Neuroimage 44, 99–111.

Rowe, J.B., Hughes, L.E., Barker, R.A., Owen, A.M., 2010. Dynamic causal modelling of effective connectivity from fMRI: are results reproducible and sensitive to Parkinson's disease and its treatment? Neuroimage 52, 1015–1026.

Schuyler, B., Ollinger, J.M., Oakes, T.R., Johnstone, T., Davidson, R.J., 2010. Dynamic causal modeling applied to fMRI data shows high reliability. Neuroimage 49, 603–611.

Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E.M., Daunizeau, J., Friston, K.J., 2010. Ten simple rules for dynamic causal modeling. Neuroimage 49, 3099–3109.

Tabibnia, G., Lieberman, M.D., Craske, M.G., 2008. The lasting effect of words on feelings: words may facilitate exposure effects to threatening images. Emotion 8, 307–317.

Volokhov, R.N., Demaree, H.A., 2010. Spontaneous emotion regulation to positive and negative stimuli. Brain Cogn. 73, 1–6.