Neural activity during affect labeling predicts expressive writing effects on well-being: GLM and SVM approaches

Negar Memarian,[1] Jared B. Torre,[1] Kate E. Haltom,[1] Annette L. Stanton,[1,2,3] and Matthew D. Lieberman[1,2]

[1]Department of Psychology, [2]Department of Psychiatry/Biobehavioral Sciences, University of California Los Angeles (UCLA), Los Angeles, CA 90095, USA, and [3]Jonsson Comprehensive Cancer Center, University of California Los Angeles (UCLA), Los Angeles, CA 90095, USA

Correspondence should be addressed to Matthew D. Lieberman, Departments of Psychology and Psychiatry and Biobehavioral Sciences, Franz Hall, University of California Los Angeles, Los Angeles, CA 90095-1563, USA. E-mail: lieber@ucla.edu

Abstract

Affect labeling (putting feelings into words) is a form of incidental emotion regulation that could underpin some benefits of expressive writing (i.e. writing about negative experiences). Here, we show that neural responses during affect labeling predicted changes in psychological and physical well-being outcome measures 3 months later. Furthermore, neural activity of specific frontal regions and amygdala predicted those outcomes as a function of expressive writing. Using supervised learning (support vector machines regression), improvements in four measures of psychological and physical health (physical symptoms, depression, anxiety and life satisfaction) after an expressive writing intervention were predicted with an average of 0.85% prediction error [root mean square error (RMSE) %]. The predictions were significantly more accurate with machine learning than with the conventional generalized linear model method (average RMSE: 1.3%). Consistent with affect labeling research, right ventrolateral prefrontal cortex (RVLPFC) and amygdalae were top predictors of improvement in the four outcomes. Moreover, RVLPFC and left amygdala predicted benefits due to expressive writing in satisfaction with life and depression outcome measures, respectively. This study demonstrates the substantial merit of supervised machine learning for real-world outcome prediction in social and affective neuroscience.

Key words: affect labeling; expressive writing; supervised learning; support vector machines; functional magnetic resonance imaging (fMRI)

Introduction

Writing about one's deepest feelings and thoughts regarding stressful experiences or trauma can bolster one's psychological and physical health (Pennebaker, 1993). Three decades of research on written expressive disclosure have documented (Pennebaker and Beall, 1986; Frattaroli, 2006) that two to four expressive writing sessions, focused on the source of one's distress, can produce demonstrable psychological and physical health benefits over the subsequent months. Health benefits of expressive writing include improvements in blood pressure (McGuire *et al.*, 2005), chronic pain (Broderick *et al.*, 2005), symptoms and medical appointments for cancer-related morbidities (Stanton *et al.*, 2002), lung function (Smyth *et al.*, 1999), liver function (Francis and Pennebaker, 1992) and immune function (Booth

et al., 1997). Improvements in mental health indicators also are demonstrated (Park and Blumberg, 2002; Hemenover, 2003). A meta-analysis confirms these effects are real, though small and heterogeneous (Frattaroli, 2006). In this article, we aimed to use neural responses during affect labeling (AL) processed using machine learning algorithms to predict improvement in four outcomes in general and associated with expressive writing.

No definitive explanations for the beneficial effects of expressive disclosure through writing have emerged. Plausible mechanisms include release from inhibition, altered cognitive appraisal/discovery of meaning, self-affirmation, narrative creation and repeated exposure; the relevant data are mixed (Pennebaker, 1993; Baikie and Wilhelm, 2005; Niles *et al.*, 2016). At this point, the literature on expressive writing and emotional disclosure can be summarized as providing clear evidence that putting feelings into words has psychological and physical health benefits but not yet establishing mechanisms through which these effects occur. Specifying these mechanisms would allow therapists and researchers to tailor the process and content of expressive writing, and other forms of emotional disclosure, to maximize its benefits.

Among the putative mechanisms of expressive writing, emotion regulation through AL has received minimal consideration, excepting research on positive and negative emotion word counts in essays as predictors of effects. The lack of investigation of emotion regulation as a mechanism of expressive writing effect makes sense given that most forms of emotion regulation are intentional and expressive writing does not feel like intentional emotion regulation. Benefits of expressive writing appear to be incidental from the perspective of the writer, rather than outcomes explicitly sought. Research on AL (i.e. the act of describing in words the emotional aspects of a stimulus or one's own emotional reaction to it) reveals that it serves as a form of incidental emotion regulation consistent with the processes likely to occur during expressive writing. Affect labeling produces many of the same effects as other forms of emotion regulation but appears to do so without intention or awareness (J. Torre and M.D. Lieberman, under review). Participants instructed to engage in AL while being exposed to negative images report less negative affect, despite endorsing a theory that AL will increase their negative affect (Lieberman *et al.*, 2011). From a mechanistic perspective, AL reduces negative affect because (i) negative affect often involves amygdala activation; (ii) AL reliably recruits right ventrolateral prefrontal cortex (RVLPFC) and (iii) RVLPFC activation reliably diminishes amygdala activity. In essence, AL reduces negative affect for the same reason that reappraisal does—they both engage prefrontal regions capable of downregulating one of the sources of negative affect (Lieberman *et al.*, 2011; Burklund *et al.*, 2014). Affect labeling is also associated with long-term reductions in electrodermal responses to negative images and feared stimuli (Tabibnia *et al.*, 2008; Kircanski *et al.*, 2012; Niles *et al.*, 2016).

Neuroimaging research using functional magnetic resonance imaging (fMRI) suggests that AL produces benefit via increased activity within RVLPFC and corresponding decreases in amygdala activity (Hariri *et al.*, 2000; Lieberman *et al.*, 2007; Payer *et al.*, 2012; Torrisi *et al.*, 2013; Burklund *et al.*, 2014). These studies have demonstrated that the act of naming the emotionally evocative aspect of an image or labeling one's own reaction to the image produces increased RVLPFC activity and decreased amygdala activity relative to a non-emotional form of labeling. Negative functional connectivity and negative correlations have also been observed between RVLPFC and amygdala responses during AL (Hariri *et al.*, 2000; Lieberman *et al.*, 2005, 2007; Foland

et al., 2008; Payer *et al.*, 2012). Consistent with the idea that RVLPFC activity is producing the relative amygdala activity reductions, dynamic causal modeling demonstrated that the best account of AL effects involved inputs to RVLPFC leading to a dampened response in the amgydala (Torrisi *et al.*, 2013).

Based on these findings, we hypothesized that the implicit emotion regulation that occurs during AL may predict: (i) improvements in measures of psychological and physical well-being in general and (ii) psychological and physical health benefits of expressive writing in particular, as it is analogous to implicit emotion regulation processes that might occur during expressive writing. We think that the same mechanism (increased response of the RVLPFC decreased amygdala response) underlies both AL and expressive writing, at least in part. If this is the case, then neural responses during AL should predict the benefits of expressive writing. Therefore, individuals who produce more robust RVLPFC and dampened amygdala activity during AL are hypothesized to benefit more from expressive writing than those with less robust RVLPFC and heightened amygdala responses during AL. Outcome measures investigated in this study are changes in self-reported anxiety, depressive symptoms and negative physical symptoms between the start and end of a 3 month period. Furthermore, these outcome changes were looked at in relation to expressive writing (controlling for self-reports prior to expressive writing). In this study, we also added life satisfaction as an outcome to explore positive adjustment.

We took two approaches to examining the predictive relationship between AL-related neural responses and the outcomes. Both procedures used the signal from a set of regions of interest (ROI) in the brain associated with AL. We identified ROIs that were more active during AL than during gender labeling (GL) of emotionally expressive faces. Parameter estimates of effect were extracted for each ROI and used to predict outcomes 3 months after expressive writing (controlling for initial values on the outcomes). We first used generalized linear modeling (GLM) with parameter estimates of activity from the ROIs serving as predictors of the four outcomes.

Our second analytic procedure is more novel within social, affective and clinical neuroscience. Here, we used support vector machine (SVM) learning algorithms to predict outcomes from the AL ROIs. The GLM approach is the typical method used when attempting to predict real-world outcomes from neural activity (Berkman and Lieberman, 2009; Berns and Moore, 2012; Falk *et al.*, 2012). If SVM systematically provides greater predictive power than GLM, it warrants use in future similar research. SVM has been previously applied to fMRI data to separate different brain functional activation patterns in different task states such as motor tasks, e.g. hand movement (Zeng *et al.*, 2008) and finger tapping (Wang *et al.*, 2007; Wang, 2009), and cognitive tasks, e.g. picture-sentence matching (Wang *et al.*, 2003), reading in different languages (Ji *et al.*, 2004), watching visual *vs* written information (Ramasangu and Sinha, 2014) and subjective experience during virtual reality (Grazia *et al.*, 2008). Our findings in this article suggest that the SVM approach may have significant implications in clinical, social and affective neuroscience as well.

Materials and methods

To investigate whether the neural correlates of AL predict (i) improvements in measures of psychological and physical well-being in general and (ii) psychological and physical health benefits of expressive writing in particular, we assessed individual differences in neurocognitive responses during AL using a variant of a published task (Lieberman *et al.*, 2007). Outcomes were
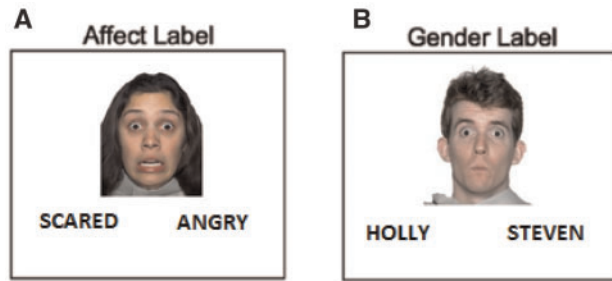
Fig. 1. Task completed by the subjects in the fMRI scanner. The paradigm for this study included two conditions: (A) affect label, (B) gender label. During the affect label condition (A), participants were shown a series of human face images showing various emotions. For each image, the participant was asked to choose one of two given label options presented on the screen that best described the facial emotion in the image (e.g. angry, fearful and happy). A non-emotional GL task (B) was also included as a control condition indexing simple cognitive responses. Functional peaks were determined based on group-level analysis of data, using the affect label–gender label contrast, and 5% significance level (FDR $P < 0.05$).

assessed prior to the randomly assigned writing procedure and at a 3 month follow-up. Through supervised machine learning, we related the neural responses during the AL tasks to level of improvement on the outcomes.

Subjects

One hundred and thirteen ($N = 113$) participants were recruited and divided into two groups of expressive writing and control. Two of the participants in the expressive writing group had corrupt fMRI scans and were excluded from the study, resulting in 58 participants (27 women; age $= 21 \pm 3.3$). The control group had 53 participants (26 women; age $= 21 \pm 2.5$). Because the protocol involved fMRI scanning, subjects were required to be scanner eligible (i.e. metal-free, right handed, not claustrophobic, not pregnant).

Protocol

Recruited from undergraduate course announcements and flyers, interested participants contacted a study coordinator for telephone screening [for methodological details, see Niles *et al.*(2016)]. Eligible participants were between 18 and 40 years old, fluent in English, had no psychological disorder or serious disease and had experienced a stressful event within the prior 5 years that they rated as 5 or greater in stressfulness on a seven-point Likert scale (1 = not at all stressful; 7 = extremely stressful).

Participants attended a baseline session during which they provided written informed consent and underwent fMRI scanning (t1). They also performed AL via an established paradigm in the fMRI scanner. Four conditions (affect label, gender label, observe and shape match) were administered over two runs, with one block per condition per run. Each block was 40 s long and comprised of eight trials (5 s each). There was a 3 s prompt before each block indicating what was coming next and a 12 s crosshair fixation/rest between the conditions. Only the data from the affect and gender label conditions were used for this study. Figure 1 shows an overview of the task. During the AL condition, participants viewed a series of human face images showing various emotions. For each image, the participant was asked to choose one of two label options presented on the screen that best described the facial emotion (e.g. angry, sad

and happy). A non-emotional GL task served as a control condition indexing simple cognitive responses.

Participants attended four 20 min writing sessions, scheduled at least 3 days apart and occurring within 8 weeks. At each of the four writing sessions, participants listened to an audio-recording of the instructions and completed the task in a private laboratory room. Participants placed their completed essays in an envelope and returned it to the experimenter. The expressive writing group participants were asked to write about past negative experiences (standard expressive writing protocol), while the other one-half wrote about non-emotional topics instead (control subjects). For the expressive writing group, the task involved describing their deepest thoughts and feelings regarding the 'most stressful or traumatic experience during the past 5 years'. Three months after the final writing session, participants completed the follow-up questionnaires via the Internet (t2).

Outcome measures

Change scores on the outcomes represent improvement in self-reported physical and psychological health indicators from the time of fMRI scanning (pre-intervention = t1) to 3 months after that (t2).

Beck Depression Inventory. The 21-item Beck Depression Inventory (BDI) (Beck and Steer, 1984) assesses symptoms of depression such as hopelessness, feelings of guilt and weight changes. Participants rated the severity of depressive symptoms from 0 to 4 in the past week. BDI items were summed; higher scores indicate more severe depressive symptoms. Level of depressive symptoms based on the BDI score is as follows: 0–13: minimal depression; 14–19: mild depression; 20–28: moderate depression; 29–63: severe depression; >63: more severe depressive symptoms. Improvement in this outcome measure was calculated as $\Delta BDI_{(t1 - t2)}$.

Satisfaction with Life Scale. The Satisfaction with Life Scale (SWLS) is a five-item instrument designed to measure global cognitive judgments of satisfaction with one's life (Diener *et al.*, 1985). Items were summed such that higher scores indicate higher life satisfaction. SWLS was the only outcome measure scored in the direction of higher scores being better. Level of satisfaction with life based on the SWLS score sum is as follows: 30–35: highly satisfied; 25–29: like their life; 20–24: generally satisfied; 15–19: slightly dissatisfied; 10–14: substantially dissatisfied; 5–9: extremely dissatisfied (Diener *et al.*, 1985). Please note that SWLS is the only outcome measure, which higher score indicates better well-being. Therefore, change in SWLS score was calculated in reverse time order (i.e. $SWLS_{t2} - SWLS_{t1}$), which indicated an improvement when positive.

Pennebaker Inventory of Limbic Languidness. The 54-item Pennebaker Inventory of Limbic Languidness (PILL) (Pennebaker, 1982) assesses common physical symptoms. Participants indicate how often they have experienced each symptom on a five-point Likert scale (1 = never or almost never, 2 = less than 3 or 4 times per year, 3 = every month or so, 4 = every week or so, 5 = more than once every week). Higher total scores indicate more frequent physical symptoms, with classification as follows: 0–21 = below normal range; 22–66 = well within normal range; 67–84 = slightly above average; within normal range; 85 or above = top 25% (Pennebaker, 1982). Improvement in this outcome measure was calculated as $\Delta PILL_{(t1 - t2)}$.
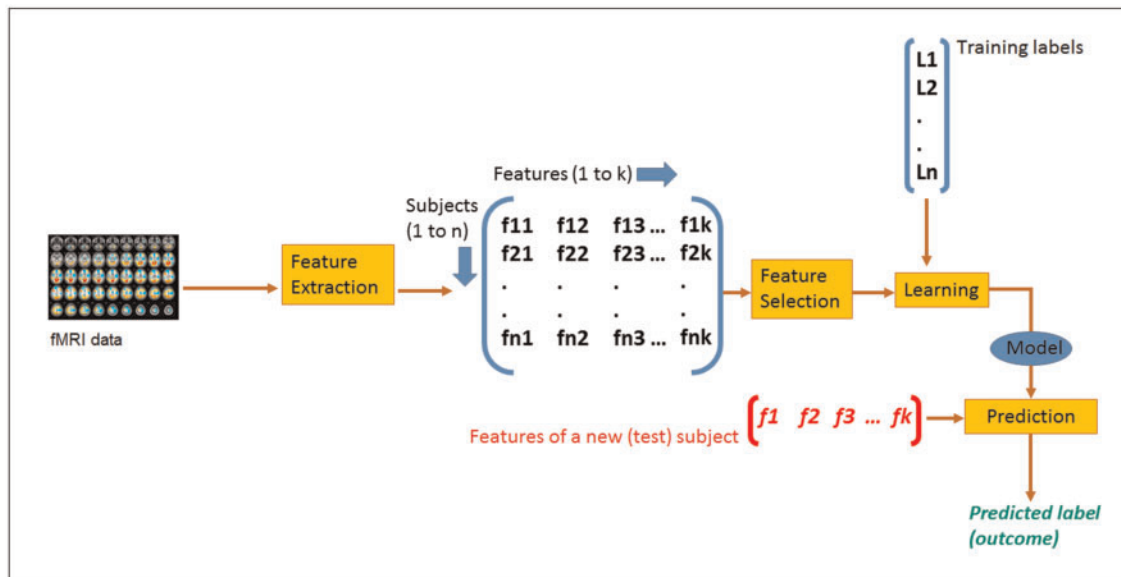
Fig. 2. System overview of the machine learning-based prediction protocol.

Brief Symptom Inventory, anxiety dimension. The 53-item Brief Symptom Inventory (BSI) was developed from its longer parent instrument, the SCL-90-R, to assess psychological symptoms. The anxiety dimension (ANX) of the BSI subsumes a set of symptoms usually associated with clinically manifest anxiety. Restlessness, nervousness and tension are all indicative of anxiety, as are experiences reflecting free-floating anxiety and panic (Derogatis and Melisaratos, 1983). The higher the value of BSI_ANX, the more anxious the subject is. Improvement in this outcome measure was calculated as $\Delta BSI_{(t1 - t2)}$.

Data acquisition

Functional MR images were acquired using a Siemens Allegra 3 Tesla MRI scanner. A 2D spin-echo image (TR = 4000 ms, TE = 40 ms, matrix size 256 × 256, 4 mm thick, 1 mm gap) was acquired in the sagittal plane to allow prescription of the slices to be obtained in the remaining scans. For each participant, a high resolution structural T2-weighted echo-planar imaging volume (spin-echo, TR = 4000 ms, TE 54 ms, matrix size 128 × 128, FOV = 20 cm, 36 slices, 1.56 mm in-plane resolution, 3 mm thick) was acquired. Foam padding was used to limit head movement.

SVM data analysis

Figure 2 shows the overview of the data analysis protocol. The premise is to train a computer algorithm by giving it features (pre-outcome data) along with corresponding responses (outcome scores) for a group of individuals. The computer algorithm tries to find patterns between features that are exclusively associated with the given response. Usually such patterns are not obvious from manual inspection of data. After the algorithm learns, it is able to predict a new subject's outcome based on his/her pre-intervention features only.

Feature extraction. Twenty-two fMRI features were extracted for each subject as shown in Table 1. These features represent the percent signal change in 6 mm radius vicinity of functional peak coordinates (except amygdala, where due to its small size,

anatomical ROI was used). Functional peaks were determined based on group-level analysis of data, using an AL–GL contrast and a false discovery rate (FDR) of $P < 0.05$. For each feature, participants with missing values were replaced with the average of that feature among all subjects. To find those brain regions with differential predictiveness as a function of condition (expressive writing/control), the fMRI predictors were multiplied by a dummy coded condition effect, i.e. +1 if they belonged to subjects from the expressive writing group or −1 if they belonged to the control group subjects.

Feature selection. An algorithm named minimal-redundancy-maximal-relevance (mRMR) was used to select the 10 most informative features (predictors). This algorithm was chosen because of its advantages in terms of both feature selection complexity and feature classification accuracy (Peng *et al.*, 2005). In this method, relevant features and redundant features are considered simultaneously, i.e. it seeks to maximize the relevance of a feature set for a specific class and minimize the redundancy of all features in the feature set. Relevance is defined by the average value of all mutual information (MI) values between the individual feature and the specific class. Redundancy is the average value of all MI values between the individual feature and every other feature in the set. Furthermore, the mRMR algorithm is suitable for unprocessed data, where the features selected in this way will have more or less correlation with each other. This is because mRMR does not intend to select features that are independent of each other. Instead, at each step, it tries to select a feature that minimizes the redundancy and maximizes the relevance (Peng *et al.*, 2005).mRMR has been tested on a number of medical datasets, namely arrhythmia (Guvenir *et al.*, 1997), cancer cell lines (Ross *et al.*, 2000; Scherf *et al.*, 2000) and lymphoma (Alizadeh *et al.*, 2000) and produced promising improvement on classification accuracy. These studies involved datasets with a large number of features, combination of discrete and continuous data and different classifiers. Because the present dataset had similar attributes, we chose to use the mRMR feature selection method.

Table 1. fMRI features representing regions of peak activity when participants performed an AL task

	Feature (Brodmann area)	Peak coordinate	Cluster size	t-value
1	Right inferior frontal gyrus pars triangularis (BA 45)	(54 27 18)	1433	8.32
2	Left middle frontal gyrus/frontal inferior triangularis (BA 45)	(−54 24 27)	2151	7.25
3	Left inferior frontal gyrus pars triangularis-a (BA 45)	(−51 33 12)	2151	8.54
4	Left inferior frontal gyrus pars triangularis-b (BA 45)	(−51 30 0)	2151	6.39
5	Right frontal inferior orbitalis (BA 47)	(48 30 −3)	1433	5.55
6	Left inferior frontal gyrus pars opercularis (BA 44)	(−48 12 18)	2151	6.30
7	Left frontal pole/frontal medial (BA 10)	(−36 57 9)	2151	3.86
8	Left insular cortex	(−30 24 −3)	2151	5.85
9	Right insular cortex	(30 24 −3)	1433	6.40
10	Right frontal pole/superior frontal (BA 10)	(15 57 18)	1136	4.03
11	Right paracingulate gyrus (BA 32)	(15 33 27)	1136	3.71
12	Left extra nuclear (BA 11)	(−15 12 3)	2151	3.21
13	Left paracingulate gyrus (BA 32)	(−12 27 27)	1136	3.60
14	Right cingulate gyrus anterior division (BA 32)	(12 27 21)	1136	3.52
15	Right frontal pole/frontal superior medial (BA 10)	(9 57 24)	1136	3.85
16	Right paracingulate gyrus/medial frontal gyrus (BA 32)	(9 18 48)	1136	6.98
17	Right ventral striatum	(9 3 −3)	2151	3.02
18	Right superior frontal gyrus/medial frontal gyrus (BA 9)	(6 48 39)	1136	4.30
19	Right superior frontal gyrus/frontal superior medial (BA 8)	(3 33 48)	1136	5.58
20	Left superior frontal gyrus/superior motor (BA 8)	(0 18 51)	1136	6.74
21	Right amygdala	N/A (anatomical ROI)	N/A	N/A
22	Left amygdala	N/A (anatomical ROI)	N/A	N/A

Peaks were determined based on group-level analysis of data, using an AL–GL contrast, and 5% significance level (FDR $P < 0.05$) in the scanner. Peak coordinate is the $X\,Y\,Z$ coordinate of the peak voxel in MNI space. Cluster size is the size of the cluster in contiguous voxels in which the peak belongs.

Learning and classification (LOOCV). Training (supervised learning) and prediction were carried out using a SVM regression model with radial basis function kernel (SVM-rbf) (Soman *et al.*, 2009). We chose SVM regression because it allowed us to perform predictive modeling with continuous outcome measures (as opposed to binary or multiple class classification which requires grouping subjects into a discrete number of classes based on arbitrary thresholds). Also, SVM is robust to a large number of variables and small samples, it can learn both simple and highly complex classification models, and it employs sophisticated mathematical principles to avoid overfitting. Previous work that used supervised learning approaches for outcome prediction in healthcare showed that SVM-based methods are the most promising based on their high classification accuracy using a small number of features. Machine learning models that achieve similar accuracy by operating on a selected set of features are preferred in investigative research over machine learning models that are saturated with input features (Memarian *et al.*, 2015). A leave-one-out cross validation (LOOCV) scheme was used to train and test the SVM model. In other words, the prediction error was computed as the average of 113 iterations (corresponding to 113 subjects), where at each iteration one participant was left out as the test subject and the remaining 57 participants were used for training the SVM regression model. This procedure was repeated 113 times until every participant was used as a test subject. It should be noted that feature selection was performed only on the training data, i.e. feature selection repeated in each leave-one-out iteration. A more detailed description of the feature selection and classification algorithms is available (Memarian *et al.*, 2015).

GLM data analysis

To compare the efficacy of SVM-based prediction against conventional univariate approaches, we also used the GLM with parameter estimates of activity from the ROIs serving as predictors and different measures of well-being serving as the outcome. GLM is the most commonly used univariate technique in the social and cognitive neuroscience literature. It has become the core tool for fMRI data analysis after its introduction into the neuroimaging community by Friston *et al.* (1994). From the perspective of multiple regression analysis, the GLM aims to 'explain' or 'predict' the variation of one dependent variable (and hence the title 'uni'variate) in terms of a linear combination (weighted sum) of several reference functions. In this study, the dependent variable is the improvement in measures of well-being and the reference functions (also known as the predictors or regressors) corresponds to the mean fMRI percent signal change in various brain regions. Similar to learning and classification with SVM-rbf, the top selected features were used as predictors and a LOOCV scheme was applied to train and test the GLM.

Results

First, we determined whether and to what extent neural activity during AL associated with changes in participants' measures of psychological and physical well-being over the course of 3 months. Second, we investigated if neural activity during AL could specifically predict benefits of expressive writing on psychological and physical health outcomes at 3 months. This was meant to give us an indication of whether AL processes could provide a mechanism for the observed benefits of expressive writing. Third, we compared the ability of GLM and SVM approaches for making this kind of prediction. GLM is the standard approach used to predict outcomes from the brain (Berkman and Lieberman, 2009; Berns and Moore, 2012; Falk *et al.*, 2012), and we were interested in whether the SVM regression predictive modeling approach would provide more accurate results (less regression error).

Table 2. Pairwise correlation coefficients of the four outcome measures

	rΔBDI	rΔSWLS	rΔPILL	rΔBSI_ANX
rΔBDI				
rΔSWLS	0.32*			
rΔPILL	0.08	0.06		
rΔBSI_ANX	0.26*	−0.04	0.06	

BDI, Beck Depression Index; SWLS, Satisfaction with Life Scale; PILL, Pennebaker Inventory of Limbic Languidness; BSI_ANX, Brief Symptom Inventory_Anxiety. Statistically significant correlations (P < 0.05) are marked with asterisk.

Pairwise correlation coefficients of the four outcomes (raw scores) are shown in Table 2. Except for the BDI and BSI_ANX (rΔBDI, ΔBSI_ANX = 0.26, P < 0.05), and the BDI and SWLS (rΔBDI, ΔSWLS = 0.32, P < 0.05), outcome measures were not significantly correlated and none shared >10% of their variance.

Expressive writing was associated with better outcome as compared with the control group, in all of the four outcome measures. For ΔBDI(t1 − t2), average improvement in the expressive writing group = 4.6, vs average improvement in the control group = 4.3. For ΔPILL(t1 − t2), average improvement in the expressive writing group = 13.5, vs average improvement in the control group = 13. For ΔBSI(t1 − t2), average improvement in the expressive writing group = 3.8, vs average improvement in the control group = 3.3. Finally, for ΔSWLS(t2 − t1), average improvement in the expressive writing group = 5.5, vs average improvement in the control group = 4.8.

The average prediction error (root mean square error: RMSE) of predicting the four outcomes using SVM-rbf regression model is shown in Table 3. In every section of the table, the leftmost data column lists the group of top 10 selected features that resulted in the reported average RMSE (specified in the second data column). The percentage of average prediction error (RMSE %) calculated from $\frac{|Actual\ outcome\ value\ -\ predicted\ outcome\ value|}{Actual\ outcome\ value \times 100}$ is presented in column three. Data column four shows the correlation sign for each of those features and the outcome measure. A positive sign of the correlation between the predictor and outcome of interest indicates that higher activity in a specific brain region correlated with improvement in the outcome measure (i.e. an increase in satisfaction with life and a decline in depressive symptoms, anxiety and physical symptoms) from t1 to t2.

In a similar fashion, the average prediction error (RMSE) and the percentage of average prediction error (RMSE %) of predicting the four outcomes were computed using the GLM model. With the same selected features, the GLM model resulted in prediction RMSE = 7.04 (RMSE % = 1.39) for the BDI outcome; prediction RMSE = 18.12 (RMSE % = 1.27) for the PILL outcome; prediction RMSE = 6.08 (RMSE % = 1.18) for the SWLS outcome and prediction RMSE = 4.72 (RMSE % = 1.34) for the BSI_ANX outcome.fMRI predictors that showed significant differential predictiveness as a function of condition (expressive writing or control) are shown in Table 4. Specifically, greater activity in right inferior frontal gyrus (IFG), a region associated with emotion regulation, predicts greater improvement in SWLS for those who engaged in expressive writing, relative to controls. In controls, activity in amygdala and subgenual cingulate, regions involved in affect generation, predict worse outcomes in depression and anxiety, respectively, for those who engaged in expressive writing, relative to controls. Using only the three predictors shown in Table 4 and Figure 3 [i.e. left amygdala, right

inferior frontal gyrus pars triangularis (BA 45) and right ventral striatum], the percentage of average prediction error of SVM-rbf for each outcome measure was 0.90% (PILL), 0.90% (SWLS), 0.87% (BDI) and 0.87% (BSI_ANX).

The SVM approach, using 10 ROIs (top 10 features as determined by the mRMR feature selection stage) per analysis, was able to predict changes in psychological and physical health outcomes with significantly lower regression error compared with GLM. A Mann–Whitney U-test between regression RMSE samples (113 samples for 113 subjects) was used to compare the performance of SVM-rbf regression predictive modeling with the GLM method and compute the Z statistic. The reported RMSE values are averages calculated based on a LOOCV scheme (Figure 4). For the BDI, average RMSE for the SVM-rbf regression model (4.29 or 0.85%) was less than for GLM (7.04 or 1.39%), Z = −3.7418, P < 0.001. For the SWLS, average RMSE for the SVM-rbf regression model (4.31 or 0.84%) was less than for GLM (6.08 or 1.18%), Z = −3.7433, P < 0.001. For the PILL, average RMSE for the SVM-rbf regression model (12.18 or 0.85%) was less than for GLM (18.12 or 1.27%), Z = −3.7418, P < 0.001. For BSI_ANX, average RMSE for the SVM-rbf regression model (3.05 or 0.87%) was less than for GLM (4.72 or 1.34%), Z = −3.7433, P < 0.001. The mean prediction error across the four outcomes was 5.95 (0.853%) for SVM and 8.99 (1.295%) for GLM.

Figure 5a–d show the GLM and SVM-rbf percentage of average prediction error trend, which is the RMSE % of models based on inclusion of each feature and its higher ranking features (as listed in Table 3) for each outcome measure. It can be seen that the SVM-rbf regression consistently yielded smaller regression error than GLM in all of the four plots.

Finally, the combination of feature selection and the SVM approach also allows us to determine which brain regions contribute most reliably across the four outcomes. For example, Figure 6 shows the contribution of the neural predictors as selected by the feature selection algorithm in predicting the four outcomes. The + (positive) sign represents a significant positive correlation between the feature and outcome, whereas the − (negative) sign represents a significant negative feature-outcome correlation. Again, it is evident that RVLPFC (BA 45) and amygdalae were among the primary predictors for each outcome. In particular, despite the four outcome measures being only modestly intercorrelated, RVLPFC was the only ROI that was a positive predictor of improvement on all four outcomes.

Discussion

There were two major goals of this investigation. First, we wanted to examine whether neural activity during an implicit emotion regulation task (AL) predicts long-term consequences (i.e. 3 months) of expressive writing for psychological and physical well-being. Second, we wanted to determine whether SVM prediction methods would be more successful than the traditional GLM approach to 'brain-as-predictor' (Berkman and Falk, 2013).

When looking at the relationship between expressive writing and well-being we did find evidence linking to neural activity during AL for three of the four outcome measures of interest. For subjective well-being, greater right IFG activity during AL predicted greater improvements over time for those who engaged in expressive writing, compared with those who did not. For depression and anxiety, reduced amygdala and subgenual cingulate activity, respectively, predicted greater improvements over time for those who engaged in expressive writing,

Table 3. RMSE of predicting the outcome measures (i.e. BDI, SWLS, PILL and BSI_ANX) using SVMs regression model with a radial basis function (SVM-rbf)

Selected features	Average SVM regression error (RMSE)	Percentage of error (RMSE %)	Correlation with outcome
BDI	4.29	0.85	
L middle frontal gyrus (BA 45)			+
L amygdala			−
R frontal pole/superior frontal (BA 10)			−
R inferior frontal gyrus pars triangularis (BA 45)			+
L paracingulate gyrus (BA 32)			−
L frontal pole/frontal medial (BA 10)			−
R superior frontal gyrus/medial frontal gyrus (BA 9)			−
R amygdala			+
L inferior frontal gyrus pars triangularis-a (BA 45)			+
R frontal inferior orbitalis (BA 47)			+
SWLS	4.31	0.84	
L superior frontal gyrus/superior motor (BA 8)			−
R frontal pole/frontal superior medial (BA 10)			−
L paracingulate gyrus (BA 32)			−
L inferior frontal gyrus pars triangularis-b (BA 45)			−
L amygdala			−
R superior frontal gyrus/medial frontal gyrus (BA 9)			+
L frontal pole (BA 10)			−
R inferior frontal gyrus pars triangularis (BA 45)			+
R frontal pole/superior frontal (BA 10)			−
R frontal inferior orbitalis (BA 47)			+
PILL	12.18	0.85	
L middle frontal gyrus (BA 45)			+
L amygdala			+
R inferior frontal gyrus pars triangularis (BA 45)			+
R frontal pole/superior frontal (BA 10)			−
R frontal inferior orbitalis (BA 47)			−
L frontal pole/frontal medial (BA 10)			−
R insular cortex (BA 47)			−
L inferior frontal gyrus pars triangularis-a (BA 45)			+
L superior frontal gyrus/superior motor (BA 8)			+
R amygdala			+
BSI_ANX	3.05	0.87	
L superior frontal gyrus/superior motor (BA 8)			+
R frontal pole/frontal superior medial (BA 10)			−
L amygdala			+
R superior frontal gyrus/medial frontal gyrus (BA 9)			−
R inferior frontal gyrus pars triangularis (BA 45)			+
L middle frontal gyrus (BA 45)			+
R amygdala			+
L inferior frontal gyrus pars triangularis-a (BA 45)			+
R ventral striatum			−
R insular cortex (BA 47)			+

The leftmost column lists the top 10 features that resulted in the reported RMSE (specified in the second column). The percentage of average prediction error (RMSE %) calculated from $\frac{|Actual\ outcome\ value\ -\ predicted\ outcome\ value|}{Actual\ outcome\ value \times 100}$ is presented in column three. The fourth column shows the correlation sign for each of those features and the outcome measure. A positive sign of correlation between the predictor and outcome indicates that higher activity in a specific brain region was correlated with improvement in the outcome measure from baseline to 3 month follow-up. L, left; R, right.

compared with those who did not. Differences in physical symptoms for the expressive writers compared with controls was not predicted by neural activity during AL in these analyses.

Given that right IFG and amygdala are the two regions reliably activated, inversely, during numerous AL studies, it is striking that these were the two of the three regions that also predicted effects of expressive writing on well-being 3 months later. The results are consistent with the hypothesis that those who are able to recruit right IFG more effectively when AL are

those who are most likely to benefit from an expressive writing intervention. It suggests further investigations into implicit emotion regulation as the basis for expressive writing benefits are warranted. In future work, to further examine the contribution of AL to expressive writing, it would be interesting to analyze the writing of the participants and see if the use of labels predicts the outcome of expressive writing.

In addition to this primary objective, we also examined neural activity during AL that predicted improvements in well-being over the 3 months following scanning, regardless of

expressive writing condition. Here again, right IFG was the only region that was consistently associated with improvements across all four measures of psychological and physical well-being. Although our results show a clear link between right IFG responses during AL and subsequent well-being, it is unclear whether the measured IFG response reflects somewhat temporary state or a true dispositional state. If the latter was true, then our results would suggest that those with high IFG responses would continually show improvement in well-being. We consider this rather unlikely. Instead, it is more likely that the IFG response represents a tendency that might last weeks or months, long enough to influence the follow-up well-being measures, but that is also likely to fluctuate over longer periods of time.
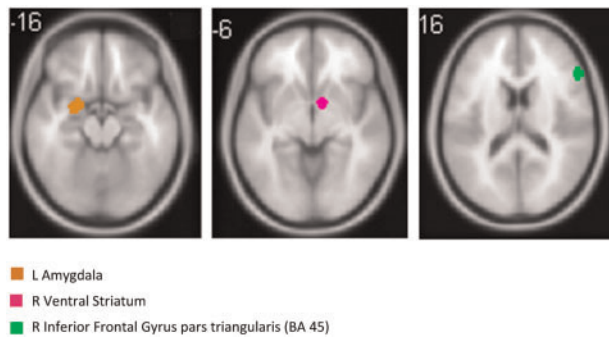

- ■ L Amygdala
- ■ R Ventral Striatum
- ■ R Inferior Frontal Gyrus pars triangularis (BA 45)

Fig. 3. RVLPFC (shown in green) and left amygdala (shown in brown) were top predictors of the four outcome measures of psychological and physical health, i.e. BDI, SWLS, PILL and BSI_ANX (axial view). Moreover, right ventral striatum (shown in purple), as well as right interior frontal gyrus pars triangularis, and left amygdala showed significant differential predictiveness as a function of condition (expressive writing or control).

Our second major goal was to investigate whether neural data analyzed with SVM produces more accurate predictions of well-being than the same neural data analyzed using a traditional GLM approach. SVM-based outcome prediction error was in all cases significantly lower than GLM-based prediction (Figure 4). With fMRI features alone, SVM achieved average prediction error of 0.87% or lower in all cases. This result highlights the superiority of machine learning-based approaches in making useful inferences from information encompassed in large datasets. GLM tries to separate the outcome responses using a linear regression model but the data are not always linearly separable. Methods such as SVM that part the data with non-linear hyperplanes show superior performance in such cases.

In addition to yielding higher predictive accuracy, the proposed SVM-based method has additional advantages over conventional univariate data analysis methods. GLM tends to be overly sensitive to outliers in the training sample. In contrast, SVM is robust to outliers and because it can use non-linear category boundaries, its prediction success is superior to multiple regression approaches (Soman *et al.*, 2009). Furthermore, GLM relates fMRI data to outcomes using the data from a single fMRI task. The SVM approach allows one to pool the predictive value of multiple tasks (along with other data modalities), into a single robust prediction analysis (Meyer *et al.*, 2003; Durrant *et al.*, 2009). This capability of inferring hidden patterns in 'multimodal' data represents a substantial strength of supervised machine learning-based outcome prediction approaches.

The SVM-based approach is an important complement to other recent popular methods in fMRI studies, such as multivoxel pattern analysis (MVPA), because it allows for the prediction of real-world outcomes. MVPA can predict whether an individual is currently viewing a picture of a house or face while in an MRI scanner but cannot predict anything outside of the scanner. Unlike MVPA, the proposed SVM-based approach is perfectly suited for prediction of outcome measures that are

Table 4. fMRI predictors that showed significant differential predictiveness as a function of condition (expressive writing or control), using a two-tailed test, significance level = 0.05

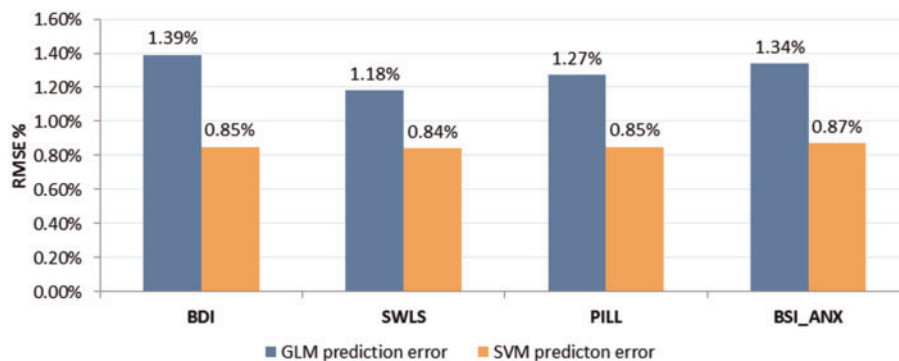| Outcome measure	Feature (fMRI predictor)	Correlation with outcome	t-statistics	P-value
BDI	L amygdala	−	−2.097	0.038
SWLS	R inferior frontal gyrus pars triangularis (BA 45)	+	2.338	0.021
PILL		N/A		
BSI_ANX	R ventral striatum	−	−2.017	0.046

Fig. 4. Performance comparison between the conventional GLM method and the SVM approach. The reported values are percentage of average prediction errors (RMSE %) calculated based on a LOOCV scheme. A Mann–Whitney U-test between prediction error samples (113 samples for 113 subjects) showed significantly lower error (superior performance) for the SVM method compared with the GLM method, in all cases ($P < 0.05$).

collected outside of the scanner. SVMs construct a non-linear parametric model that can predict a real-world outcome for a participant based on multimodal data gathered prior to the intervention. Parameters of this model are learned from a vetted set of data based on outcomes given by other participants who have undergone the same intervention, meaning that the approach employs true supervised learning.

By using SVM-regression for prediction, there is no need to discretize the outcome measure scores into two or more classes and it is possible to predict the level of improvement in continuous outcome measures of interest, like the four studied in this article. This attribute renders this technique much more clinically useful.

A limitation of the current version is that it requires having a priori ROIs. We are currently working on using tools that take a searchlight approach (Kriegeskorte *et al.*, 2006) to perform feature selection across the entire brain, which should provide even more robust effects. In future work, we aim to extend this approach to prediction with multimodal data (e.g. neural, self-report, medical history, genetics, etc.) to take advantage of the full potential of supervised learning for robust outcome prediction.
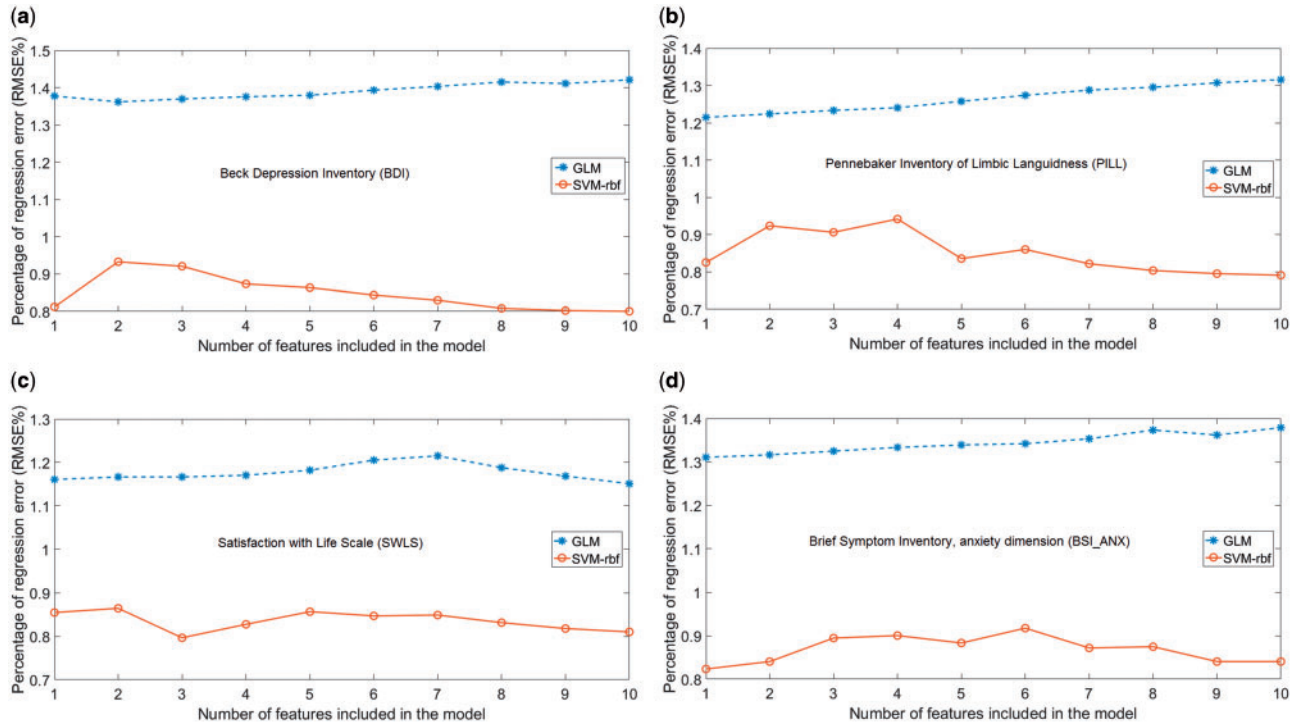

Fig. 5. Percentage of average prediction error (RMSE %) for GLM and SVM-rbf predictive models based on inclusion of each feature and its higher ranking features (as listed in Table 3) for (a) BDI, (b) SWLS, (c) PILL and (d) BSI_ANX.
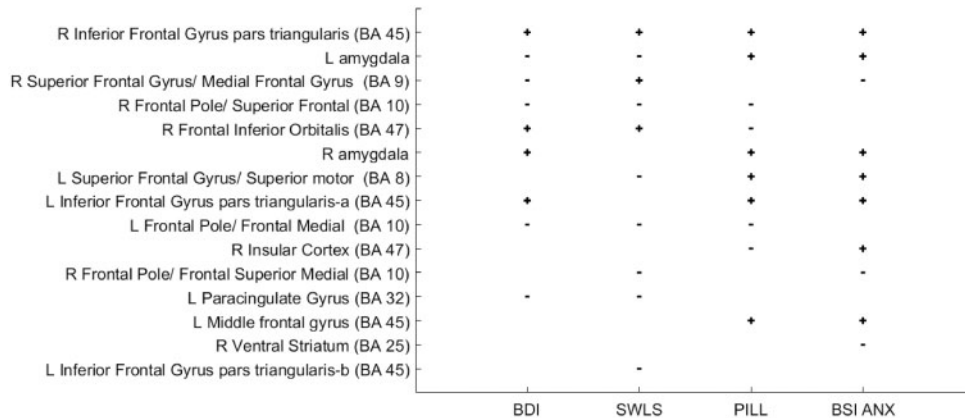

Fig. 6. Contribution of the neural predictors as selected by the feature selection algorithm in predicting the four outcomes (BDI, Beck Depression Inventory score; SWLS, Satisfaction with Life Scale; PILL, Pennebaker Inventory of Limbic Languidness score) BSI_ANX, Brief Symptom Inventory Anxiety score. Plus sign represents positive correlations between the activity of the brain region and improvement on the outcome, whereas a minus sign represents a negative correlation between the feature and the outcome improvement.

Conclusion

In this study, we showed that neural responses during AL can predict changes in psychological and physical health measures, in particular the subsequent benefits of expressive disclosure. In our study of 113 patients, using fMRI data, SVM regression predicted improvements in participants' self-reported physical and psychological health with ≤0.87% prediction error. SVM was significantly more accurate than the classic GLM method for outcome prediction. Greater improvements in the outcomes due to the expressive writing intervention were associated with higher activity in the RVLPFC and lower activity in L amygdala.

References

Alizadeh, A.A., *et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–11.

Baikie, K.A., Wilhelm, K. (2005). Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment*, **11**(5), 338–46.

Beck, A.T., Steer, R.A. (1984). Internal consistencies of the original and revised Beck Depression Inventory. *Journal of Clinical Psychology*, **40**, 1365–7.

Berkman, E.T., Lieberman, M.D. (2009). Using neuroscience to broaden emotion regulation: theoretical and methodological considerations. *Social and Personality Psychology Compass*, **3**(4), 475–93.

Berkman, E.T., Falk, E.B. (2013). Beyond Brain Mapping: Using Neural Measures to Predict Real-World Outcomes. *Current Directions in Psychological Science*, **22**(1), 45–50.

Berns, G.S., Moore, S.E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology*, **22**(1), 154–60.

Booth, R.J., Petrie, K.J., Pennebaker, J.W. (1997). Changes in circulating lymphocyte numbers following emotional disclosure: evidence of buffering? *Stress Medicine*, **13**(1), 23–9.

Broderick, J.E., Junghaenel, D.U., Schwartz, J.E. (2005). Written emotional expression produces health benefits in fibromyalgia patients. *Psychosomatic Medicine*, **67**(2), 326–34.

Burklund, L.J., David Creswell, J., Irwin, M.R., Lieberman, M.D. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology*, **5**, 1–10.

Derogatis, L., Melisaratos, N. (1983). The Brief Symptom Inventory: an introductory report. *Psychological Medicine*, **13**, 595–605.

Diener, E., Emmons, R.A., Larsen, R.J., Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, **49**, 71–5.

Durrant, S., Hardoon, D.R., Brechmann, A., Shawe-Taylor, J., Miranda, E.R., Scheich, H. (2009). GLM and SVM analyses of neural response to tonal and atonal stimuli: new techniques and a comparison. *Connection Science*, **21**(2–3), 161–75.

Falk, E.B., Berkman, E.T., Lieberman, M.D. (2012). From neural responses to population behavior: neural focus group predicts population-level media effects. *Psychological Science*, **23**(5), 439–45.

Foland, L.C., Altshuler, L.L., Bookheimer, S.Y., Eisenberger, N., Townsend, J., Thompson, P.M. (2008). Evidence for deficient modulation of amygdala response by prefrontal cortex in bipolar mania. *Psychiatry Research*, **162**(1), 27–37.

Francis, M.E., Pennebaker, J.W. (1992). Putting stress into words: the impact of writing on physiological, absentee, and self-reported emotional well-being measures. *American Journal of Health Promotion: AJHP*, **6**(4), 280–7.

Frattaroli, J. (2006). Experimental disclosure and its moderators: a meta-analysis. *Psychological Bulletin*, **132**(6), 823–65.

Friston, K.J., Jezzard, P., Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, **1**(2), 153–71.

Grazia, M., Bono, D., Zorzi, M. (2008). Decoding cognitive states from fMRI data using support vector regression. *Psychology Journal*, **6**(2), 189–201.

Guvenir, H.A., Acar, B., Demiroz, G., Cekin, A. (1997). A supervised machine learning algorithm for arrhythmia analysis. *Computers in Cardiology 1997*, Sweden, pp. 433–36, IEEE.

Hariri, A.R., Bookheimer, S.Y., Mazziotta, J.C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport*, **11**(1), 43–8.

Hemenover, S.H. (2003). The good, the bad, and the healthy: impacts of emotional disclosure of trauma on resilient self-concept and psychological distress. *Personality and Social Psychology Bulletin*, **29**(10), 1236–44.

Ji, Y., Liu, H.-B., Wang, X.-K., Tang, Y.-Y. (2004). Cognitive states classification from fMRI data using support vector machines. In: Proceedings of 2004 International Conference on Machine Learning and Cybernetics, 2919–23.

Kircanski, K., Lieberman, M.D., Craske, M.G. (2012). Feelings into words: contributions of language to exposure therapy. *Psychological Science*, **23**(10), 1086–91.

Kriegeskorte, N., Goebel, R., Bandettini, P. (2006). Information-based functional brain mapping. *Proc Natl Acad Sci USA*, **103**(10), 3863–8.

Lieberman, M.D., Eisenberger, N.I., Crockett, M.J., Tom, S.M., Pfeifer, J.H., Way, B.M. (2007). Putting feelings into words. *Psychological Science*, **18**(5), 421–8.

Lieberman, M.D., Hariri, A., 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. *Nature Neuroscience*, **8**(6), 720–2.

Lieberman, M.D., Inagaki, T.K., Tabibnia, G., Crockett, M.J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion*, **11**(3), 468–80.

McGuire, K.M.B., Greenberg, M.A., Gevirtz, R. (2005). Autonomic effects of expressive writing in individuals with elevated blood pressure. *Journal of Health Psychology*, **10**(2), 197–209.

Memarian, N., Kim, S., Dewar, S., Engel, J., Staba, R.J. (2015). Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. *Computers in Biology and Medicine*, **64**, 67–78.

Meyer, D., Leisch, F., Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, **55**(1–2), 169–86.

Niles, A.N., Haltom, K.E.B., Lieberman, M.D., *et al.* (2016). Writing content predicts benefit from written expressive disclosure: evidence for repeated exposure and self-affirmation. *Cognition and Emotion*, **30**(2), 258–74.

Park, C.L., Blumberg, C.J. (2002). Disclosing trauma through writing: testing the meaning-making hypothesis. *Cognitive Therapy and Research*, **26**(5), 597–616.

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**(2), 229–35.

Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(8), 1226–38.

Pennebaker. (1982). *The Psychology of Physical Symptoms*. New York: Springer-Verlag.

Pennebaker. (1993). Putting stress into words: health, linguistic, and therapeutic implications. *Behaviour Research and Therapy*, **31**(6), 539–48.

Pennebaker, Beall. (1986). Confronting a traumatic event: toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, **95**(3), 274–81.

Ramasangu, H., Sinha, N. (2014). Cognitive state classification using transformed fMRI data. In: International Conference on Signal Processing and Communications (SPCOM), 1–5.

Ross, D.T., Scherf, U., Eisen, M.B., *et al.* (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics*, **24**(3), 227–35.

Scherf, U., *et al.* (2000). A gene expression database for the molecular pharmacology of cancer. *Nature Genetics*, **24**(3), 236–44.

Smyth, J.M., Stone, A.A., Hurewitz, A., Kaell, A. (1999). Effects of writing about stressful experiences on symptom reduction in patients with asthma or rheumatoid arthritis: a randomized trial. *JAMA*, **281**(14), 1304–9.

Soman, K., Loganathan, R., Ajay, V. (2009). *Machine Learning with SVM and Other Kernel Methods*. New Delhi, India: PHI Learning Pvt. Ltd.

Stanton, A.L., Danoff-Burg, S., Sworowski, L.A., *et al.* (2002). Randomized, controlled trial of written emotional expression and benefit finding in breast cancer patients. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **20**(20), 4160–8.

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**(3), 307–17.

Torrisi, S.J., Lieberman, M.D., Bookheimer, S.Y., Altshuler, L.L. (2013). Advancing understanding of affect labeling with dynamic causal modeling. *NeuroImage*, **82**, 481–8.

Wang, X., Hutchinson, R., Mitchell, T.M., Wang, X., Hutchinson, R., Mitchell, T.M. (2003). Training fMRI classifiers to detect cognitive states across multiple human subjects. In: Advances in Neural Information Processing Systems 16 (NIPS 2003), 1–8.

Wang, Z. (2009). A hybrid SVM – GLM approach for fMRI data analysis. *NeuroImage*, **46**(3), 608–15.

Wang, Z., Childress, A.R., Wang, J., Detre, J.A. (2007). Support vector machine learning-based fMRI data group analysis. **36**(4), 1139–51.

Zeng, L., Liu, Q., Xiao, H., Chen, H. (2008). Support vector machine on functional MRI. In: Wang, R., editor. *Advances in Cognitive Neurodynamics*, Dordrecht: Springer ScienceþBusiness Media B.V., pp. 915–8.