The Representation of Emotion in Autonomic and Central Nervous System Activity

by

Philip A. Kragel

Department of Psychology & Neuroscience
Duke University

Date: ______________________

Approved:

___________________________
Kevin S. LaBar, Supervisor

___________________________
Ahmad R. Hariri

___________________________
Scott A. Huettel

___________________________
Marc A. Sommer

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology & Neuroscience in the Graduate School of Duke University

2015
ABSTRACT

The Representation of Emotion in Autonomic and Central Nervous System Activity

by

Philip A. Kragel

Department of Psychology & Neuroscience
Duke University

Date:_______________________

Approved:

___________________________
Kevin S. LaBar, Supervisor

___________________________
Ahmad R. Hariri

___________________________
Scott A. Huettel

___________________________
Marc A. Sommer

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Psychology & Neuroscience in the Graduate School
of Duke University

2015
Abstract

Phenomenologically, humans effectively label and report feeling distinct emotions, yet the extent to which emotions are represented categorically in nervous system activity is controversial. Theoretical accounts differ in this regard, some positing distinct emotional experiences emerge from a dimensional representation (e.g., along axes of valence and arousal) whereas others propose emotions are natural categories, with dedicated neural bases and associated response profiles. This dissertation aims to empirically assess these theoretical accounts by examining how emotions are represented (either as disjoint categories or as points along continuous dimensions) in autonomic and central nervous system activity by integrating psychophysiological recording and functional neuroimaging with machine-learning based analytical methods. Results demonstrate that experientially, emotional events are well-characterized both along dimensional and categorical frameworks. Measures of central and peripheral responding discriminate among emotion categories, but are largely independent of valence and arousal. These findings suggest dimensional and categorical aspects of emotional experience are driven by separable neural substrates and demonstrate that emotional states can be objectively quantified on the basis of nervous system activity.
Contents

Abstract ............................................................................................................................................... iv

List of Tables ..................................................................................................................................... x

List of Figures ................................................................................................................................... xi

1. Introduction ...................................................................................................................................... 1

  1.1 Drawing formal inferences with multivariate pattern classification ............................................ 2

  1.2 Decoding neural representations of emotion with MVPC .......................................................... 5

    1.2.1 Establishing ground truth ........................................................................................................ 6

    1.2.2 Feature selection .................................................................................................................... 7

    1.2.3 Algorithm selection ................................................................................................................. 8

    1.2.4 Generalization testing .......................................................................................................... 9

    1.2.5 Quantifying performance ...................................................................................................... 10

    1.2.6 Model comparison ............................................................................................................... 11

    1.2.7 Limitations .......................................................................................................................... 14

  1.3 Classification of emotional responding in the central nervous system ...................................... 15

    1.3.1 Decoding emotional perception ........................................................................................... 15

    1.3.2 Decoding emotional experience .......................................................................................... 21

  1.4 Classification of emotional responding in the autonomic nervous system ................................. 29

  1.5 Summary ..................................................................................................................................... 38
2. Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions

2.1 Introduction

2.2 Methods

2.2.1 Participants

2.2.2 Materials and procedure

2.2.3 Psychophysiological recording and feature extraction

2.2.4 Statistical Analysis

2.3 Results

2.3.1 Manipulation check

2.3.2 Decoding distinct emotions from response patterns

2.4 Discussion

3. Multivariate neural biomarkers of emotional states are categorically distinct

3.1 Introduction

3.2 Methods

3.2.1 Participants

3.2.2 Experimental Paradigm

3.2.3 Image acquisition

3.2.4 Preprocessing and estimating neural activation

3.2.5 Assessment of self-report

3.2.6 Pattern classification
6.2.3 Predictions and implications ................................................................. 154
6.3 Conclusion and future directions .............................................................. 156
Appendix A .................................................................................................. 159
Appendix B .................................................................................................. 173
Appendix C .................................................................................................. 174
References ................................................................................................. 176
Biography ................................................................................................. 204
## List of Tables

Table 1: Studies decoding patterns of BOLD fMRI response into distinct emotional states
.............................................................................................................................................................................. 16

Table 2: Studies decoding patterns of autonomic nervous system activity into distinct emotional states..................................................................................................................................................31

Table 3: Validation of emotion induction......................................................................................................................... 56
List of Figures

Figure 1: Different approaches for testing the functional organization of emotion in the brain with MVPC. ................................................................. 12

Figure 2: Redundant encoding of subjective valence in BOLD response patterns........ 23

Figure 3: Inputs for pattern classifiers averaged across subjects from peripheral physiology (left) and self-report (right)........................................................... 58

Figure 4: Performance metrics for classification of emotions from peripheral physiology (left) and self-report (right). ...................................................... 59

Figure 5: Dimensional sampling from self-report of emotion.................................. 60

Figure 6: Relationship between classification errors and experienced emotion along dimensions of valence and arousal (left) and seven discrete categories (right)......... 62

Figure 7: Dimensional sampling from self-report of emotion.................................. 76

Figure 8: Performance measures of multi-voxel pattern classification. ....................... 88

Figure 9: Distributed patterns of neural activity predict discrete emotional states. ...... 92

Figure 10: Stacked bar plot showing the coverage of all regions in the AAL atlas by each emotion-predictive pattern......................................................... 93

Figure 11: Categorical and dimensional models of emotion. ..................................... 96

Figure 12: Emotional states emerge spontaneously during resting-state scans......... 120

Figure 13: Coherent fluctuations in emotional state during resting-state scans.......... 122

Figure 14: Individual differences in mood and personality modulate the occurrence of spontaneous emotional brain states................................................. 124

Figure 15: Spontaneous emotional brain states exhibit correspondence with self-report. ................................................................................................. 126
Figure 16: Experimental paradigm and behavioral results................................................. 139

Figure 17: Multi-voxel pattern classification of BOLD response to facial and vocal
expressions of emotion............................................................................................................ 140

Figure 18: Topographic organization of emotion-predictive patterns............................ 142

Figure 19: A neural model of emotion representation.................................................... 150
1. Introduction

Emotion plays a prominent role in human experience. It changes the way we see the world, remember past events, interact with others, and make decisions. Emotional experiences produce feelings ranging from tranquility to full-fledged rage. Despite the ease with which these experiences are categorized and labeled, emotional feelings are often characterized as private, subjective, and notoriously difficult to quantify. As stated by Fehr and Russell, “Everyone knows what an emotion is, until asked to give a definition” (1984 p. 484). The scientific investigation of emotion has utilized behavioral observation, experiential self-report, psychophysiological monitoring, and functional neuroimaging to understand the nature of emotions.

Currently, there is considerable debate as to whether the categorical structure of experienced emotions is respected by the brain and body, stems from cognitive appraisals, or is a construction of the mind. One key issue at hand concerns whether the autonomic responses, central nervous system activity, and subjective feelings that occur during an emotional episode invariantly reflect distinct emotions or map on to more basic psychological processes from which emotions are constructed. Although reviews and meta-analyses of univariate studies have not consistently supported emotion specific patterning within these response systems (L. F. Barrett, 2006a), recent proposals
have suggested that multivariate analysis of autonomic and neuroimaging data may be better suited to identifying biomarkers for specific emotions compared to univariate approaches (Friedman, 2010; S. Hamann, 2012). It is possible that statistical limitations of univariate approaches have led to the premature dismissal of emotions as having distinct neural signatures. The remainder of this chapter will consider the promises and potential pitfalls of multivariate pattern classification and review how multivariate tools have been applied so far to test the structure of emotion in both central and autonomic nervous systems.

1.1 Drawing formal inferences with multivariate pattern classification

The crux of the problem in defining distinct emotions is that, while many changes occur throughout the body during events given an emotional label, are they consistent or specific enough to indicate the occurrence of naturally separable phenomena? For example, research has shown that fearful stimuli often elicit increases in heart rate, sympathetic nervous system activity, and neural activation within the amygdala. If fear is categorically represented in the nervous system, then it should be possible to reliably infer its occurrence given such a response profile. This type of reverse inference is notoriously difficult to make accurately if the response profile is not highly selective, as any individual measure may change with the occurrence of any
number of stimuli or psychological processes (Poldrack, 2006). For instance, the
frequency of insula activation in neuroimaging studies has been reported to be almost
one in five studies (Nelson et al., 2010; Yarkoni, Poldrack, Nichols, Van Essen, & Wager,
2011), making it difficult to make meaningful inferences based on the presence of its
activation alone. Within the domain of emotion, meta-analyses have suggested that no
individual region responds both consistently and specifically to distinct emotions (L. F.
Barrett & Wager, 2006; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012), which
has been taken as disconfirming evidence that they are represented distinctly in the
brain (L. F. Barrett, 2006a).

Utilizing machine learning algorithms for assessing the discriminability and
specificity of multivariate patterns (Jain, Duin, & Mao, 2000) is a promising new
approach for formally inferring mental states. This method, commonly called
multivariate pattern classification (MVPC; Haynes & Rees, 2006; Norman, Polyn, Detre,
& Haxby, 2006) was first employed using functional magnetic resonance imaging (fMRI)
by Haxby and colleagues (Haxby et al., 2001) to characterize the functional response of
the ventral temporal cortex during the perception of objects. This work aimed to
advance a debate centered on the organization of neural systems involved in visual
perception. While one prevalent model predicted the representation of objects was
encoded broadly throughout ventral temporal cortex in a distributed fashion, alternative theories posed a modular functional organization of the region based on the category of object, including regions selectively responsive to faces (Kanwisher, McDermott, & Chun, 1997). By constructing a model to predict the class of perceived objects from multivariate samples of imaging data, pattern classification demonstrated that overlapping distributions of neural activation encoded stimulus categories. Beyond its characterization of object selective cortex, this study highlighted that individual voxels may exhibit increased activation to many stimuli and concurrently belong to a larger pattern of activation which is functionally specific to a mental percept. In this way, the method can go beyond mean signal intensity changes averaged over voxels in univariate designs, offering a new means of drawing inferences.

Since its introduction to human neuroimaging, pattern classification has been used to infer mental states previously considered beyond the scope of fMRI. Some examples include the perception of complex scenes (Kay, Naselaris, Prenger, & Gallant, 2008), color (Brouwer & Heeger, 2009), the orientation of lines (Kamitani & Tong, 2005), the direction of motion (Kamitani & Tong, 2006), the source and content of spoken sound (Formisano, De Martino, Bonte, & Goebel, 2008), the contents of working memory (S. A. Harrison & Tong, 2009), and more broadly the occurrence of different cognitive (or
subjective) processes (Poldrack, Halchenko, & Hanson, 2009). Because of its increased sensitivity and ability to quantify the specificity of subjective mental states, MVPC overcomes the shortcomings of univariate methods that complicate studying the nature of emotion in the nervous system. Thus, adopting the multivariate framework allows researchers to formally test the extent to which emotional states are discretely represented and whether such states are organized along dimensional constructs.

1.2 Decoding neural representations of emotion with MVPC

Before reviewing how MVPC studies may inform models of emotion, here we outline the steps involved in pattern classification, as there are multiple decisions to be made in the analysis pipeline for MVPC that constrain what information can be utilized in the formulation of a classification model. These steps include how ground truth is defined for different data points, how features are selected for use in classification, what algorithm is used to distinguish between classes, how data are split between training and testing, and how to quantify performance. The selection of these analysis parameters bears particular importance when assessing how distinct emotions are represented in the nervous system (Fig 1).
1.2.1 Establishing ground truth

As pattern classification is a supervised learning technique (Duda, Hart, & Stork, 2001), knowledge about each instance of data used to construct a classifier is required in order to assign class membership. In the case of classifying emotional states, different theories offer a variety of possibilities. Theories positing a small number of categorically separate emotions (on the basis of overt behavior, or specific neural circuitry) can be directly translated into the labeling of multiple classes. Although cognitive appraisal and constructionist theories do not consider emotions to be purely categorical in nature, they do propose that distinct emotions emerge from more basic dimensions or unique patterns of appraisal (Russell, 1980; Scherer, 1984). Another alternative is to consider regularities in states produced by reinforcement contingencies (Rolls, 1990) or action tendencies (Frijda, 1987) to guide the definition of emotional states. In this sense, using categorical labels is appropriate for comparing different theories, although alternative labeling based on dimensions of valence, arousal, or components of cognitive appraisal may improve classification performance if responses cluster along constructs beyond basic emotion categories.
1.2.2 Feature selection

Determining which combination of response variables (referred to as 'features') to use in pattern classification is critical because selecting too many or the wrong ones can result in poor performance. Due to the high dimensionality of fMRI datasets (in the tens of thousands for typical datasets), reducing the number of uninformative inputs is important because the predictive power of a classifier decreases as a function of the number of features for a fixed number of instances (Hughes, 1968). While this is less of a concern when studying relatively few signals measured from the autonomic nervous system, its importance increases when confronted with thousands of measures of brain activity in the context of neuroimaging experiments. Common methods for feature selection in neuroimaging include the use of a priori regions of interest, spherical or cuboid searchlights centered around individual voxels (Kriegeskorte, Goebel, & Bandettini, 2006), masks of thresholded univariate statistics (e.g., an F map from an ANOVA), independent or principal component analysis, or a restricted field of view at high resolution. These different approaches extract information at different spatial scales and are accordingly useful for testing alternative hypothesis about the representation of emotional states in the brain. For instance, examining local patterning with searchlights may be more appropriate for identifying categorical representation of emotion in
subcortical circuits as postulated by some categorical emotion theories (Ekman, 1992; J. Panksepp, 1982). Additionally, the ability of dimensionality reduction methods to extract functionally coherent networks associated with basic psychological processes is well matched to test the constructionist hypothesis that emotions emerge from the interaction of large scale networks (Kober et al., 2008; Lindquist et al., 2012). Thus, different methods for reducing the dimensionality of fMRI data may reveal complementary ways in which emotion is represented in the brain.

1.2.3 Algorithm selection

The ability of a classifier to create an accurate mapping between input data and the ground truth class labels is dependent upon the learning algorithm used. While the varieties in learning algorithms are vast, they can broadly be designated as either linear or nonlinear (i.e. a curved boundary) based on the shape of the decision surface used to classify data. Linear methods (e.g. linear discriminant analysis (LDA) or support vector machines (SVM) with linear kernels) are more frequently used with fMRI as they have been shown to perform similar to, if not better than, nonlinear approaches (Misaki, Kim, Bandettini, & Kriegeskorte, 2010) while being straightforward to interpret. Although nonlinear algorithms are more complex, they are capable of solving a broader range of classification problems and mapping higher order representations (Hanson, Matsuka, &
Haxby, 2004). Considering different organizations of affective space, there may be instances where nonlinear approaches are advantageous for decoding distinct emotional states. In particular, if the hypothesized dimensionality of the data is smaller than the number of invariant emotional states, as is the case in most constructionist and appraisal theories, then mapping data to a higher dimensional space with a nonlinear algorithm may be advantageous.

1.2.4 Generalization testing

In order to determine if the decision rule learned by a classifier generalizes beyond the data with which it was constructed, classification must be tested on an independent set of data (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Typically, generalization involves the partitioning of a few trials within a run, different runs of data acquisition, or independent subjects for use in testing (beyond those used for training the classification model) to produce an unbiased estimate of accuracy. The way in which data are split into training and testing sets provides an opportunity to examine different claims of various emotion theories. For example, a fundamental aspect of some categorical emotion theories is the universality of emotional expressions and the conservation of their underlying neural mechanisms (Ekman & Cordaro, 2011). Between subject classification is one way of providing evidence for this claim (although negative
results would be difficult to interpret as poor generalization may result from differences in fine scale neural anatomy, temporal dynamics, or other factors unrelated to representational content). Alternatively, testing on independent trials or runs may prove more fruitful if emotional states are hypothesized to be idiosyncratic to the individual or particular context as emphasized by appraisal and constructionist theories (L. F. Barrett, 2006b; Scherer, 1984) because these testing schemes relax assumptions of response invariance.

1.2.5 Quantifying performance

A classifier is generally considered to exhibit learning if it is capable of classifying independent data at levels beyond chance. While this criterion is important when testing performance, it does little to differentiate between alternative emotion theories. For example, two classifiers could exhibit the same levels of overall accuracy, sensitivity, and specificity, yet have uniquely different error distributions. Testing how the structure of errors corresponds to predictions of different emotion theories is particularly relevant in decoding emotional states, because distinct emotions have been hypothesized by different theories to be completely orthogonal (Ekman, 1992; Tomkins, 1962), clustered within a multidimensional space (Arnold, 1960; Scherer, 1984; C. A. Smith & Ellsworth, 1985), or arranged in a circumplex about dimensions of valence and
arousal (L. F. Barrett, 2006b; Russell, 1980). Identifying the structure of information used in classification provides a link between neural data and conceptual and computational theories of emotion. In this way, multivariate classification can reveal how different constructs are instantiated in the brain or periphery instead of focusing on whether certain emotional states are biologically basic.

1.2.6 Model comparison

Given that psychological models propose that distinct emotional states emerge through different processes, the way in which a classifier performs can be used to test predictions of alternative theoretical models. While both categorical emotion theories and appraisal models assume some degree of consistency in the neural mechanisms underlying emotional states, constructionist models claim that distinct emotions do not have invariant neural representations. Altering the generalization testing in a classification pipeline is one means of testing this hypothesis. For instance, a classifier trained during the experience of core disgust and tested on moral disgust occurring from the violation of social norms could reveal which response components are shared between the two emotional states. Additionally, appraisal models and constructionist models propose that different factors organize distinct emotions, and as such the similarity of neural activation patterns should reflect this structure. If there is increased
Figure 1: Different approaches for testing the functional organization of emotion in the brain with MVPC.

The flexibility of analysis parameters in pattern classification permits posing different experimental questions about the neural basis of emotion under the same framework (left column). By adjusting the method of feature selection employed (middle column), information contained in the activity of large scale networks, specific regions, or searchlights throughout the brain focus analysis onto different spatial scales. Adopting different learning algorithms and classification schema (right column) permits testing of diverse hypotheses.
similarity for emotional states that are more closely related along dimensions of valence and arousal (e.g., fear and anger), then evidence would support a constructionist model of emotion, whereas if they are related along appraisal dimensions such as novelty, valence, agency, or goals, then appraisal models would be favored. Generally, the similarity of emotional states reflected in neural activation patterns, quantified with classification errors, can be related to the similarity proposed by competing models (for an example using models of perception see Kriegeskorte, Mur, & Bandettini, 2008).

While different models of emotion are often characterized as competing, classification results supporting one model are not necessarily evidence against alternative models. It is possible that different branches of the nervous system, separate brain regions, or neural activation at diverse spatial scales, better conform to alternative conceptions of emotion. Moreover, given separate bodies of behavioral work supporting both dimensional and categorical aspects of emotion, it may not be advantageous to search for any single ‘best’ model. Investigating in what neural systems or under which contexts dimensional or categorical constructs are observed may provide more leverage for advancing theories of emotion. Thus, depending on the structure of information revealed by classification, MVPC is capable of discerning how distinct emotions are
reflected in activity across multiple levels of the neuraxis – either as constructions, patterns of appraisal, or independent categories.

1.2.7 Limitations

While multivariate classification approaches are theoretically promising in terms of identifying emotion specific patterns of neural activity, there are several caveats to their use which limit the inferences that can be drawn from significant results. Importantly, classifiers will utilize any information which is capable of differentiating the cases to be labeled. In the context of classifying patterns of neural activity into emotional states it is possible that results are driven by factors unrelated to emotion per se, such as motor behavior, task induced artifact (e.g., movement-related or psychophysiological confounds across conditions), perceptual or semantic differences in stimuli. When the goal of a study is to identify the patterns of BOLD response which best represent a single construct, adequately accounting for potential confounds is necessary – especially given the increased sensitivity of the method relative to conventional regression models (Todd, Nystrom, & Cohen, 2013). Both carefully designed experiments and appropriate inferential models are thus critical in using MVPC to inform theories of emotion.
1.3 Classification of emotional responding in the central nervous system

Relative to the large body of univariate work examining the relationship between fMRI activation and distinct emotions, a small number of studies are beginning to use MVPC to map neural activation onto specific emotional states. Here, the extent to which emotional states, as defined by either dimensional or categorical models (i.e. classifying subjective valence/arousal or distinct emotions such as fear and anger), are represented by specific patterns of neural activation is reviewed. Because emotional perception and experience engage different processes and are subserved by separate neural systems (T. Wager et al., 2008), the studies reviewed are organized based on the process most likely engaged during functional imaging (Table 1).

1.3.1 Decoding emotional perception

The majority of studies using MVPC to decode emotions have investigated the perception of emotional stimuli. The goal of these studies is to use blood-oxygen level dependent (BOLD) activation patterns to infer the perceived emotional content in a stimulus, as determined by subjective evaluative judgments. This approach differs from predicting the physical similarity of stimuli, which may not necessarily correspond to the perceptual state of participants (for a review and discussion, see Tong & Pratte, 2012). Stimulus sets are often comprised of facial, vocal, and gestural expressions due to
Table 1: Studies decoding patterns of BOLD fMRI response into distinct emotional states

<table>
<thead>
<tr>
<th>Study</th>
<th>Emotion Model</th>
<th>Class Labels</th>
<th>Component Process</th>
<th>Stimulus Modality</th>
<th>Feature Selection</th>
<th>Classification Algorithm</th>
<th>Cross-Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Said et al. 2010</td>
<td>Categorical</td>
<td>Anger, Disgust, Perception</td>
<td>Visual</td>
<td>Multiple</td>
<td>Logistic regression (linear)</td>
<td>Run</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fear, Sadness, Happiness,</td>
<td></td>
<td>a priori</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surprise</td>
<td></td>
<td>regions of interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pessoa and Padmala 2007</td>
<td>Categorical</td>
<td>Fear, No Fear</td>
<td>Perception</td>
<td>Visual</td>
<td>SVM (linear and nonlinear)</td>
<td>Trial</td>
<td></td>
</tr>
<tr>
<td>Ethofer et al. 2009</td>
<td>Categorical</td>
<td>Anger, Sadness, Joy, Relief, Neutral</td>
<td>Perception</td>
<td>Auditory</td>
<td>SVM (linear)</td>
<td>Trial</td>
<td></td>
</tr>
<tr>
<td>Kotz et al. 2012</td>
<td>Categorical</td>
<td>Anger, Sadness, Happiness,</td>
<td>Perception</td>
<td>Auditory</td>
<td>SVM (linear)</td>
<td>Run</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surprise, Neutral</td>
<td></td>
<td>Whole brain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>searchlight</td>
<td></td>
<td>searchlight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baucom et al. 2012</td>
<td>Dimensional</td>
<td>P, N or LA, HA, HAN, LAN,</td>
<td>Experience</td>
<td>Visual</td>
<td>Logistic regression (linear)</td>
<td>Subject, run</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LAP</td>
<td></td>
<td>Whole brain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolls et al. 2009</td>
<td>Dimensional</td>
<td>P, N</td>
<td>Experience</td>
<td>Thermal</td>
<td>Probability estimation, multilayer perceptron, SVM (linear)</td>
<td>Trial</td>
<td></td>
</tr>
<tr>
<td>Sitaram et al. 2011</td>
<td>Categorical</td>
<td>Happiness, Disgust, Sadness,</td>
<td>Experience</td>
<td>Imagery</td>
<td>SVM (linear)</td>
<td>Subject, run</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disgust, Happiness</td>
<td></td>
<td>Whole brain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>and multiple</td>
<td></td>
<td>and multiple</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a priori regions of interest</td>
<td></td>
<td>a priori regions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>of interest</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P = positive, N = negative, LA = low arousal, HA = high arousal, HAN = high arousal negative, LAN = low arousal negative, LAP = low arousal positive, HAP = high arousal positive, SVM = support vector machine
their ability to elicit discrete emotional responses and convey unique information critical for motivating behavior and social communication (Adolphs, 2002).

Studies examining the perception of emotion in facial expressions are beginning to reveal how discrete emotion categories are represented neurally. One investigation (Said, Moore, Engell, Todorov, & Haxby, 2010) tested whether multivariate patterns of activation in regions that commonly respond to facial expressions, i.e., frontal operculum, along with anterior and posterior aspects of the superior temporal sulcus (STS), could be rated by an independent group of subjects. These results demonstrate that fine scale patterns of activation within STS contain movement related information capable of specifying the category of perceived facial affect. Most importantly, univariate analysis of the same activation patterns yielded a nonspecific mapping to the perception of emotion. Thus, the information carried within the STS depends on the underlying spatial distribution of neural activation in a fashion univariate analysis cannot detect. Therefore, the functional specificity of the STS may be improperly assessed if only tested with a univariate approach.

Another study (Pessoa & Padmala, 2007) examined emotional perception by showing participants faces with happy, neutral, and fearful facial expressions at near subliminal durations (33 or 67 ms) followed by neutral face masks and subsequently testing if a fearful face was perceived. The aim of the study was to determine if patterns
of fMRI activation could predict the response of a participant, thus representing a mental state of perceived fear. Combinations of voxels from multiple *a priori* regions of interest (found to discriminate behavioral choice when assessed with a univariate analysis; Pessoa & Padmala, 2005) predicted the choice with increasing accuracy as more regions were included in the analysis. Further, the information contributed when incorporating multiple regions was found to add synergistically, such that the total decision-related information was greater when two regions were considered jointly compared to the sum of the information when each was considered individually. Information contributed by the amygdala in particular was found to most significantly improve the performance of classifiers. These findings indicate that regions implicated in different aspects of affective processing (e.g., representing the stimulus, or modulating autonomic responses) form a distributed representation of an emotional percept, and that conjoint analysis of independent information can more accurately characterize emotional states. While studies examining emotional perception using facial expressions have found discriminable patterns of BOLD response, the reported studies used posed rather than naturalistic stimuli. Understanding whether response patterns to naturally occurring facial expressions are similar to those for staged or prototypical expressions remains an open area of research.
Studies investigating emotion perception of human vocalizations have similarly established categorical specificity of fMRI activation patterns. One such study (Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009) demonstrated that spatially distinct patterns of activation within voice sensitive regions of auditory cortex, established with an independent localizer, could discriminate between pseudowords spoken with emotional prosody of anger, sadness, neutral, relief, and joy. While voice sensitive cortex was consistently activated across all stimuli in the experiment, no differences in the average response amplitude were observed between categories in a univariate analysis. Furthermore, errors committed by classifiers were most frequent for stimuli which shared similar arousal, but not valence. The authors interpreted this relationship to be driven by alterations in the fundamental frequency of vocalizations, which may be captured in the spatial variability of activation patterns. Other work aiming to decode the emotional prosody of speech from local patterns of fMRI activation (Kotz, Kalberlah, Bahlmann, Friederici, & Haynes, 2012) yielded similar results. Using the searchlight approach, these authors found that a distributed network of regions discriminated among the perception of angry, happy, neutral, sad and surprised vocal expressions. In addition to superior temporal gyrus, this predominantly right-lateralized network included inferior frontal operculum, inferior frontal gyrus, anterior superior temporal sulcus, and middle frontal gyrus. While univariate studies have implicated these regions
in a number of aspects of vocal processing (e.g., speaker identity, fundamental frequency, or semantic content), future work is needed to confirm that they contribute distinct information to a distributed representation of perceived emotion.

In sum, MVPC studies of emotional perception have demonstrated that information carried in fine-grain structure across multiple cortical sites is capable of signifying affective information in the environment in distinct ways. Findings that information reflected in local patterns (e.g., Ethofer et al., 2009; Said et al., 2010), but not mean activation levels at the centimeter scale, is consistent with a growing body of evidence from multivariate decoding of perception. Together, these findings suggest that the perceptual qualities of a stimulus that confer emotional significance are reflected in patterns of BOLD activation within a number of cortical regions. Further, classification utilizing voxels spanning multiple distributed regions showed that different types of information could be combined synergistically to represent a perceptual state. By considering multivariate information at various scales, MVPC studies have revealed patterning specific to the perception of distinct emotional stimuli, a difficult goal to achieve with conventional univariate fMRI (Winston, O’Doherty, & Dolan, 2003) given its relatively limited functional specificity and spatial resolution.
1.3.2 Decoding emotional experience

In addition to examining emotional percepts, MVPC has been used to infer the subjective experience of emotion on the basis of functional activation patterns. Although the distinction between emotional perception and experience is often blurred, as the perception of emotional stimuli can lead to the experience of emotional feelings, here studies that attempted to decode subjective states on the basis of introspective self-report are reviewed. Work examining the neural underpinnings of emotional feelings using univariate approaches (A. R. Damasio et al., 2000) has qualitatively characterized the similarities and differences between distributed patterns of activation during emotional episodes. While comparisons of this nature are capable of grossly characterizing which brain regions are engaged during the experience of a specific emotional feeling, they lack the quantitative precision necessary to map patterns of neural activation to specific components of an emotional episode. For instance, Damasio and colleagues (2000) examined response patterns within subcortical structures (e.g. hypothalamus, brainstem nuclei) and cortical regions (insula, anterior cingulate, SII), which collectively exhibited altered response profiles during the experience of different emotional feelings. Some regions were selectively engaged during one emotion relative to others (for instance, happiness and sadness produced activation in different portions of the insula - among many other regions), but the extent to which these differences
were specific to a particular emotion was not assessed with the univariate approach employed.

The application of MVPC can precisely characterize how patterns of neural activation signify distinct emotional experiences by quantifying where in the brain information specific to a particular emotion is represented, and the extent to which neural activation patterns specifically and accurately reflect a given emotional state. Initial MVPC studies decoding the experience of emotion have often utilized dimensional models, primarily focusing on the classification of positive and negative episodes. One such study (Baucom, Wedell, Wang, Blitzer, & Shinkareva, 2012) implemented an implicit emotion induction where participants incidentally viewed blocked presentations of images that had been previously validated to elicit positive and negative affect at high and low levels of arousal. Using classifiers constructed to decode the experience of valence and or arousal, whole brain activation patterns revealed a common neural representation of emotional experience sparsely distributed throughout the brain. Further, classification accuracies were above chance levels when cross-validation was performed within and between subjects. Additional multivariate analyses utilizing multidimensional scaling of the most stable 400 voxels identified from pattern classification revealed that the most variance in fMRI activation could be explained along dimensions of valence and arousal. While findings from this work
Figure 2: Redundant encoding of subjective valence in BOLD response patterns.

Classification accuracy obtained using an increasing number of voxels in discriminating positive and negative valence. (A) Solid lines depict gains in information and accuracy (left panels), which increase asymptotically in a nonlinear fashion when voxels from mPFC (right panel) are added to classification\(^1\). (B) Box-plot (left) illustrates that including additional voxels in a whole brain feature search improves the classification accuracy of valence in a sub-linear fashion\(^2\).


demonstrate the feasibility of inferring emotional states from fMRI activation patterns, one main issue limits the implications of the study. Although the stimuli used were previously shown to produce emotional responses differing in valence and arousal, no on-line reports of emotional experience were conducted. As such, it remains unclear whether information related to the perceptual processing of stimuli or the affective experience of participants was informing the classification of fMRI data.

Additional work examining the neural representation of positive and negative experience has focused on the subjective pleasantness of thermal stimulation (Rolls, Grabenhorst, & Franco, 2009). In this work, the subjective pleasantness of thermal stimulation to the hand was successfully inferred from local patterns of fMRI activation in frontal cortex. Consistent with results from the whole brain classification conducted by Baucom and colleagues (2012; see Fig 2), information was redundantly pooled across local voxels such that classification performance increased sub-linearly as the number of voxels used increased, even when adding voxels from different regions. This finding suggests that the information encoded by each voxel was not independent, but rather was redundantly carried through multiple voxels. Taken together, these initial studies suggest widely distributed and highly redundant patterns of BOLD response contain valence-related information, possibly implicating a common neuromodulatory source,
such as dopamine, given its role in appetitive motivation and its broad projections to prefrontal cortex (Ashby, Isen, & Turken, 1999). A regional modulation of neural activity due to transient dopamine release (Phillips, Ahn, & Howland, 2003) is consistent with the location and redundancy of information observed. Given the causal role of neurotransmitter systems in generating affective states proposed by neurobiological models of emotion (Jaak Panksepp, 1998), understanding the link between these systems and patterns of neural activation is a critical area for future research.

In addition to decoding the experience of emotion along affective dimensions, categorical emotion models have been used to guide MVPC. In one such study (Sitaram et al., 2011), pattern classifiers implementing support vector machine algorithms to classify patterns of whole brain activity were shown to accurately discriminate among the experience of happiness, sadness, and disgust during imagery in real time for individual participants. Off-line analysis of classifiers demonstrated that a number of cortical and subcortical regions each contributed to classification performance. In this analysis, regions of interest from prior meta-analyses (Frith & Frith, 2006; Ochsner, 2004; Ochsner & Gross, 2005; Phan, Wager, Taylor, & Liberzon, 2002), including cortical (middle frontal gyrus, superior frontal gyrus, and superior temporal gyrus) and subcortical (amygdala, caudate, putamen, and thalamus) regions, could be used to decode the experience of disgust versus happiness. These results are striking in
comparison to univariate meta-analytic work (Lindquist et al., 2012; Phan et al., 2002) in which several of these regions, the medial prefrontal cortex in particular, exhibited little to no specificity for distinct emotions when treated as homogeneous functional units. One important limitation of this study, however, was the relatively small number of emotions sampled. Because often only a single positive and negative emotion pair was decoded (i.e. disgust and happiness), information used during classification could either reflect differences in basic emotion categories or affective dimensions such as valence.

Other work using imagery to induce distinct emotional states by Kassam and colleagues (2013) identified the most stable voxels from whole brain acquisitions of fMRI data to classify states of anger, disgust, envy, fear, happiness, lust, pride, sadness, and shame. Participants were presented with two cue words for each emotion which remained in view for 9 seconds during while imagery was performed. The authors demonstrated the localization, generalizability, and underlying structure of emotion-related information. Multi-voxel patterns drawn from the most stable 240 voxels commonly included a distributed set of areas spanning frontal, temporal, parietal, occipital, and subcortical regions. Using Gaussian Naïve Bayes classifiers, the nine states were classified with 84% accuracy when generalization was performed within subjects and 71% accuracy for between subject classification. Furthermore, when generalizing to the perception of images, disgusting and neutral stimuli were classified within
individuals at 91% accuracy. Investigating the occurrence classification errors revealed similarities between positive states of happiness and pride and isolated states of lust as being the least similar to the other elicited states. Consistent with these findings, a factor analysis conducted on neural activation patterns revealed constructs interpreted as valence, arousal, sociality, and lust to underlie the variability in patterns. While comprehensive in its characterization of emotion patterning, it is unclear to what extent the classification results were dependent on semantic processing. Given the repeated presentation of related words (e.g. repeated presentations of ‘afraid’ and ‘frightened’ were discriminated from those of ‘gloomy’ and ‘sad’) and the presence of a left lateralized prefrontal network implicated in semantic processing (Martin & Chao, 2001), it is possible that the observed results are partly due to the classification of semantic information.

In sum, the few studies that have conducted multivariate classification of emotional experience suggest that regions conventionally considered to be nonspecific in a univariate sense are capable of specifying the experience of a distinct emotion at a multivariate level. While this line of investigation is in its infancy, it shows promise for moving beyond limitations of univariate, locationist approaches to analyzing brain function. Consider, for instance, activation in medial prefrontal cortex (mPFC) to emotional stimuli. Meta-analyses of fMRI activation show that this region is commonly
engaged during multiple emotional states, implicating the mPFC as a functional unit engaged by a broad array of emotional content (Phan et al., 2002). On the other hand, evidence from MVPC reviewed above suggests that at finer spatial scales, activity within mPFC is capable of discriminating between emotional states. As this region of medial frontal cortex has been hypothesized to play many roles, including emotional appraisal (Kalisch, Wiech, Critchley, & Dolan, 2006), attribution (Ochsner et al., 2004) and regulation (Goldin, McRae, Ramel, & Gross, 2008), it is possible that these processes are differentially engaged during the experience of different emotions. Such varied recruitment may be evident in the structure of spatial activation patterns that predict subjective emotional experience. Identifying which processes contribute to distinct emotional representations in mPFC remains an open area of research.

In addition to mPFC, activation patterns within the amygdala could consistently be used to decode the subjective experience of emotion along dimensions of valence and arousal (Baucom et al., 2012; Sitaram et al., 2011), converging with a large body of univariate work. The amygdala has been implicated in numerous affective processes (Davis & Whalen, 2001); more specifically it has been found to functionally covary with the subjective experience of arousal (A. K. Anderson et al., 2003; Colibazzi et al., 2010; Liberzon, Phan, Decker, & Taylor, 2003; Phan et al., 2003) and valence (Anders, Lotze, Erb, Grodd, & Birbaumer, 2004; Gerber et al., 2008; Posner et al., 2009). One possibility is
that pattern classifiers read out information from distinct populations of neurons
specific to positive and negative valence (Paton, Belova, Morrison, & Salzman, 2006) to
predict the ongoing affective state. Given the potential of MVPC to utilize information
beyond the conventional resolution of fMRI by capitalizing on changes in the
distribution of neurons from voxel to voxel (Brouwer & Heeger, 2009; Freeman,
Brouwer, Heeger, & Merriam, 2011; Kamitani & Tong, 2005; but see Op de Beeck, 2010),
these findings show promise in advancing our understanding of the functional
architecture of the amygdala at increasingly fine resolution. Future work is required to
determine whether there are neural populations within the amygdala that specifically
code for valence and arousal, and whether distinct neural populations within the
amygdala are essential for experiences commonly labeled as fear.

1.4 Classification of emotional responding in the autonomic nervous system

In addition to work investigating emotion specificity of central nervous system
activation, a growing body of evidence is supporting the ability of MVPC to extract
emotion specific patterns from multivariate sampling of the autonomic nervous system
(ANS). While the emotions induced, stimuli used, and general experimental protocol are
similar between studies examining central and autonomic patterning, the response
variables being classified are inherently quite different. Psychophysiological studies
typically measure both the sympathetic and parasympathetic branches of the ANS,
including measures of cardiovascular, electrodermal, respiratory, thermoregulatory, and gastric activity. Here we review psychophysiological studies whose primary goal is to use pattern classification to predict the emotional state of participants (Table 2). In particular we focus on the extent to which information carried in the activity of the ANS is capable of discriminating distinct emotional states and how this information is organized (e.g., as discrete emotion categories or along dimensional constructs such as valence or arousal).

Early work applying pattern classification methods to test the emotional specificity of autonomic responding performed by Schwartz and colleagues (1981) sought to differentiate states of happiness, sadness, fear, anger, relaxation, and a control condition using cardiovascular measures. Entering measures of heart rate and systolic/diastolic blood pressure into a classifier utilizing stepwise discriminant analysis, linear combinations of cardiovascular response were found to significantly differentiate the six conditions with an accuracy of 42.6%, while chance was approximately 16.6%. Of the five discriminant functions produced to differentiate between conditions, two accounted for 96% of the total explained variance. Notably, the first discriminant function (the linear combination of heart rate and blood pressure changes which best
Table 2: Studies decoding patterns of autonomic nervous system activity into distinct emotional states

<table>
<thead>
<tr>
<th>Study</th>
<th>Class Labels</th>
<th>Stimulus Modality</th>
<th>ANS Measures</th>
<th>Classification Algorithm</th>
<th>Classification Accuracy (Chance; I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwartz et al., 1981</td>
<td>Happiness, Sadness, Anger, Fear, Relaxation, Neutral</td>
<td>Imagery</td>
<td>Cardiovascular</td>
<td>LDA</td>
<td>42.6% (16.6%; .312)</td>
</tr>
<tr>
<td>Sinha and Parsons, 1996</td>
<td>Fear, Anger, Neutral</td>
<td>Imagery</td>
<td>Cardiovascular, electrophysiological and thermal</td>
<td>LDA</td>
<td>66.5% (33.3%; .498)</td>
</tr>
<tr>
<td>Rainville et al., 2006</td>
<td>Fear, Anger, Sadness, Happiness</td>
<td>Imagery</td>
<td>Cardiovascular and respiratory</td>
<td>LDA</td>
<td>49.0% (25.0%; .320)</td>
</tr>
<tr>
<td>Christie and Friedman, 2004</td>
<td>Amusement, Anger, Contentment, Disgust, Fear, Sadness, Neutral</td>
<td>Audiovisual</td>
<td>Cardiovascular and electrophysiological</td>
<td>LDA</td>
<td>37.4% (14.3%; .270)</td>
</tr>
<tr>
<td>Kreibig et al., 2007</td>
<td>Fear, Sadness, Neutral</td>
<td>Audiovisual</td>
<td>Cardiovascular, electrophysiological and respiratory</td>
<td>LDA³</td>
<td>69.0% (33.3%; .535)</td>
</tr>
<tr>
<td>Nyklicek et al., 1997</td>
<td>Happiness, Serenity, Sadness, Agitation</td>
<td>Auditory</td>
<td>Cardiovascular and respiratory</td>
<td>LDA</td>
<td>46.5% (20%; .331)</td>
</tr>
<tr>
<td>Stephens et al., 2010</td>
<td>Amusement, Contentment, Surprise, Fear, Anger, Sadness, Neutral</td>
<td>Audiovisual</td>
<td>Cardiovascular, electrophysiological and respiratory</td>
<td>LDA</td>
<td>44.6% (14.3%; .354)</td>
</tr>
</tbody>
</table>

LDA = linear discriminant analysis, SVM = support vector machine, I = proportional reduction in error rate.

³ Kolodyazhni and colleagues (2011) reanalyzed this dataset using quadratic discriminant analysis, radial basis function neural networks, multilayer perceptrons, and k-nearest neighbors clustering – with nonlinear algorithms showing improvements over linear methods.
differentiated the six conditions) corresponded to broad cardiovascular activation, positively weighting all three measures. This function generally corresponded to arousal, wherein discriminant weights increased along a spectrum from relaxation to anger. The second discriminant function less clearly mapped onto a psychological construct, having positive weights for both fear and sadness, and negative weights for happiness and a neutral control condition. While this study sampled a single branch of the ANS and accordingly was limited to three inputs for use in classification, it clearly demonstrates that differences in cardiovascular arousal are capable of partially differentiating distinct emotional states.

Subsequent studies examining response patterning during emotional imagery have further established the specificity of autonomic responding by incorporating a number of additional physiological measures beyond cardiovascular activity. In one such study (Sinha & Parsons, 1996), the extent to which fear, anger, action, and neutral states induced with imagery could be differentiated using measures of electrodermal, cardiovascular, thermal, and facial muscle activity was tested using discriminant function analysis. While this work did not exclusively test the specificity of autonomic responses (as classification was conducted using both autonomic and somatic responses), the largest reported differences between the induction of fear and anger were cardiovascular changes – suggesting that autonomic responding likely contributed
considerably to classification. Classification of peripheral responses from two separate sessions yielded an internal accuracy of 84% and a cross-validated accuracy of 66.5% when generalizing from one session to the other (in both cases chance was approximately 33.3%). Although little can be inferred about the structure of affective information because only two negative emotional states were decoded, this work further established the predictive capacity of autonomic responses.

Work by Rainville and colleagues (2006) additionally examined the specificity of autonomic responses, sampling cardiovascular and respiratory changes during states of fear, anger, sadness, and happiness induced via imagery. To more precisely characterize the contribution of sympathetic and parasympathetic branches of the ANS, the authors extracted multiple temporal and spectral measures of cardiac and respiratory activity from the raw electrocardiogram and respiratory data. This feature extraction produced a set of 18 variables which were subsequently decomposed into a smaller number of dimensions using principal component analysis, producing five components explaining 91% of the variance in the original data. Loadings of the five components showed that the principal component analysis separated variance related to parasympathetic activity coupled with respiration, parasympathetic activity independent of respiration, and sympathetic activity (as well as respiration amplitude and variability). Pattern classification utilizing discriminant analysis on component loadings yielded an accuracy
of 65.3% using resubstitution and 49.0% with leave-one-out cross validation, with chance accuracy equal to 25%. By using more innovative feature extraction and reduction methods, this work suggests that multiple independent mechanisms underlie autonomic changes capable of discriminating emotional states in a specific manner.

In addition to imagery, instrumental music or cinematic film clips are commonly used to induce emotional states in research investigating autonomic patterning, due to the specificity of experiences produced and the availability of validated stimuli. One such study by Christie and Friedman (2004) used film clips to induce distinct emotional states of amusement, contentment, anger, fear, sadness, disgust, and neutral during concurrent acquisition of cardiovascular and electrodermal responses. Feature extraction from raw measures yielded 6 variables for classification: heart period, mean successive difference of heart period, tonic skin conductance level, systolic blood pressure, diastolic blood pressure, and mean arterial pressure. With the exception of disgust, all conditions could be predicted above chance levels from these variables with an average correct classification rate of 37.4% (chance being 14.3%). This absolute level of accuracy is somewhat lower than that observed in studies classifying a smaller number of emotional states (e.g., Kreibig, Wilhelm, Roth, & Gross, 2007), due to both lower chance rates and increased difficulty of the classification problem. A separate discriminant function analysis excluding response patterns for disgust, as it was not discriminated above
chance levels in pattern classification, revealed that self-report variables differentiated emotional states along a structure organized by valence and activation (66.11% and 14.31% of explained variance), whereas autonomic variables better mapped on to dimensions of activation and action tendency (58.29% and 14.4% of explained variance). Discriminant weights on the primary autonomic factor revealed that high activation was accompanied by high values of skin conductance and negative values for mean successive difference in heart period, suggestive of sympathetic activation and decreased vagal influences. This work suggests that the information content of autonomic signals predicts emotional states in a manner distinct from the self-report of emotional experience.

The predominance of broad physiological activation in the ANS is consistent with other work decoding cardiac and respiratory responses into the experience of happiness, sadness, serenity, agitation, and neutral states elicited with instrumental music (Nyklicek, Thayer, & van Doornen, 1997). In this study, discriminant analysis capable of classifying five emotional states with an average accuracy of 46.5% produced two functions interpreted as arousal and valence (62.5% and 10.0% of explained variance, respectively). The first discriminant function differentiated happiness and agitation from sadness and serenity. This function was weighted negatively for respiration rate and weighted positively for inhale time, exhale time, inter-beat interval,
and respiratory sinus arrhythmia measures. Although the second discriminant function was considered uninterpretable as it did not meaningfully correspond to any psychological construct, the third discriminant function mapped on to valence, differentiating happiness and serenity from sadness and agitation. The physiological weights of the function were negatively associated with inter-beat interval, diastolic blood pressure, and left ventricular ejection time. Contrary to other studies reviewed here, this work identified valence and arousal as the predominant factors in differentiating emotional states using autonomic measures – a structure commonly found in self-reported affect. It is possible, however, that the selection of stimuli drove these differences as the emotional states examined all varied either on dimensions of valence or arousal.

Studies investigating autonomic patterning across both film and instrumental music have examined how well classification models can find solutions that are capable of differentiating emotional states induced by stimuli with vastly different perceptual properties. Stephens and colleagues (2010) examined the extent to which measures of heart rate variability, peripheral vascular changes, systolic time intervals, respiratory changes, and electrodermal activity could be used to predict the induction of contentment, amusement, surprise, fear, anger, sadness and a neutral state from across
both music and film. Discriminant analysis using ANS measures could correctly classify emotional state with an accuracy of 44.6% given chance rates of 14.3%.

While accuracy rates were above chance for both music and film inductions, the weights of discriminant functions were not examined, possibly missing the role of constructs such as valence or arousal in the differentiation of autonomic responses.

In sum, this body of work using MVPC shows there is information in patterns of ANS activity that can be mapped to distinct emotions, although the true nature of this information is far from clear. There is mixed evidence concerning the extent to which emotion specific patterns are organized along a dimension of arousal, as only 3 of the 7 studies reviewed identified such a relationship. Only a single study identified discriminatory patterns which corresponded to the experience of valence, which only accounted for 10% of variance in autonomic responses (Nyklicek et al., 1997). Given this evidence, it is unlikely that information contained in the ANS is represented equivalently as emotional states are experienced (either as valence and arousal or discrete categories), but that information is organized in an alternative low dimensional embedding, or configuration, of affective space. One possibility is that ANS activity follows along such dimensions within regions of affective space, for example sympathetic activation tracking experienced arousal, but interactions between
sympathetic and parasympathetic branches produce nonlinear regions (Berntson, Cacioppo, Quigley, & Fabro, 1994), which complicate a direct mapping.

1.5 Summary

Studies decoding emotional states have demonstrated that patterns of central and autonomic nervous system activity carry emotion-specific information, yet much remains to be understood about how this information is integrated into a distinct, coherent emotional experience. In particular, the representational format of information remains to be precisely characterized according to categorical and dimensional models of emotion. This thesis examines how the representational content of neural activity maps on to that of unique emotional experiences, allowing theoretical models of emotion to be effectively compared, addressing longstanding debates, and advancing the field.
2. Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions

This chapter describes a psychophysiological experiment examining whether information conveyed in autonomic signals is represented along categorical or dimensional models of emotion. In this experiment, cinematic film clips and instrumental music were used to induce emotional states of anger, fear, sadness, surprise, amusement, contentment, in addition to a neutral state (Gross & Levenson, 1995; Stephens et al., 2010). Peripheral measures of electrodermal, cardiac, respiratory, and gastric activity were acquired to broadly sample autonomic responding. On-line sampling of dimensional and categorical aspects of emotional experience was performed for the purpose of model comparison. Quantifying the accuracy with which patterns of experiential and autonomic responding can be classified along discrete emotions approximates the information-content of response patterns. Examining precisely how classification errors are made provides insight as to how information is represented in autonomic signals – either along categorical or dimensional models. The findings presented in this chapter have been published as a research article in the journal *Emotion*. Copyright © 2013 by the American Psychological Association. Reproduced with permission. The official citation that should be used in referencing this material is Kragel, P. A., & LaBar, K. S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. Emotion, 13(4), 681-690. doi:
A core theoretical question in emotion research concerns understanding how emotions are organized and represented in behavior and bodily responses. Two predominant views on the structure of emotions include dimensional and categorical models. Dimensional models organize emotions based on their coordinates in an affective space commonly anchored using valence and arousal (or related constructs) as dimensional axes (Lang, 1994; Russell, 1980, 2003). The spatial representation, or affective space, posited by dimensional theories typically contains two orthogonal axes, arousal and valence, from which a specific emotion can be charted. This organization implies that distinct emotions are interrelated along these two dimensions. For example, fear and disgust bear similarity to each other because they both lie on the negative hemiplane of affective space. Alternatively, categorical theories propose that emotions form distinct clusters of response properties and often assume a core set of basic emotions from which more complex representations can be built (Ekman, 1992; C. E. Izard, 1992; R. Plutchik, 2001). Categorical theories posit that affective space is constructed from a small set of orthogonal monopolar axes, each corresponding to a distinct emotion. A critical aspect of this characterization is the independence between
emotions. For instance, fear and disgust are considered independent phenomena. While they may share some common features, such as being hedonically negative, the occurrence of one emotion does not inform the likelihood of experiencing another.

Categorical and dimensional theories predict different relationships between emotions and the structure of affective space. For categorical theories, each emotion is hypothesized to be governed by a distinct biological mechanism and specific physiological state. This response configuration should produce a sparse structure in which emotions are distant from each other. Dimensional accounts of emotion suggest responses should be interrelated along either valence or arousal dimensions (although the general consensus is that both these dimensions are required; for a review of models see Feldman Barrett & Russell, 1999). Thus, affective space should be densely packed as it has fewer orthogonal dimensions. While these conceptualizations of affective space differ in their theoretical underpinnings, they are not entirely contradictory. For example, a circumplex model (Russell, 1980) has certain features similar to those suggested by categorical emotion theory. A circumplex configuration in a dimensional model orders emotions along the perimeter of a circle contained by the plane created by the valence and arousal dimensions. As circumplex configurations typically require equal spacing along their perimeter (Acton & Revelle, 2004), placing more intense instances of affect further from each other increases the sparsity of affective space while
maintaining two bipolar dimensions. In this way, both sparsity and clustering of emotions can be common to both categorical and dimensional emotion theories. It is therefore essential to determine how emotional responses align to both models, as they are critically different yet bear some commonalities.

There is no consensus on the structure of emotions, and synthesis of current evidence has led some researchers (e.g. L. F. Barrett, 2006a) to abandon the notion of emotion categories as “natural kinds.” A primary argument made to support this claim is the lack of evidence from psychophysiological studies and meta-analyses of functional neuroimaging data. Despite a long history of accumulating evidence supporting autonomic specificity between emotions (Ax, 1953; Ekman, Levenson, & Friesen, 1983; Levenson, Ekman, & Friesen, 1990; Stemmler, 1989), such reviews suggest that there is more consistency of findings differentiating positive and negative valence, as opposed to distinct indices of sub-categories of emotion in both autonomic (Cacioppo, Tassinary, & Berntson, 2000) and central nervous system activity (Lindquist et al., 2012; Phan et al., 2002; T. D. Wager, Phan, Liberzon, & Taylor, 2003). On the opposite side of the argument, others interpret empirical evidence to support emotion categories as natural kinds (Friedman, 2010; C. E. Izard, 2007), pointing to methodological and conceptual issues as reasons for observed variability in emotional responses.
A key criticism of the available evidence to date is that most empirical studies have used univariate statistical approaches to address this theoretical question. Univariate approaches test the relationship between a single dependent measure and experimental manipulations with one or more independent variables. In contrast, multivariate statistical approaches are able to jointly consider multiple dependent variables and experimental manipulations. Given that emotions engage a complex set of physiological components that likely make unique contributions to different affective states, univariate approaches are not optimal for identifying such interactions that are critical in characterizing emotions. Even when multiple dependent variables are measured and used to construct an analysis of variance (for example if a dependent variable is used to construct an additional factor to test an interaction with an independent variable), they are assumed to be independent. In multivariate approaches, the correlation between both independent and dependent variables is utilized in the analysis. By jointly considering multiple variables, multivariate approaches can reveal organization in data that is lost when response variables are treated independently or examined one at a time.

A series of investigations using multivariate pattern classification techniques shows promise for testing the autonomic specificity of distinct emotions. Similar to conventional univariate approaches, these studies have elicited distinct affective states
using a range of induction methods such as film (Christie & Friedman, 2004; Kreibig et al., 2007; Stephens et al., 2010), music (Nyklicek et al., 1997; Stephens et al., 2010), and autobiographical recall (Rainville et al., 2006) while autonomic measures such as cardiac, respiratory, and electrodermal activity were recorded. Patterns of physiological responses were then shown to differentiate the induced states by constructing pattern classifiers capable of using relationships between the multiple responses in order to infer which emotion had been elicited (for a review of pattern classification see Jain et al., 2000). Generally, these studies show that autonomic patterning differentiates distinct emotions, but not as accurately as self-report of the experience. While these findings demonstrate autonomic patterning of distinct emotions, they fail to determine whether information utilized by pattern classification is organized around multiple categories or more basic dimensions, limiting the conclusions which can be made about the structure of emotions.

The goal of the present study was to investigate the structure of emotion representations in subjective experience and physiological expression, by using pattern classification methods to decode emotional state. The term “decode” refers to the use of multivariate pattern classifiers to assign a class label to a set of dependent measures. Within the field of cognitive neuroscience, this approach has been widely used to infer the mental state of a participant from patterns of neural activity, termed “mind reading.”
or “decoding” (Norman et al., 2006). The approach used here is analogous, only the affective state of participants is predicted using patterns of self-report and autonomic responses. We adapted the method of emotion induction from Stephens et al. (2010) and used machine learning algorithms to label the experience of fear, anger, sadness, surprise, contentment, amusement and a neutral state. We employed a nonlinear machine learning algorithm -- a support vector machine using a Gaussian kernel-- as it is capable of detecting more subtle and complex patterns and may result in improved performance. We compared classification accuracy against chance levels to test the hypothesis that categorical responding occurs in peripheral autonomic systems and self-report. This method of characterizing emotions as natural kinds tests for the presence of projectable property clusters. More specifically, each emotion should have definitive features that co-occur and reliably observed for every instance of the category (L. F. Barrett, 2006a). Thus, the accuracy of a pattern classifier can test natural kind status by quantifying to what extent patterns of autonomic responses are unique and differentiate emotions.

To test the organization of emotion evidenced in self-report and peripheral autonomic expression, we compared the distribution of observed classification errors to those predicted by categorical versus dimensional models of emotion. This approach parallels the well-established use of confusion data in psychophysics studies of
perceptual categorization and recognition, where participants label stimuli and the
distribution of errors is used to characterize the mental representation of stimuli (e.g.
Loomis, 1982; Townsend, 1971). Equivalently, examining the structure of errors from a
pattern classifier will reflect how classes are represented by the input variables. If
emotions are organized categorically, errors should be randomly distributed and
unrelated to dimensions such as valence and arousal. Conversely if responses are not
specific to any emotion but map to general locations in affective space, classification
errors should increase with the proximity of stimuli along dimensions of arousal and
valence.

2.2 Methods

2.2.1 Participants

Twenty healthy volunteers (10 women, 10 men, 15 White, three Black, two Asian, Mage
= 23.5 years, age range: 19–36 years) gave written informed consent and participated in
the study. The study was approved in accordance with the Institutional Review Board at
Duke University. Participants were compensated either $10 per hour or with course
credit.

2.2.2 Materials and procedure

Standardized music and film clips were presented to elicit the discrete emotions
of fear, anger, sadness, surprise, contentment, amusement, and a neutral control
condition. Two emotion induction techniques were used to ensure the patterning of responses was not specific to the method of elicitation. The stimuli selected were the same as those used and validated in Stephens et al. (2010), with the exception of the film clips intended to induce contentment. Because these films lacked audio and were of generally low quality, new clips for the contentment category were created from commercially available videos of beach scenes (Powerfloe Network; Portola Valley, CA). In total, the stimuli consisted of two music clips and two films for each intended emotion category, yielding a total of 28 stimuli. Participants viewed stimuli in one of four pseudo-random counterbalanced orderings, which alternated between positive and negative valence. Preceding each emotion induction stimulus, participants were presented with a standardized washout video consisting of colored bars and patterned crosshairs to reduce carry-over between stimuli (Gross & Levenson, 1995). Immediately following each emotion induction, a 23-item affect self-report scale was administered (ASR; Christie & Friedman, 2004; Nyklicek et al., 1997; Stephens et al., 2010). This questionnaire contains items about both categorical and dimensional aspects of emotion experience, allowing comparisons between basic and dimensional models of emotion. Participants were instructed to “Select the number on the scale that best describes how you felt during the music/film clip that you just listened to/viewed. If the word does not at all describe how you felt during the clip, select ‘1’. If the word very accurately
describes how you felt, select ‘7’; for an intermediate amount, select ‘4’, etc.” Thus, the instructions emphasized self-experienced affect in response to the emotion induction rather than sensory-evaluative processes regarding the intended emotion based on physical properties of the stimuli. The specific items on the questionnaire include: content, amused, surprised, fearful, angry, sad, neutral, good, bad, positive, negative, calm, agitated, pleasant, unpleasant, passive, active, relaxed, excited, indifferent, intensity, enjoyment, and familiarity. Due to the length of this within-subjects procedure, films and music were presented on two separate days within the same week in a counterbalanced order. Each experimental session lasted approximately two hours. Responses were made by pressing the corresponding button on a keyboard. Presentation of stimuli and acquisition of behavioral responses was controlled using Cogent 2000 software (Wellcome Department of Imaging Neuroscience, http://www.vislab.ucl.ac.uk/cogent.php).

2.2.3 Psychophysiological recording and feature extraction

Psychophysiological data were acquired using a BIOPAC MP-150 data acquisition system and subsequently processed using AcqKnowledge software (BIOPAC Systems Inc., Goleta, CA) and custom in-house scripts (MATLAB 2010a, The MathWorks Inc., Natick, MA, 2010). Physiological activity was recorded as an analog signal and digitized at a frequency of 200 Hz. Finite sampling of any continuous measure can
introduce error in its reconstruction, and selecting a sufficiently high sampling rate is critical. While the selected sampling rate may contribute as a potential source of error in the computation of measures requiring temporal precision such as heart-rate variability (Merri, Farden, Mottley, & Titlebaum, 1990), it is sufficient for capturing the gestalt of peripheral changes occurring in different biological systems at varying timescales. Electrodermal, cardiac, respiratory, and gastric activity were targeted for recording to provide a broad sampling of autonomic responding. Electrodermal activity was acquired by placing two Ag-AgCl electrodes on the hypothenar eminence of the left palm and amplified using the GSR 100C module. Changes in cardiac activity were measured using a lead II electrocardiogram (ECG) configuration with electrodes placed on the right wrist and left ankle and amplified using the ECG 100C module. Respiratory activity was measured using a BIOPAC SS5LB respiratory effort transducer placed just below the sternum and amplified using the RSP 100C module. Gastric activity was recorded using a monopolar cutaneous electrode placement along the transpyloric plane and amplified using the EGG 100C module.

Digitized physiological data were processed on a trial-by-trial basis for each instance of emotion induction to extract features relevant to autonomic patterning. Electrodermal data were processed using custom MATLAB scripts to extract the tonic skin conductance level (SCL) and the skin conductance response rate (SRR). SCL was
obtained by calculating the mean level of skin conductance. SRR was derived by counting the number of individual skin conductance responses during the emotion induction, indicated by continuous increases in electrodermal activity with amplitude greater than .02 μS. ECG data were processed using heart rate variability software (Acharya, Joseph, Kannathal, Lim, & Suri, 2006; Brennan, Palaniswami, & Kamen, 2001) in the BIOSIG-toolbox (Schlogl & Brunner, 2008) to obtain heart rate (HR), heart period (HP), high frequency cardiac power (HF; .15 to .40 Hz), low frequency cardiac power (LF; .04 to .15 Hz), very low frequency cardiac power (VLF; 0 to .04 Hz), and total cardiac power (TP). AcqKnowledge software was used to decompose digitized respiratory data to measures of inhale time (IT), exhale time (ET), respiratory depth (RD), respiration rate (RR), respiratory sinus arrhythmia (RSA; the breath to breath difference between minimum and maximum R-R interval), peak inspiratory flow (PIF), and peak expiratory flow (PIF) for each trial. The relative power of the electrogastrogram (EGG) at a frequency of 3 cycles per minute was computed as an index of normogastria (N. A. Harrison, Gray, Gianaros, & Critchley, 2010). Given that these peripheral measures stem from common sources, considerable correlation between variables was expected. While this issue of multicollinearity is a common issue in univariate regression based analysis, linear pattern classifiers can pool evidence offered
by correlated inputs and nonlinear classifiers can utilize the presence or absence of interactions between variables in constructing a classification rule.

These 16 autonomic measures were adjusted to reflect relative change from the washout preceding the trial by subtracting the mean value of the 60 sec prior to the end of the emotion induction from the last 60 sec of the preceding washout. This averaging served as a temporal filter, reducing the impact of artifacts and ensuring a stable measure of peripheral responses. As the stimuli had variable duration and on average lasted 124 seconds, averaging during this window ensures the emotions have had sufficient time to emerge and offers an even sampling for all stimuli. Beyond this windowed averaging, no data were explicitly removed due to the presence of artifacts, as the preprocessing and analytical approach employed should minimize their impact. Prior to pattern classification analysis, autonomic measures were converted to a standardized z-score within subjects to remove variability related to overall levels of autonomic responsivity between subjects.

2.2.4 Statistical Analysis

The effectiveness of the emotion induction was assessed by performing an analysis of variance (ANOVA) using categorical items from the self-report scales. Post-hoc contrasts were constructed to compare the targeted emotional response against all
alternatives. This manipulation check was employed to ensure the experience of emotion occurred as intended.

The amount of emotion-related information present in autonomic responding and self-report was investigated using pattern classification analysis. In this type of analysis, a classifier is constructed which determines the categorical grouping of response variables, called features, by creating a decision rule. We constructed classifiers using either autonomic responses or self-report as input in an attempt to decode the categorical emotion label for each trial. Ground truth for classification was determined a priori based on previous work validating the stimuli (Gross & Levenson, 1995; Nyklicek et al., 1997). In this respect, the labeling of ground truth was independent of the measures used as input for pattern classification. Nevertheless, given that prior work has demonstrated categorical experience of these stimuli from single items of self-report, we predict pattern classifiers should be able to decode the emotional state of participants from self-report with high accuracy. As there were 20 participants and 28 instances of emotion induction, a total of 560 data points were available for training and testing the classifier. The number of possible features to use as input to the classifiers was 23 in the case of self-report and 16 for autonomic activity.

We selected a support vector machine algorithm, implemented in LIBSVM (Chang & Lin, 2011), utilizing both linear and radial basis functions (Boser, Guyon, &
Vapnik, 1992) to classify the data. As classification involved labeling seven classes, the default “one-against-one” multiclass algorithm of LIBSVM was used (Hsu & Lin, 2002). Kernel parameters were automatically tuned for the classifiers using gradient descent (Chapelle, Vapnik, Bousquet, & Mukherjee, 2002). To estimate classifier performance in a manner that generalizes between subjects, repeated random subsampling was performed. This form of validation involves randomly dividing the data into training and testing sets multiple times to estimate the error rate of a classifier. This approach is advantageous because the number of training and testing samples can be matched, which provides balance between model accuracy and low variance in estimating the classification error rate. Unlike other methods such as k-fold cross validation, the amount and proportion of training and testing data is not dependent on the number of folds. In this case, half of the subjects were chosen at random for training (the remainder used for testing) and the trials were stratified so an equal number of instances from each emotion category (30) was used for training and testing. The subsampling procedure was repeated 100 times, yielding predicted labels made by the classifiers which could be compared to the true categorical label of each instance. This approach is advantageous because using 210 instances for training and testing relative to a much smaller number of features aids against overfitting – a common problem when feature sizes are larger, such as in classification of genomics or neuroimaging data. Thus, the ratio between the
number of instances used and the number of features in training was kept large, improving the predictive power of the classifier. Further, the approach ensures that a large number of independent samples were used to test the classifier, critical factors to minimizing bias in the classification of small samples (Raudys & Jain, 1991).

To evaluate our first hypothesis that distinct affective states are represented in patterns of peripheral autonomic activity and self-report, we conducted the following statistical analyses. We calculated the accuracy of the classifier for each repetition by comparing the categories predicted by the classifier to the true categories in the testing set. Non-parametric statistical tests were performed to ensure robustness against violations of normal distribution assumptions. Chance performance was estimated by permuting emotion labels and running the random subsampling procedure on the same data. Statistical significance of performance was determined by entering the multiple repeated estimates of classifier accuracy using true labels against those with the random labels with a Wilcoxon signed rank test. This approach has been suggested to overcome violations of the assumptions of parametric tests created when making inference on cross validated data (Bouckaert & Frank, 2004; Demsar, 2006). By adopting a subsampling approach that tested across different subjects and induction methods, we aimed to minimize dependence between samples on factors other than the induced emotion – which could possibly bias estimates of classification accuracy. While
classification was run with support vector machines using both linear and Gaussian kernels, nonlinear classification outperformed simpler linear approaches using both self-report (10.9% improvement, z = 8.14, p < .001) and peripheral responses (27.1% improvement, z = 8.68, p < .001). To simplify the presentation of results, we report only results from nonlinear classification, given its superior performance.

To investigate the degree to which response patterning supports different theoretical organizations of emotion, we examined the distribution of errors produced by pattern classifiers. Using the true and predicted labels from classification, we constructed a confusion matrix to characterize the structure of performance on each repetition. The confusion matrix was then used to tally the number of errors made for the 21 possible pair-wise combinations of emotions that could constitute an error, for example mistaking fear and anger. The distribution of errors on each repetition was compared to the difference in subjective experience along discrete categories and dimensions of arousal and valence. Categorical items directly corresponded to the seven self-report measures for each emotion: ‘content’, ‘amused’, ‘surprised’, ‘fearful’, ‘angry’, ‘sad’, and ‘neutral’. Measures of valence were computed by averaging the values from the self-report ratings of ‘good’, ‘positive’, and ‘pleasant’ while subtracting the average of the values from the ‘bad’, ‘negative’, and ‘unpleasant’ ratings for each emotion category. Similarly, arousal indices were calculated by averaging ratings for the items
‘agitated’, ‘active’, and ‘excited’, while subtracting the average values of ‘calm’, ‘passive’, and ‘relaxed’. We calculated the Euclidean distance between each pair of emotion categories, yielding 21 measures of proximity of the induced emotions in an affective space characterized either by valence and arousal or discrete categories. Distances were calculated on test data of each repetition, ensuring independence from model construction. Correlations between these proximity values and the number of classification errors were computed on each repetition of the subsampling procedure to test whether information captured in pattern classification is driven by these affective dimensions.

| Self-Report of Experienced Emotion | Target vs. Other |

### 2.3 Results

#### 2.3.1 Manipulation check

Contrasts from the ANOVA of self-report variables revealed that the target emotions were successfully induced. Post hoc analysis of planned contrasts revealed that, on average, the experience of each category of stimuli was judged to be greater for the intended emotion than alternative categories (see Table 3).

Table 3: Validation of emotion induction
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>t(133)</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Contentment</td>
<td><strong>2.20</strong></td>
<td>0.62</td>
<td>-1.12</td>
<td>-1.38</td>
<td>-1.33</td>
<td>-0.72</td>
<td>1.54</td>
<td>8.46</td>
<td>1.47</td>
</tr>
<tr>
<td>2. Amusement</td>
<td>1.62</td>
<td><strong>2.93</strong></td>
<td>0.22</td>
<td>-1.31</td>
<td>-1.01</td>
<td>-1.15</td>
<td>-0.30</td>
<td>12.52</td>
<td>2.17</td>
</tr>
<tr>
<td>3. Surprise</td>
<td>-0.80</td>
<td>-0.17</td>
<td><strong>2.27</strong></td>
<td>0.67</td>
<td>-0.29</td>
<td>-0.39</td>
<td>0.32</td>
<td>9.10</td>
<td>1.58</td>
</tr>
<tr>
<td>4. Fear</td>
<td>-1.20</td>
<td>-0.65</td>
<td>0.03</td>
<td><strong>2.27</strong></td>
<td>0.87</td>
<td>0.44</td>
<td>-0.74</td>
<td>10.13</td>
<td>1.76</td>
</tr>
<tr>
<td>5. Anger</td>
<td>-1.05</td>
<td>-0.76</td>
<td>0.39</td>
<td>0.65</td>
<td><strong>1.51</strong></td>
<td>1.02</td>
<td>-1.12</td>
<td>6.40</td>
<td>1.11</td>
</tr>
<tr>
<td>6. Sadness</td>
<td>-0.29</td>
<td>-0.24</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.37</td>
<td><strong>1.47</strong></td>
<td>-0.78</td>
<td>5.71</td>
<td>0.99</td>
</tr>
<tr>
<td>7. Neutral</td>
<td>-0.48</td>
<td>-1.73</td>
<td>-1.74</td>
<td>-0.89</td>
<td>0.62</td>
<td>-0.67</td>
<td><strong>1.07</strong></td>
<td>3.58</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Note. Self-report of emotion conveys the standardized scores for categorical items, averaged across participants. Numbers in bold font indicate scores for the target emotion. The columns on the right indicate the $t$ statistic and effect size (Cohen’s $d$) for post-hoc contrasts of the target emotion versus all alternatives.

### 2.3.2 Decoding distinct emotions from response patterns

Feature extraction produced an array of 23 self-report and 16 autonomic activity measures (see Figure 3 for mean patterns of self-report and autonomic activity). The confusion matrices in Figure 4 depict the hits, misses, and false alarms of the classifiers, from which statistical measures of performance were computed.
Figure 3: Inputs for pattern classifiers averaged across subjects from peripheral physiology (left) and self-report (right).

Colorbars indicate average z-scores standardized within subjects. SCL = tonic skin conductance level; SRR = number of phasic skin conductance responses; HR = heart rate; HF = high frequency cardiac power; LF = low frequency cardiac power; VLF = very low frequency cardiac power; TP = total cardiac power; RSA = respiratory sinus arrhythmia; RR = respiration rate; RD = respiratory depth; EGG = three cycle per minute gastric activity; HP = heart period; PIF = peak inspiratory flow; PEF = peak expiratory flow; IT = inhale time; ET = exhale time.
Figure 4: Performance metrics for classification of emotions from peripheral physiology (left) and self-report (right).

Confusion matrices (top) illustrate the distribution of classification errors. Colorbars indicate the number of trials in each cell. The diagonal of a confusion matrix contains the number of hits, where the predicted emotion matches the target emotion. The number of misses for a given category is indicated by the sum of off-diagonal values for the corresponding column in the matrix. Similarly, the number of false alarms for a category is indicated by the sum of off-diagonal values in a given row. Note the increased clustering of errors for similar emotions in self-report versus peripheral physiology. As depicted in histograms of accuracy (bottom), overall classification performance greatly exceeds chance levels estimated by running the same classification on randomly permuted labels.

Classification using self-report measures yielded an overall accuracy of 88.2%, which was significantly greater than the expected chance rate of 14.3% (z = 8.68, p < .001). Classifiers using autonomic measures to decode emotional state exhibited similar results, with a classification rate of 58.0% (z = 8.68, p < .001). Post hoc tests comparing the
accuracy of classification on instances of film versus music induction did not reveal a significant difference ($z = 0.89, p = .374$), indicating that emotion-specific response patterns did not vary significantly as a function of induction method. These results suggest that configurations of peripheral physiology are diagnostic of affective states across participants, although to a lesser degree than experiential self-report ($z = 8.67, p < .001$).

Figure 5: Dimensional sampling from self-report of emotion.

Scatterplot shows the distribution of experienced emotion along dimensions of arousal and valence for all 560 trials. The location for trials along the valence axis reflects the average response of self-report items “good,” “positive,” and “pleasant” minus the scores for items “bad,” “negative,” and “unpleasant.” The location along the arousal axis is indicative of the mean value of items “agitated,” “active,” and “excited,” while subtracting the average value of responses for “calm,” “passive,” and “relaxed.”
Examining the relationship between the number of classification errors and location in a two-dimensional affective space revealed that valence and arousal contributed selectively to classification of self-report compared to physiological responses. Plotting the location of all 560 trials based on self-report of valence and arousal revealed a broad sampling of affective space, with considerable overlap between distinct emotions (see Figure 5).

Consistent with models proposing a dimensional organization of emotion, classification of self-report produced fewer errors when the experienced emotions were more distant in affective space \((r = -0.34, z = -8.50, p < .001; \text{see Figure 6})\). For example, errors between fear and anger \((M = 3.00, SD = 0.98)\) were more frequent than those between contentment and fear \((M = 0.31, SD = 0.28)\) as assessed by a Wilcoxon signed-ranks test \((z = 8.66, p < .001)\). In contrast, the structure of errors from classifying physiological responses was not correlated with distance along dimensions of valence and arousal \((r = 0.03, z = 1.19, p = .235)\). In this case, confusions between fear and anger \((M = 3.74, SD = 1.56)\) were not more frequent than those between contentment and fear \((M = 4.59, SD = 1.43, z = 0.23, p = .81)\). More broadly, errors between emotions differing in valence and arousal such as anger and contentment \((M = 3.48, SD = 1.66)\) or sadness and amusement \((M = 4.25, SD = 1.52)\) did not produce fewer errors on average than those between fear and anger \((z = -0.78, p = .433)\). These findings indicate that peripheral
Figure 6: Relationship between classification errors and experienced emotion along dimensions of valence and arousal (left) and seven discrete categories (right).

Scatterplots (top) depict the relationship between the number of classification errors made as a function of the distance between experienced emotion in an affective space with dimensions corresponding to either seven categories or valence and arousal (averaged across 100 repetitions). Each point reflects one of 21 possible pairwise combinations between emotions. In the dimensional configuration, for instance, fear versus contentment has a large value because the two emotions differ in both valence and arousal. The distance between emotions differs when computed on the basis of categorical items, as each emotion is predominantly mapped to a single measure (see Table 1). Histograms (bottom) show the full distribution of correlation coefficients computed on all repetitions. In the dimensional configuration of affective space, classification using self-report produced fewer errors as the distance between emotions increased, whereas the performance of classifiers using peripheral autonomic measures did not exhibit this relationship. Alternatively, when a categorical conception of affective space was considered, classification errors using both self-report and peripheral measures were significantly fewer as the distance between emotions increased.
autonomic responses are specific to distinct emotions irrespective of an underlying structure based on dimensions of valence and arousal.

Testing the relationship between the similarity of experienced emotion on the basis of categorical self-report items and classification errors indicated that both experiential and physiological responding were organized in a categorical manner. Classification errors were fewer when the Euclidean distance increased in an affective space constructed from seven distinct categories for both self-report ($r = -0.52$, $z = -8.68$, $p < .001$) and peripheral physiological responses ($r = -0.17$, $z = -1.19$, $p < .001$). Further, direct comparisons revealed both self-report ($z = -8.51$, $p < .001$) and peripheral autonomic measures ($z = -7.93$, $p < .001$) exhibited a closer correspondence to categorical than dimensional arrangement of affective space. Together, the present results demonstrate that information in peripheral physiological signals capable of discriminating between distinct emotional states primarily reflects categorical aspects of emotional experience, whereas self-reports incorporate aspects of both categorical and dimensional structure.

2.4 Discussion

Characterizing the organization of affective representations in behavioral, psychophysiological and neural metrics of emotion is a central goal of affective neuroscience. Recent advances in applying multivariate pattern classification
approaches to such complex data sets holds much promise to elucidate the nature of affective space and to advance an understanding of how relevant information for discriminating emotions is decoded. Consistent with Stephens et al. (2010), we found that music and film could be used to elicit specific emotions and that physiological patterning during these experiences could be used to infer the emotion induced. Beyond replicating previous findings, the present study provides novel evidence that emotion specific information contained in patterns of autonomic activity is not structured around dimensions of valence and arousal. These findings advance knowledge of the structure of physiological responding and self-report in emotion and point to both commonalities and differences between predominant theoretical models.

While the replication of previous work was notable, there were some methodological and analytical differences worthy of mention. The autonomic measures we recorded differed slightly from those used by Stephens et al. (2010). We did not employ impedance cardiography or blood pressure monitors, and we included the use of an electrogastrogram. While this reduced set of input features potentially limited the capability of classifiers to decode emotional state, the replication of significant classifier performance suggests the sampling of autonomic activity was sufficiently broad for this task. The success of autonomic classification was likely due in part from sampling both
branches of the autonomic nervous system simultaneously, which has been identified as a critical aspect of studying autonomic response patterning (Cacioppo et al., 2000).

In addition to these methodological differences, we employed more advanced machine learning algorithms than Stephens et al. (2010). By applying principles and methods commonly used in the emerging field of affective computing (Picard, Vyzas, & Healey, 2001) as suggested by Kolodyazhniy et al. (2011), we improved on the performance of classifiers. The use of nonlinear algorithms improved classification rates beyond those of linear classifiers using both autonomic and self-report measures. Peripheral information signifying distinct emotions could be summed across multiple autonomic systems using linear methods, but also the covariation between multiple autonomic systems was further captured by nonlinear relationships. In this respect the information conveyed by the autonomic nervous system is more than the sum of its parts. Nonlinear classification algorithms are particularly well suited to decoding distinct affective states using psychophysiological approaches for several reasons. Firstly, nonlinear classifiers are capable of utilizing information that is carried in the interaction of variables, which can be entirely missed using linear methods. Second, the number of features employed (16) was relatively small compared to the number of instances or data points used for training and testing (210). Finally, decoding seven classes of stimuli is a complex task. Increasing the dimensionality of the data using a
nonlinear kernel aids in separating classes by increasing the sparsity of the data in multidimensional space. Thus, while prior pattern classification efforts have provided empirical support for categorical models, advances in machine learning techniques show promise in improving the detection of affective state from autonomic activity.

In addition to showing that autonomic measures are diagnostic of the affective state of an individual, we verified that information used in pattern classification was independent of affective dimensions of valence and arousal. This finding supports emotion theories which suggest information carried in autonomic responses are capable of differentiating distinct emotions, namely basic (Ekman, 1992; C. E. Izard, 1992) and appraisal theories (Scherer, 1984) of emotion. These results stand in contrast to meta-analysis of studies relying on univariate methods (Cacioppo et al., 2000) which suggest there is little consistency in autonomic responding for specific emotions, as well as theories suggesting that physiological responses are organized along valence and arousal from which the experience of discrete emotions are constructed (L. F. Barrett, 2006b). It is important to note that this result is not contrary to the common finding that specific peripheral physiological responses (e.g. electrodermal activity, heart rate, or facial expression) are related to the experience of arousal or valence (M. M. Bradley & Lang, 2000). Rather, it demonstrates that the information carried in autonomic signals separate from these dimensions is specific to distinct affective states.
While we found that autonomic and experiential measures of affective state do not share the same structure, this does not necessitate that a peripheral, autonomic representation of emotion cannot be acted upon by cognitive processes to yield a more distinct conscious experience. This transformation between peripheral physiology and self-report is evident through development (Posner, Russell, & Peterson, 2005), whereby children have difficulty differentiating between distinct emotions but can identify “happy” or “sad” faces (Bullock & Russell, 1984). This difference is also evident in individuals with alexithymia (Sifneos, 1972), who have difficulty reporting their affective state, if only using basic words to describe their feelings. In this regard, cognitive processes engendered through development may be responsible for shaping a peripheral representation of emotion into one that is experienced consciously. Our findings suggest that peripheral autonomic activity contains information capable of representing discrete emotions, and the readout of this information may be partially reorganized, leading to somewhat different manifestations of emotion in peripheral autonomic signals and self-report.

We have demonstrated the use of multivariate pattern classifiers to test the ability of different models to organize distinct emotion states. This novel approach offers affective neuroscience a framework for comparing and validating models of emotion in different response systems. To more fully capture the organizational properties of
emotional representation, future assessment of response patterning should directly compare a wider array of models, for example those based on cognitive appraisal theory (Scherer, 1984), include other measures of emotional responding (such as neural indices), and use additional induction methods such as autobiographical recall or visual imagery. Additionally, as some classification parameters were optimized within the present sample, it will be important to test the accuracy of classification models on separate generalization sets in the future. Finally, it would be important to compare explicit emotion labeling with free viewing or listening of stimuli. Testing the structure proposed by competing models highlights the commonalities and differences between them, hopefully enabling advances in emotion theory to be made. In turn, this endeavor should elucidate the core properties of emotion organization, be they conceptualized as dimensions, clusters, or a series of events.
3. Multivariate neural biomarkers of emotional states are categorically distinct

This chapter describes a functional magnetic resonance imaging (fMRI) experiment examining whether information conveyed in central nervous system activation is represented along categorical or dimensional models of emotion, utilizing the paradigm employed in Chapter 2. Cinematic film clips and instrumental music were used to induce emotional states of anger, fear, sadness, surprise, amusement, contentment, in addition to a neutral state (P. A. Kragel & LaBar, 2013). Patterns of blood-oxygen-level dependent (BOLD) response to these stimuli were acquired using fMRI. On-line sampling of dimensional and categorical aspects of emotional experience was performed for the purpose of model comparison. Examining precisely how classification errors are made provides insight as to how information is represented in distributed patterns of brain activation. The findings presented in this chapter have been published as a research article in the journal Social Cognitive & Affective Neuroscience (P. A. Kragel & LaBar, 2015).

3.1 Introduction

It has long been hypothesized that emotions involve coordinated physiological, psychological, and behavioral changes that support survival (Robert Plutchik & Kellerman, 1980). Emotion directs attention, modulates memory, biases perception, and guides decision making. In humans, these changes are often concomitant with emotional
feelings that inherently reflect their utility in maintaining homeostasis (A. Damasio & Carvalho, 2013a) and are shaped both by biological and social influences (L. F. Barrett, 2012). Accordingly, emotional states are multifaceted phenomena that are supported through the integration of information pertaining to the environment, homeostatic state, and relevant adjustments in behavior.

There is contention regarding how emotions are represented within the brain: either as discrete categories or points in a continuous dimensional space (L. F. Barrett & Wager, 2006; S. Hamann, 2012; Lindquist et al., 2012). While both types of models posit that emotional states result from processes mediated by the central nervous system, they characterize their emergence via different mechanisms. Dimensional models propose that emotional states differ fundamentally along a small number of factors, e.g. pleasantness and activation (Feldman Barrett & Russell, 1999; Russell, 1980), although alternative interpretations and more complex models exist (Fontaine, Scherer, Roesch, & Ellsworth, 2007; Mehrabian, 1996; Watson & Tellegen, 1985). Differences in these core dimensions, along with past experience and contextual factors, are thought to contribute to the construction of distinct emotional states (L. F. Barrett, 2006b; Lindquist, 2013). Functional magnetic resonance imaging (fMRI) studies have identified putative neural correlates of some of these dimensions, implicating medial prefrontal structures in the processing of valence and the amygdala in mediating arousal or emotional salience (A.
K. Anderson et al., 2003; S. B. Hamann, Ely, Hoffman, & Kilts, 2002; Lewis, Critchley, Rotshstein, & Dolan, 2007; Wilson-Mendenhall, Barrett, & Barsalou, 2013; Winston, Gottfried, Kilner, & Dolan, 2005). Categorical emotion models (Ekman & Cordaro, 2011; Carroll E. Izard, 1971; Johnson-Laird & Oatley, 1992; J. Panksepp, 1982) alternatively hypothesize that a finite number of independent emotions, often thought to derive from survival-related functions, are sufficient in characterizing the structure of emotion. Meta-analyses suggest discrete emotional experiences can be discriminated on the basis of neural activation in humans (Murphy, Nimmo-Smith, & Lawrence, 2003; Vytal & Hamann, 2010), although neural substrates specific to any given emotion remain elusive (Lindquist et al., 2012).

Given the lack of consensus from univariate studies regarding the representation of emotion in the brain, multivariate methods have been adopted in attempts to identify more specific mappings between patterns of neural activation and a variety of affective states (for a review, see Philip A. Kragel & LaBar, 2014). Brain states (measured with functional magnetic resonance imaging, fMRI) have been classified along both dimensional (e.g., Baucom et al., 2012; Rolls et al., 2009) and categorical (e.g., Kassam et al., 2013; Sitaram et al., 2011) models of emotion. While such research has demonstrated the capacity of multivariate methods to isolate patterns of neural activity associated with different affective states, several factors limit the conclusions that can be drawn from
prior research. As prior studies have focused on either on categorical or a dimensional framework for classification, evidence favoring either model is difficult to ascertain. Most studies only utilize a single stimulus modality (e.g. visual images, mental imagery, facial expressions or vocalizations), which limits the generalizability of results. Assessment of classification models in the literature thus far tends to focus on predictive capacity and fails to characterize the localization and spatial characteristics of emotion-predictive patterns (i.e. does each emotion have a unique substrate? Are they organized modularly or are they more widely distributed?). Further, studies commonly sample a limited number of emotions, which fail to adequately sample the dimensions of valence and arousal. Finally, and perhaps most importantly, no prior fMRI pattern classification research has examined on-line measures of subjective experience. This is a critical issue, as it is essential to determine whether specific emotions were elicited by the stimuli presented, and whether the information decoded by the pattern classifiers contributes to the experience of emotions as distinct entities or as points along broader affective dimensions.

In the present work, we tested how multi-voxel patterns of blood-oxygen level-dependent (BOLD) response predict discrete emotional states and whether the characteristics of these patterns better conform to categorical or dimensional models of emotion. Because successful classification of emotional states could equally support
either model (as both models commonly predict that emotional states are experienced
discretely), we examined the manner in which emotions were misclassified to test
whether emotions are better characterized along valence and arousal dimensions or as
independent emotion categories. To induce discrete emotions, we presented participants
cinematic film clips and instrumental music immediately followed by a period of online
self-report to sample categorical and dimensional aspects of emotion experience during
scanning. We then conducted multivariate pattern classification followed by Bayesian
model comparison to assess the discriminability and organization of emotional states in
the brain.

3.2 Methods

3.2.1 Participants

Thirty-two healthy, right-handed individuals (Mage = 26 years, age range = 19 to 38
years, 13 males) completed the study. Two additional participants were run (one male,
one female), but technical issues precluded analysis of their data. In particular, one
participant moved excessively during scanning (exceeding 5 mm of movement within a
single run), and technical issues with behavioral response hardware prevented
recording of self-report in one individual. The proportion of males and females was not
found to significantly differ in the final sample ($\chi^2 = 1.125, p = 0.289$). All participants
provided written informed consent to participate in accordance with the Duke
University Institutional Review Board. All participants received $20 per hour as monetary compensation for completion of the study.

3.2.2 Experimental Paradigm

During scanning, participants engaged in an emotion induction procedure (P. A. Kragel & LaBar, 2013; Stephens et al., 2010) adapted from standardized instrumental music (Nyklicek et al., 1997) and cinematic film clips (Gross & Levenson, 1995). Inducing emotions with two stimulus modalities permits generalization beyond stimulus features and enables classification across stimulus types. While both types of stimuli contained auditory information, the nature of this information differed as the music clips did not contain speech. In this task, participants were presented with either a film or music clip (with a mean duration of 2.2 minutes), immediately followed by a 23-item affect self-report scale (Nyklicek et al., 1997; Stephens et al., 2010) lasting 1.9 minutes followed by a 1.5 minute washout clip to minimize overlapping responses. On-line measurement of self-report, which has been shown to minimally alter fMRI responses to emotionally engaging stimuli (Hutcherson et al., 2005), was critical in quantifying dimensional and categorical aspects of emotional experience, as post-scan assessments rely on memory judgments which may inaccurately reflect moment-to-moment fluctuations in affect.

The stimuli used were the same as those in Kragel and LaBar (2013), which have been shown to adequately sample affective space to evaluate dimensional models and to
produce discrete experience of contentment, amusement, surprise, fear, anger, sadness, and a neutral state (included to serve as an experimental control; see Figure 7). Given that prior validation using this procedure reliably induced discrete emotional experience and differentiated autonomic responding (P. A. Kragel & LaBar, 2013; Stephens et al., 2010), these emotions can be considered as a priori ground truth for classification. These emotions were chosen because they vary in both valence and arousal, whereas using the standard basic emotions (i.e. happiness, surprise, fear, anger, disgust, sadness and neutral) would bias sampling towards negative emotions with high arousal. Two music and two film clips were used to induce each target emotion, resulting in a total of 28 unique stimuli – each presented a single time. Participants viewed stimuli in one of four pseudorandom counterbalanced orderings, which alternated between positive/neutral and negative valence to minimize the persistence of mood across trials. Preceding each emotion induction stimulus, participants were presented with a standardized washout video consisting of colored bars and patterned crosshairs to reduce carry-over between stimuli (Gross & Levenson, 1995). The self-report phase included items querying both categorical and dimensional aspects of emotion experience, facilitating comparisons between discrete and dimensional emotion models. The item labels on the questionnaire consisted of the following: content, amused, surprised, fearful, angry, sad, neutral, good, bad, positive, negative, calm, agitated, pleasant, unpleasant, passive, active, relaxed,
Scatterplot shows the distribution of self-reported emotion along dimensions of arousal and valence (standardized within subjects). The x-axis is the average of valence-related self-report items (‘good’, ‘positive’, and ‘pleasant’ minus the scores for ‘bad’, ‘negative’, and ‘unpleasant’) and the y axis is the average of arousal-related items (‘agitated’, ‘active’, and ‘excited’ minus ‘calm’, ‘passive’, and ‘relaxed’). Marginal histograms depict distributions of valence and arousal for all emotions. The distributions generally conform to dimensional models of emotion: contentment and amusement oppose fear, anger, and sadness along the valence dimension whereas contentment and sadness oppose fear, anger, surprise, and amusement in terms of arousal.

excited, indifferent, intensity, enjoyment, and familiarity. By sampling a number of emotions and affective dimensions on every trial, the self-report procedure minimizes conceptual biases towards one characterization of emotion over the other. Participants
were instructed to “Select the number on the scale that best describes how you felt
during the music/film clip that you just listened to/viewed. If the word does not at all
describe how you felt during the clip, select ‘1.’ If the word very accurately describes
how you felt, select ‘4’ and so forth.” Thus, the instructions emphasized experienced
affect during the induction rather than sensory evaluative processes regarding the
intended emotion based on physical properties of the stimuli.

Each experimental session consisted of seven runs of data acquisition, each
comprised of two trials, lasting a total of approximately 12 min. The total duration of the
scanning session was accordingly 2 hrs. Films and music were presented on two
separate days within a two week period in a counterbalanced order due to the length of
this within-subjects procedure, to minimize the amount of instructions, and to provide
seamless transitions between trials.

Presentation of stimuli and acquisition of behavioral responses was controlled
using Cogent 2000 software (Wellcome Department of Imaging Neuroscience,
http://www.vislab.ucl.ac.uk/cogent.php). Participants viewed stimuli on mirrors aligned
with a LCD screen upon which images were projected from a stimulus control
computer. Audio stimulation was presented using MR-compatible headphones
(Resonance Technology, Northridge California). Responses were made by pressing the
corresponding key on a fiber optic response box connected to the stimulus control
computer using a universal serial bus interface (Current Designs, Philadelphia Pennsylvania).

### 3.2.3 Image acquisition

Scanning was performed on a 3 Tesla General Electric MR 750 system with 50-mT/m gradients and an eight channel head coil for parallel imaging (General Electric, Waukesha, Wisconsin, USA). High-resolution images were acquired using a 3D fast SPGR BRAVO pulse sequence: repetition time (TR) = 7.58 ms; echo time (TE) = 2.936 ms; image matrix = 256\(^2\); \(\alpha = \) 12\(^\circ\); voxel size = 1 × 1 × 1 mm; 206 contiguous slices) for coregistration with the functional data. These structural images were aligned in the near-axial plane defined by the anterior and posterior commissures. Whole-brain functional images were acquired using a spiral-in pulse sequence with sensitivity encoding along the axial plane (TR = 2000 ms; TE = 30 ms; image matrix = 64 x 128; \(\alpha = \) 70\(^\circ\); voxel size = 3.8 x 3.8 x 3.8 mm; 34 contiguous slices). The first five images of each run were excluded from analyses to ensure the magnet had reached steady state.

### 3.2.4 Preprocessing and estimating neural activation

Processing of MR data was performed using Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience). Functional images were spatially realigned to correct for motion artifacts (Friston et al., 1995), coregistered to high resolution anatomical scans (Collignon et al., 1995), normalized to Montreal...
Neurologic Institute (MNI) space using high-dimensional warping implemented in the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm.html). In addition to normalizing anatomical and functional data, this routine provided intensity maps of gray matter, white matter, and cerebral spinal fluid for each participant. Because of the relatively long duration of stimuli, conventional high pass filtering was not performed during preprocessing data, as it would attenuate task-related signal of interest.

Whole-brain patterns of neural activation were estimated using the general linear model approach implemented in SPM8. For each subject, BOLD responses were modeled by convolving box-car functions with a canonical hemodynamic response function separately for the two induction trials in all 14 runs. Two additional regressors modelling the washout phases and self-report trials were included in each run. To model nuisance effects, six motion parameters (roll, pitch, yaw, in addition to translation (in x, y, and z dimensions) and session constants, were incorporated into the model. Following voxel-wise model estimation, parameter estimates for induction trials (28 in total) were inclusively masked with a gray matter mask constructed from the average modulated gray matter output from VBM8. No smoothing of data was performed. This procedure produced an 896 trial by 19,962 voxel pattern matrix for classification after concatenating across all 32 participants.
3.2.5 Assessment of self-report

To validate that the emotion induction procedure elicited experiences that were both discrete, yet subjectively varied in terms of valence and arousal, planned t- and F-tests were conducted on self-report data. Tests for discreteness were conducted by comparing the mean of the self-report item corresponding to the targeted induction against all other categories using a paired t-test (e.g. the mean rating of ‘amused’ was compared against the mean of the other six categorical items in response to the four amusing stimuli). Separate F-tests were performed for valence and arousal in one-way repeated measures ANOVAs with one measure per subject for each emotion category (seven in total). Statistical thresholds were corrected for multiple comparisons by using false discovery rate correction (FDR; Benjamini & Hochberg, 1995).

3.2.6 Pattern classification

Decoding of emotional states was performed via partial least squares Discriminant Analysis (PLS-DA; Wold, Sjostrom, & Eriksson, 2001) using the NIPALS algorithm (Martens & Næs, 1989) as implemented in the libPLS toolbox (http://www.libpls.net). This approach identifies a linear combination of features (in this case voxels) which maximally discriminate between classes. By effectively reducing the dimensionality of input features in classification, this algorithm is particularly well
suited in the present study as the number of features exceeds the number of observations.

Classification of multiple categories was performed using a winner-takes-all approach, wherein one class is classified against all others. Because this approach creates an uneven proportion of classes (1:6), a weighted approach was taken for discriminant analysis to minimize bias due to class imbalance. Input data (896 total instances, 28 from each participant) were mean centered before conducting the analysis. As the emotion induction procedure implemented in this study involved only four instances of each category per subject, classification was performed across rather than within subjects. We estimated classification error using 8-fold subject independent cross validation, wherein data from 28 subjects were used to build classification models and data from 4 subjects were used to test performance until all subjects have been used for testing. To protect against overfitting, our primary results used 5-fold cross validation on the training data to select number of latent variables in each model. To determine the information content of patterns, cross-validated measures of signal detection (sensitivity, specificity, and the area under the receiver operating characteristic curve) were averaged across all categories. To make statistical inference on these signal detection measures and estimate 95% confidence intervals, 10 iterations of cross-validation were performed to estimate the standard deviation of each metric. Wilson score centers and intervals were calculated.
to estimate confidence intervals for accuracy across all categories (Wilson, 1927). To test the robustness of our results, we additionally performed standard 5-fold cross validation while fixing the number of latent variables between one and three or using an inner 5-fold cross validation for model selection. Models using three latent variables were found to minimize generalization error, although using a smaller number produced accuracy above chance levels (see Appendix A.1).

The significance of PLS regression coefficients, which indicate the importance of each voxel in predicting emotions, was determined by bootstrap resampling of the full dataset using 10,000 replicates and three latent variables. This large number of iterations produced fine-grained estimates of confidence intervals and z statistics. Statistical maps were thresholded using a voxel-wise threshold of $P < .05$ and extent of 14 voxels, which were shown to control the false-positive rate of $\alpha < .05$ in Monte Carlo simulations (Forman et al., 1995). The resultant maps thus indicate in which regions activation reliably predicts a particular emotion. To verify that these maps reflected the influence of a voxel towards classification rather than the intensity of BOLD activation, we constructed scatterplots of PLS regression coefficients against the intensity range (maximum minus minimum) of parameter estimates used as input for classification, which showed minimal bias due to the scaling of the data (see Appendix A.2-3). To confirm that these maps were not driven by modality-specific information, z maps were
created on film and music trials separately and combined using Stouffer’s method. These maps were highly correlated with those generated using the full dataset (all $r > .79$), indicating that bootstrapping across all trials identified regions which were generally modality-independent.

The extent of overlap between thresholded maps was determined by calculating the proportion of voxels common to each unique pair of emotions relative to the total number of voxels above threshold. The average percent overlap across all 21 unique pairs was compared using a one-sample t-test against the average overlap created by 10,000 iterations of randomly shuffling the location of above threshold voxels and computing the same metric of relative overlap, which produced a value of 1.78%.

To test the relationship between classification errors and the subjective experience of emotion, separate generalized linear models were created for each emotion category using ratings of categorical items (e.g. ‘content’, ‘amused’, ‘surprised’, etc.) as independent variables and the number of predictions for each category of emotion induction as the dependent variable. As the number of predictions constitutes count data, a binomial link function was used. Statistical significance was conducted using t-tests on the single parameter in each model, correcting for multiple comparisons using FDR correction.
3.2.5 Model comparison

The organization of information in neural activation patterns was examined relative to dimensional and categorical models of emotion. As the categorical model assumes an independent axis for each emotion, the Euclidean distance between two emotions \((i, j)\) was calculated as:

$$\text{Dcat}_{i,j} = \sqrt{\sum_{e=1}^{7} (x_{e,i} - x_{e,j})^2}$$

Where \(x_{e,i}\) reflects the average of self-report item \(e\) for instances of emotion category \(i\). The distance between two emotions in an affective space of valence and arousal dimensions was defined as:

$$\text{Ddim}_{i,j} = \sqrt{(x_{\text{val},i} - x_{\text{val},j})^2 + (x_{\text{arous},i} - x_{\text{arous},j})^2}$$

Where \(x_{\text{val},i}\) reflects the average of valence-related self-report items (‘good’, ‘positive’, and ‘pleasant’ minus the scores for ‘bad’, ‘negative’, and ‘unpleasant’) across instances of emotion category \(i\) and \(x_{\text{arous},i}\) reflects the average of arousal related self-report items (‘agitated’, ‘active’, and ‘excited’ minus ‘calm’, ‘passive’, and ‘relaxed’) across instances of emotion category \(i\). Distances were computed for all 21 unique pairwise combinations of emotions for each model.

Models were fitted using distances as independent variables in a Poisson regression, as the total number of errors in classification are count data and can be
modeled as a linear combination of categorical and dimensional distance. Four models were constructed: one including only a constant term, one with a constant and categorical term, one with a constant and dimensional term, and one with constant, categorical, and dimensional terms. Including both dimensional and categorical terms in the combined model allows the regression to assess unique variance for each term, serving to highlight differences between the two accounts.

Bayesian model comparison was performed to determine which model was most likely to produce the distribution of errors resulting from classification of neural data. The Bayesian information criterion (BIC) score was computed for each model (which includes a penalty term for additional parameters), and Bayes factors (\( B \)) were subsequently estimated using the Schwarz approximation (Schwarz, 1978). We additionally converted raw BIC scores to BIC weights to visualize the relative evidence for each model (Buckland, Burnham, & Augustin, 1997). To determine the relationship between classification errors and distance between instances in each model space, t-tests were performed on parameter estimates in the favored model. Because dimensional and categorical distances were correlated (\( r = .778, P < .0001 \)), multicollinearity between regressors could drive changes in the direction of observed effects. To rule out this possibility, we additionally inspected the parameter estimates in models with single terms to ensure the direction of effects did not change when including both terms.
3.2.7 Univariate models

To examine brain activity that varied as a function of self-reported valence and arousal, a group-level univariate ANOVA with factors of emotion (seven levels) and stimulus modality (2 levels) was performed. Contrasts for each emotion (versus implicit baseline) were separately created for film and music trials at the subject level and used as input. Orthogonalized contrasts corresponding to the group average self-report scores for valence and arousal were performed at each gray matter voxel. To assess the similarity of PLS regression coefficients and univariate effects, separate contrasts for each emotion versus all others were performed. Because multiple inputs were used from the same subjects, independence was not assumed for emotion and modality factors. A correction for non-sphericity was applied for to account for potential differences in variance across levels of emotion and modality. Statistical maps were thresholded using a voxel-wise threshold of $P < .05$ and extent of 85 voxels, which were shown to control the false-positive rate of $\alpha < .05$ in Monte Carlo simulations.

3.3 Results

3.3.1 Decoding emotional states from BOLD activation patterns

Before weighing the evidence of dimensional and categorical models, we first needed to demonstrate that discrete emotional states could be predicted on the basis of brain activation. To this end, we conducted pattern classification using single-trial
estimates of neural activation from fMRI data acquired during the presentation of film and music clips. Utilizing Partial Least Squares Discriminant Analysis (PLS-DA; Wold et al., 2001) with 8-fold subject independent cross-validation, we classified seven emotional states (contentment, amusement, surprise, fear, anger, sadness, and a neutral state) with 37.3% accuracy (where chance is 14.3%, $P < .001$, 95% CI [34.2% 40.5%]) (Figure 8a). Pair-wise classification performance was above chance levels for sensitivity ($M = .736, P < .001$, 95% CI [.655 .816]), specificity ($M = .560, P = .034$, 95% CI [.506 .614]), and area under the receiver operating characteristic curve ($M = .652, P < .001$, 95% CI [.610 .694]). Beyond accuracy, which is not best suited to characterize classification performance as a single number (A. P. Bradley, 1997), these measures show reliable detection of emotion-related information. High levels of sensitivity compared to specificity indicate that the classifier is better at identifying the presence of an emotion rather than the absence of an emotion, indicating that classification was not biased toward the more frequent negative class. Further, these performance metrics did not differ between film and music inductions (all $P$’s > .204), suggesting that learning was not biased towards information in one modality.
Figure 8: Performance measures of multi-voxel pattern classification.

(A) Confusion matrix for classifying all seven emotional states (top) and receiver operating characteristic (ROC) curves for one-vs-all classifications (bottom). Increasing classification performance would yield a confusion matrix with more predictions falling along the diagonal and push ROC curves towards the upper left corner, away from chance. (B) Binomial regression predicting the number of instances assigned to each emotion category based on the self-reported experience of the targeted emotion. The x-axis reflects standardized scores of self-report for the target emotion and the y-axis indicates the number of trials labeled as the target emotion. Note the curves approach 18 predictions (near random assignment of 128 trials into 7 classes) as self-report scores near zero.

To explicitly test how well the neural classification models generalized across stimulus modalities, we performed cross-classification between responses to music and films. This procedure involved training a classification model on patterns evoked using film stimuli and testing on the responses to music clips. Next, music clips were used for training models and film stimuli were used to test the performance of classification. Cross-classification accuracy of all seven emotional states was 28.38% (binomial test against chance of 14.3%, $P < .001$), indicating that a substantial amount of pattern-information generalized across the stimulus modality used to induce emotions.
Beyond testing performance measures for stimulus decoding, we examined whether the predictions from pattern classification tracked the subjective experience of the emotion. Binomial regression was performed to assess whether the average rating for each categorical self-report item obtained during scanning was associated with the number of predictions made during classification across all seven emotion inductions (Figure 8b). This analysis revealed the extent to which participants felt a given emotion followed the number of times that emotion was predicted during classification: ‘content’ ($\beta = .990$, $t = 4.13$, $P_{adj} < .0001$, 95% CI [.519 1.46]), ‘amused’ ($\beta = 1.49$, $t = 9.03$, $P_{adj} < .0001$, 95% CI [1.17 1.82]), ‘surprised’ ($\beta = 1.49$, $t = 5.40$, $P_{adj} < .0001$, 95% CI [.950 2.03]), ‘fearful’ ($\beta = 1.24$, $t = 5.74$, $P_{adj} < .0001$, 95% CI [.817 1.66]), ‘angry’ ($\beta = 1.72$, $t = 8.04$, $P_{adj} < .0001$, 95% CI [1.30 2.14]), ‘sad’ ($\beta = 0.79$, $t = 3.29$, $P_{adj} = .00099$, 95% CI [.321 1.26]), and ‘neutral’ ($\beta = 4.20$, $t = 9.01$, $P_{adj} < .0001$, 95% CI [3.29 5.12]). The average coefficient of determination ($R^2$, computed as 1 minus the ratio of the residual sum of squares and the sum squared deviation in the dependent measure) for these regression models was .57 (min: .21 for surprise, max: .94 for anger), indicating that a considerable portion of the variance in the number of predictions made for different emotions could be explained by subjective ratings of the target emotion.

Given that pattern classification identified emotion-specific information in the fMRI data, we next identified the voxels in which increased levels of activation reliably
predicted each emotion. We found that PLS-DA regression coefficients, which indicate
the contribution of each voxel in predicting the classification outcome, were reliably
greater than zero in a number of clusters distributed throughout the brain (Figure 9; see
Appendix A.4-10). To provide approximate localization, we computed the percent
overlap of each predictive pattern with all regions from a macroscopic atlas of brain
anatomy (Tzourio-Mazoyer et al., 2002). Activation in a distributed set of clusters
predicted states of contentment, including dorsal precuneus, bilateral postcentral gyrus,
and mid-cingulate gyrus. A more focal set of regions informed the classification of
amusement: bilateral superior- and middle-temporal lobe, bilateral early visual cortex,
and bilateral supplementary motor area, among other frontal regions. Several clusters of
activation led to labelling instances as surprise, including much of mid- and anterior-
cingulate gyrus, bilateral anterior insula, bilateral mid-occipital cortex, bilateral
thalamus, and bilateral cerebellum. Predictions of fear were derived from activation
within bilateral lingual gyrus, bilateral fusiform gyrus, dorsal/anterior precuneus, and a
number of bilateral medial temporal lobe structures: amygdala, hippocampus, and
parahippocampal cortex. Activation within bilateral superior temporal gyrus, ventral
precuneus, and right angular gyrus primarily informed predictions of anger. Patterns of
activity including left cerebellum, bilateral superior temporal gyrus, and bilateral
temporal pole led to sad emotional states. Bilateral activation in angular gyrus,
supramarginal gyrus, postcentral gyrus, and lingual gyrus contributed to classification of neutral states.

Overall, we found the patterns to be distinct yet partially overlapping at this macro-scale (Figure 10), as activation within many of the same structures contributed to the prediction of several emotions. Given dimensional and categorical models predict differing degrees of differentiation in the neural substrates underlying emotions, we quantified the spatial overlap of emotion-predictive patterns. Although different emotional states were predicted by activation in some of the same regions when considered at a macro scale of brain parcellation, the average voxel-wise overlap between patterns ($M = 2.66\%, \ 95\% \ CI \ [1.55\% \ 4.17\%]$) did not vary from random permutations of clusters of the same size within the same volume ($M = 1.78\%, \ P = .2386, 95\% \ CI \ [1.04\% \ 2.02\%]$). This result shows that despite being localized in similar macro-scale structures, the patterns were relatively non-overlapping at the voxel level. Such little overlap indicates that discrete emotional states are best differentiated on the basis of activation in a number of spatially disjoint brain regions.

Comparisons of PLS regression coefficients and analogous contrasts from univariate GLMs (i.e., each emotion vs. the average of all others) revealed moderate spatial correspondence of parameter estimates. Pearson correlations were highly significant for estimates of contentment ($r = .428$), amusement ($r = .644$), surprise ($r = .
Figure 9: Distributed patterns of neural activity predict discrete emotional states.

Each map depicts voxels whose activation consistently led to predictions of the related emotion across all subjects and stimulus modalities ($P < .05$, FWE corrected), overlaid on the group-average anatomical image.
Figure 10: Stacked bar plot showing the coverage of all regions in the AAL atlas by each emotion-predictive pattern.

.368), fear (r = .828), anger (r = .531), sad (r = .612), and neutral (r = .716) states (all P < .0001). Averaging across emotions, 5.53% of total gray matter voxels were predictive in both univariate and multivariate maps, while 11.05% were predictive only in multivariate maps, and 9.68% of voxels were predictive only in univariate maps. Thus, the two methods provide unique and complementary insight into neural activity associated with distinct emotions.
3.3.2 Testing categorical and dimensional models of emotion

In order to assess the structure of information within emotion-specific patterns, it was first necessary to validate that the emotion induction paradigm produced emotional experiences that varied along both emotion categories and affective dimensions of valence and arousal. Planned contrasts of categorical self-report items demonstrated that participants experienced the targeted emotion to a greater extent than alternative emotions (‘content’: $t_{31} = 4.19$, $P_{adj} = .0002$, 95% CI [1.870 2.522]; ‘amused’: $t_{31} = 8.32$, $P_{adj} < .0001$, 95% CI [1.26 3.43]; ‘surprised’: $t_{31} = 5.70$, $P_{adj} < .0001$, 95% CI [1.28 2.70]; ‘afraid’: $t_{31} = 7.40$, $P_{adj} < .0001$, 95% CI [1.61 3.84]; ‘angry’: $t_{31} = 6.50$, $P_{adj} < .0001$, 95% CI [1.56 3.00]; ‘sad’: $t_{31} = 4.57$, $P_{adj} = .0001$, 95% CI [1.01 2.63]; ‘neutral’: $t_{31} = 2.66$, $P_{adj} = .0123$, 95% CI [1.22 1.68]).

One-way repeated measures ANOVAs performed on dimensional self-report items revealed that emotional states could be differentiated in terms of valence (‘good’, ‘positive’, and ‘pleasant’ minus the scores for ‘bad’, ‘negative’, and ‘unpleasant’: $F_{6,186} = 50.503$, $P_{adj} < .0001$, 95% CI [20.39 246.7]) and arousal (‘agitated’, ‘active’, and ‘excited’ minus ‘calm’, ‘passive’, and ‘relaxed’: $F_{6,186} = 25.305$, $P_{adj} < .0001$, 95% CI [10.217 123.6]) dimensions. Further, plotting individual trials in an affective space organized by valence and arousal showed broad coverage and relatively uniform sampling (Figure 7), replicating prior work using the same stimuli (P. A. Kragel & LaBar, 2013). These results...
suggest the instances of emotion experienced in the scanner were adequate for use in testing both categorical and dimensional models.

Next, we quantified differences in emotional experience along categorical and dimensional models by computing the Euclidean distance in self-report ratings for each of 21 unique pairs of emotional states. Distances spanned seven dimensions in the categorical model, with each dimension directly corresponding to each emotion category. The dimensional model was constructed using self-report to parameterize dimensions from negative to positive valence and low to high arousal (P. A. Kragel & LaBar, 2013). The average distance between emotions did not differ between dimensional ($M = 1.37$, 95% CI [1.05 1.69]) and categorical models ($M = 1.55$, 95% CI [1.38 1.72]), indicating that the emotions selected for study were equally separable along either model.

To visualize the spatial homogeneity of each model, we performed hierarchical clustering of distances using Ward’s method in each model’s affective space. This analysis revealed a more uniform distribution of emotional states in the categorical space whereas states were more tightly clustered in the dimensional model ($F_{1,40} = 3.71$, $P = .0051$, 95% CI [1.50 9.13]) (Figure 11a). These results show the models we constructed were consistent with their theoretical underpinnings: emotional states are relatively
Figure 11: Categorical and dimensional models of emotion.

(A) Hierarchical clustering of self-report (left) and radar plots of categorical and dimensional terms (right). Note the increased distance between positive and negative emotions in the dimensional model and the reduced overlap between emotions in the categorical model. (B) BIC weights suggest a combined model with both categorical and dimensional terms was most likely to produce the observed distribution of classification errors (using subject-independent cross-validation). (C) Parameter estimates from the combined model show errors are more frequent as distance increases along valence and arousal, and that errors are less frequent as categorical distance increases. Error bars reflect 95% confidence intervals.
sparse and equidistant in the categorical account and more clustered and overlapping in
the dimensional account.

As participants experienced emotional states that could be characterized as both
independent categories and points in a low dimensional space, it is possible that pattern
classifiers utilized information organized according to either or both models to classify
emotional states. For instance, if the information in neural activation were to conform to
a single model (e.g. the categorical model with 7 axes) then the number of classification
errors should decrease with distance in the model space. Alternatively, if dimensions of
valence and arousal drove the classification results, then errors should decrease for
inductions rated differently in terms of valence and arousal. As the dimensional and
categorical models can be viewed as complementary, it is possible that increasing
distance along both categorical and dimensional models could contribute to fewer errors
in classification. To test these hypotheses, we conducted Bayesian model comparison
(see Methods; Kass & Raftery, 1995) to identify whether categorical, dimensional, or a
combination of the models was most likely to produce the observed classification
results.

Model comparison using Bayes factors strongly favored the combined model
across all other alternatives ($B_{comb-dim} = 6.43, B_{comb-categ} = 2632.3, B_{comb-con} = 1269.7$), indicating
that jointly considering dimensional and categorical conceptions of emotional states best
characterized classification errors (BIC weights for individual models are plotted in Figure 5b). Testing the parameter estimate for distance in categorical space indicated that the number of errors decreased with distance ($\beta = -.573$, $t = -2.69$, $P = .0073$, 95% CI [-0.992 -0.155]), suggesting increased experience of categorical emotions other than the target emotion increased classification errors. Surprisingly, inspection of parameter estimates for the dimensional term revealed the opposite relationship between distance in model space and classification errors ($\beta = 0.533$, $t = 4.56$, $P < .0001$, 95% CI [.304 0.762]). This positive effect remained significant when categorical distance was excluded from the model ($\beta = 0.293$, $t = 3.956$, $P < .0001$; see Appendix A.11), indicating that it was not the result of multicollinearity in the combined model. Furthermore, similar effects were observed for categorical ($\beta = -1.377$, $t = -6.63$, $P < .0001$, 95% CI [-1.785 -0.970]) and dimensional terms ($\beta = 0.5137$, $t = 4.23$, $P < .0001$, 95% CI [.276 .752]) when predicting the distribution of errors when modality-independent cross-classification was performed. Together, these results show that instances differing more in terms of valence and arousal were more difficult to classify, which is opposite to the relationship proposed by dimensional accounts of emotion. Together, these findings demonstrate that emotional states, as represented by neural activation patterns, are most separable during categorically distinct instances and are obscured when experiences differ in terms of valence and arousal.
3.3.3 Examining effects of valence and arousal

Given the detrimental effect of valence and arousal on classification performance, we examined their effect on neural activity at the group level using contrasts based on self-report (see Appendix A.12). Increasing valence was associated with activation in a number of brain regions including bilateral inferior parietal lobule, anterior dorsal cingulate, dorsomedial prefrontal cortex, and left somatosensory cortex. Decreasing valence was associated with activation in posterior cingulate, precuneus, middle and lateral occipital gyri, inferior temporal gyrus, and right middle frontal gyrus. Increases in self-reported arousal were associated with activation in numerous cortical and subcortical regions including thalamus, caudate, bilateral amygdala, bilateral insula, bilateral superior temporal gyrus, and right inferior frontal gyrus. Decreases in arousal were associated with activity in right parietal lobe.

3.4 Discussion

We have successfully predicted the occurrence of seven distinct emotional states using multivariate pattern classification across two different induction methods. In acquiring self-report ratings during scanning, we demonstrated for the first time a mapping between patterns of neural activation and an on-line assay of subjective feeling states. Relating subjective experience to neural activity is a critical advance over prior fMRI studies classifying emotional states in which experimental factors other than
emotion, such as low level stimulus properties or task demands, may have contributed to classification. These findings further establish the use of multivariate classification to study emotion in brain and bodily responses (Philip A. Kragel & LaBar, 2014; Nummenmaa, Glerean, Hari, & Hietanen, 2014; Shinkareva et al., 2013) and support the notion that emotions are categorically organized response patterns suited to bias behavior in functionally distinct ways.

Given the broad array of emotions induced, successful classification likely required considerable information beyond valence (or arousal) to differentiate emotional states. Our analysis of regression coefficients revealed this information was contained within diverse patterns of activation, spanning a number of cortical and subcortical brain regions. For instance, states of contentment and amusement were both rated as highly pleasant and positive, yet quantitatively disjoint patterns of activity formed their prediction. Maps for contentment included precuneus, medial prefrontal, cingulate, and primary somatosensory cortices -- regions implicated in self-referential processing (Northoff et al., 2006) and topographic representation of bodily sensations. Predictions of amusement were largely informed by activation in temporal cortex, supplementary motor area, and thalamus, which have been linked to processing of humor, laughter, and smiling (Wild, Rodden, Grodd, & Ruch, 2003; Wild et al., 2006). While the regions contained within these patterns are not individually specific to either emotional state (or
emotion in general), accumulating information across these disparate regions specified emotions as they were experienced as distinct categories.

Despite engaging partially overlapping neural substrates at the macro-scale, emotion-predictive patterns were largely non-overlapping at the voxel level. Such separability of emotional states at the voxel level may explain why meta-analytic works (Murphy et al., 2003; Phan et al., 2002; Vytal & Hamann, 2010) have associated neural activity with discrete emotions (e.g. correspondence between activation within the amygdala and fear or dorsal anterior cingulate and happiness), yet have failed to consistently identify emotion-specific neural substrates (Lindquist et al., 2012). The specificity and spatial separability of patterns identified in the present study suggest that methodological limitations inherent to univariate meta-analyses (variability of spatial normalization, focus on individual brain regions, and necessity of spatial smoothing) may have complicated the localization of emotion specific substrates in prior work.

We found that univariate and multivariate maps exhibited moderately similar spatial organization, although a sizeable number of voxels were differentially predictive across the two methods. Because multivariate classifiers pool information carried across multiple voxels and are generally more sensitive than univariate approaches (Jimura & Poldrack, 2012), the multivariate patterns presented here complement and extend prior
univariate findings. Further, it is important to note that PLS estimates better reflect the underlying latent structure in the fMRI data compared to ordinary least squares estimates (i.e. conventional GLMs), as they are less influenced by error variances at each voxel (Burnham, MacGregor, & Viveros, 2001). Comparisons and interpretations of the present results against prior and future univariate studies should accordingly be made with these methodological differences in mind.

Through Bayesian model comparison, we demonstrated that integrating categorical and dimensional configurations of affect best modeled the information contained in neural patterns during the experience of discrete emotional states. Although ratings made using categorical labels were sufficient to differentiate the emotional states, self-reports of emotion could also be distinguished along dimensions of valence and arousal, concordant with prior research on the similarity of emotion words (Bush, 1973; Russell, 1980), emotional experiences (Lisa Feldman Barrett & Russell, 1999), and facial expressions of emotion (Schlosberg, 1952; Susskind, Littlewort, Bartlett, Movellan, & Anderson, 2007). In this sense, classification of neural activation could be driven by either dimensional or categorical aspects of experience (P. A. Kragel & LaBar, 2013). Strikingly, this was not the case in the neural data: the optimal model showed that emotional states that differed the most in terms of valence and arousal were the most difficult to predict, whereas differences on categorical aspects of experience led
to better classification. Neural responses related to valence or arousal may have obscured patterns of neural activation that more effectively differentiate emotional states.

From a machine learning perspective, neural variance related to valence or arousal may have acted more as noise (which led to poorer classification) as opposed to informative features that differentiate emotions. This is not to say that neural correlates of valence and arousal are unstructured or unrelated to the experience of emotion. Rather, valence or arousal related activity likely co-occurs with emotional experiences but poorly discriminates among emotions similar along such dimensions, leading to more classification errors. For example, our univariate analyses revealed activation in middle and superior temporal gyrus associated with increasing arousal (consistent with Kensinger & Schacter, 2006; Mather et al., 2006; Mourao-Miranda et al., 2003), whereas classification models identified distinct patterns of activation in this region that contributed to prediction of amusement, anger, and sadness. Widespread activation associated with arousal may have masked information specific to each emotion, possibly explaining higher rates of confusion between these emotions. Such opposition between categorical and dimensional representation is incompatible with models suggesting emotions emerge from undifferentiated states of valence and arousal, as such accounts
suggest differences in affective dimensions should aid in discrimination among emotions.

Although the present findings refute the idea that valence and arousal serve to differentiate emotional states as reflected in patterns of neural activation, such affective dimensions have been linked to distinct neural systems (A. K. Anderson et al., 2003; Chikazoe, Lee, Kriegeskorte, & Anderson, 2014; Lang & Bradley, 2010; Lewis et al., 2007; Small et al., 2003). This body of research localized the neural basis of these affective properties using pre-specified regions of interest. It is possible that using whole-brain data for classification of emotional state in the present study overlooked information represented locally in these regions, due to the presence of other signals that better distinguished emotions. Alternatively, the sampling of emotions could have prevented differences in valence and arousal from informing classification. As multiple emotions were similar in terms of valence and/or arousal (e.g. fear and anger or content and neutral states), pattern classifiers may have underutilized activation related to these dimensions if it was present in the data. For these reasons, the present results should not be taken as de facto evidence against dimensional representations of affect in the brain. Rather, these findings implicate separable neural substrates in multi-faceted categorical experience of emotion beyond a continuous mapping of valence and arousal.
We found that neutral states were among those most accurately classified. This finding is consistent with recent studies examining the perception of facial expressions (Said et al., 2010) and bodily sensations (Nummenmaa et al., 2014). Neutral states are likely to have high classification accuracy because they are more likely to differ from all other emotions along either dimensional or categorical factors. In a circumplex dimensional account, for instance, the neutral origin is equidistant from emotions organized along the perimeter of affective space. Emotions that are clustered about the perimeter are thus more difficult to differentiate, as they share components of valence and arousal. Additionally, neutral states can be differentiated on the basis of activation at the superordinate non-emotional vs. emotional level, whereas discrete emotions must be classified only at the subordinate level as they all contain emotional content.

The present findings extend related research classifying emotional states on the basis of brain activity. In recent work by Kassam and colleagues (2013), states of anger, disgust, envy, fear, happiness, lust, pride, sadness, and shame were self-induced by method actors during the presentation of an emotional words (e.g. afraid or frightened). These states were identified at accuracies well above chance levels of .50 using classification models both within (mean rank accuracy of .84) and across subjects (mean rank accuracy of .70), and an exploratory factor analysis revealed that neural activation during the self-induction period was organized along dimensions interpreted as valence,
arousal, sociality, and lust. While the present results are consistent their findings of above chance accuracy levels, we did not find evidence indicating that neural activation was organized along dimensions of valence and arousal. It is possible that the presence of emotional words, which have a semantic structure well characterized along such dimensions (Storm & Storm, 1987) and have been shown to elicit responses in frontal and limbic structures (Lewis et al., 2007), may explain this difference in results.

As emotion induction trials lasted several minutes, the activation patterns we classified may have been driven by multiple processes which ultimately lead to categorically experienced states. The acquired time-course of neural activation likely reflected perceptual evaluation of stimuli, physiological expression of emotion, and construction of emotional experiences. Functional neuroimaging studies of emotion perception (Peelen, Atkinson, & Vuilleumier, 2010; Said et al., 2010) have implicated the superior temporal sulcus in representing the emotional content of stimuli categorically. Psychophysiological research has demonstrated categorical specificity in peripheral physiological responses (P. A. Kragel & LaBar, 2013; Stephens et al., 2010), which have been shown to bi-directionally predict subjective feelings (Friedman, Stephens, & Thayer, 2014). Given the numerous and interrelated processes that unfold during an emotional episode, it is unlikely that the patterns we identified perfectly capture the essence an emotion, but are better characterized as an amalgamation of components
which make emotions unique. Future research examining which aspects of the task are essential for categorical experience of emotion will be necessary to further evaluate theories concerning the neural basis of emotions, and whether they sufficiently meet the criteria for being considered biologically basic.

In conclusion, we have found mappings between neural activation patterns and categorically distinct emotional experiences, raising numerous questions central to the representation of emotion in the brain. The patterns identified here clarify the relationship between brain activity and the experience of a subset of emotions, but they do not fully define the emotional experience of participants. Quantifying other factors, such as motivational state or social context, could further elucidate the structural organization of emotions in the brain. Assessing generalization to other forms of induction is another important avenue to further validate and refine neural representations of emotion. Examining whether the experience of emotion can be modeled at more fine-grained intervals through parametric manipulations and continuous prediction models (e.g. SVR, LASSO-PCR, or PLS regression, see Favilla et al., 2014; Marquand et al., 2010; T. D. Wager et al., 2013) may yield more sensitive accounts of emotion experience. Beyond informing neural and theoretical models of emotion, biomarkers of emotional states could improve our understanding of mood and affective disorders, aid in diagnosis, and serve as novel outcome measures for
interventions. Key diagnostic criteria for these disorders include subjective feelings of fear, anxiety, and sadness. Such feelings are complex and difficult to quantify objectively. Therefore, examining the prevalence and congruence of these patterns in healthy and clinical populations may further clarify the order and disorder of emotion in the human brain.
4. Decoding spontaneous emotional states in the human brain

This chapter describes a functional magnetic resonance imaging (fMRI) experiment examining whether emotional states, as reflected by distinct patterns of central nervous system activation, emerge from intrinsic brain activity during wakeful rest. Resting state fMRI time-series were drawn from the Duke Neurogenetics Study and decoded into distinct emotional states utilizing the classification models developed in Chapter 3. To test the extent to which such spontaneous emotional states are consistent with on-line self-report of emotion, a follow-up experience sampling study was conducted in which participants rated their own emotional feelings following periods of unconstrained, wakeful rest.

4.1 Introduction

The experience of emotional feelings elicits widespread patterns of neural activity spanning cortical and subcortical brain regions (A. R. Damasio et al., 2000; P. A. Kragel & LaBar, 2015). Research using machine-learning based approaches has mapped such multivariate patterns onto categorical labels of such experiences including happiness, fear, anger, and sadness (Kassam et al., 2013; P. A. Kragel & LaBar, 2015; Sitaram et al., 2011). A common assumption in neurobiological models of emotion (A. Damasio & Carvalho, 2013b; Oatley & Johnson-Laird, 1987; J. Panksepp, 1982) is that these brain states are responses to events with inherent or learned significance (Ekman
& Cordaro, 2011; LeDoux, 2000). However, it is unclear whether emotional brain states occur intrinsically (Fox & Raichle, 2007; Vincent et al., 2007), in the absence of external eliciting stimuli, and what factors modulate the rate at which emotion-specific patterns arise spontaneously during rest.

We adapted the logic of other cognitive imaging studies (Kenet, Bibitchkov, Tsodyks, Grinvald, & Arieli, 2003; Vincent et al., 2007) to postulate that the presence of spontaneous emotional brain states should be detectable using models derived from prior investigations of emotion elicitation. We previously developed decoding algorithms to classify stimulus-evoked responses to emotionally evocative cinematic films and instrumental music (P. A. Kragel & LaBar, 2015). These models accurately classify patterns of neural activation across induction methods associated with six different emotions (contentment, amusement, surprise, fear, anger, and sadness) and a neutral control state in independent subjects and track self-reported emotional experience. By indexing the extent to which a pattern of neural activation to extrinsic stimuli reflects a specific emotion, these models can be used to test whether intrinsic patterns of brain activity correspond to similar emotional states.
4.2 Methods

4.2.1 Resting-state experiment

4.2.1.1 Participants.

A total of 499 subjects (age = 19.65 ± 1.22 years (mean ± s.d.), 274 women) were included from an ongoing parent protocol, the Duke Neurogenetics Study (DNS), which assesses a wide range of behavioral and biological traits among healthy, young adult university students. This sample size is sufficient to reliably detect ($\beta = .01$) a moderate effect ($r = .2$) with a type-I error rate of .05, which is particularly important when studying individual differences in neural activity. All participants provided informed consent in accordance with Duke University guidelines and were in good general health. The participants were free of the following study exclusions: (1) medical diagnoses of cancer, stroke, head injury with loss of consciousness, untreated migraine headaches, diabetes requiring insulin treatment, chronic kidney or liver disease, or lifetime history of psychotic symptoms; (2) use of psychotropic, glucocorticoid, or hypolipidemic medication; and (3) conditions affecting cerebral blood flow and metabolism (e.g. hypertension). Diagnosis of any current DSM-IV Axis I disorder or select Axis II disorders (antisocial personality disorder and borderline personality disorder), assessed with the electronic Mini International Neuropsychiatric Interview (Sheehan et al., 1998) and Structured Clinical Interview for the DSM-IV subtests (First, 1996) were not an exclusion, as the DNS seeks to establish broad variability in multiple behavioral
phenotypes related to psychopathology. No participants met criteria for a personality disorder and 72 (14.4%) participants from our final sample met criteria for at least one Axis I disorder (10 Agoraphobia, 33 Alcohol Abuse, 3 Substance Abuse, 25 Past Major Depressive Episode, 5 Social Phobia). However, as noted above none of the participants were using psychotropic medication during the course of the DNS.

4.2.1.2 Data acquisition and pre-processing.

Participants were scanned on a 3 Tesla General Electric MR 750 system with 50-mT/m gradients and an eight channel head coil for parallel imaging (General Electric, Waukesha, Wisconsin, USA). High-resolution 3-dimensional structural images were acquired coplanar with the functional scans (repetition time (TR) = 7.7 s; echo time (TE) = 3.0 ms; flip angle (α) = 12°; voxel size = 0.9 × 0.9 × 4 mm; field of view (FOV) = 240 mm; 34 contiguous slices). For the two 4 min, 16 s resting-state scans, a series of interleaved axial functional slices aligned with the anterior commissure–posterior commissure plane were acquired for whole-brain coverage using an inverse-spiral pulse sequence to reduce susceptibility artifact (TR = 2000 ms; TE = 30 ms; α = 60°; FOV = 240 mm; voxel size = 3.75 × 3.75 × 4 mm; 34 contiguous slices). Four initial radiofrequency excitations were performed (and discarded) to achieve steady-state equilibrium. Participants were shown a blank gray screen and instructed to lie still with their eyes open, think about nothing in particular, and remain awake.
Preprocessing of all resting-state fMRI data was conducted using SPM8 (Wellcome Department of Imaging Neuroscience). Images for each subject were slice-time-corrected, realigned to the first volume in the time series to correct for head motion, spatially normalized into a standard stereotactic space (Montreal Neurological Institute template) using a 12-parameter affine model (final resolution of functional images = 2 mm isotropic voxels), and smoothed with a 6mm FWHM Gaussian filter. Low-frequency noise was attenuated by high-pass filtering with a 0.0078 Hz cut-off.

### 4.2.2 Experience sampling experiment

#### 4.2.2.1 Participants

A total of 22 subjects (age = 26.04 ± 5.16 years (mean ± s.d.), 11 women) provided informed consent and participated in the study. Data from one participant was excluded from analyses because of excessive head movement (in excess of 1 cm) during scanning. While no statistical test we performed to determine sample size a priori, this sample size is similar to those demonstrating a correspondence between self-report of affect and neural activity (Chikazoe et al., 2014; P. A. Kragel & LaBar, 2015; Tusche, Smallwood, Bernhardt, & Singer, 2014).

#### 4.2.2.2 Experimental procedure

Participants engaged in an experience sampling task in which they rated their current feelings during unconstrained rest. Participants were instructed to keep their
eyes open and let their mind wander freely and that a rating screen (Scherer, 2005) would occasionally appear, which they should use to indicate the intensity of the emotion that best describes how they currently feel. This validated assay of emotional self-report consists of 16 emotion words organized radially about the center of the screen. Four circles emanate from the center of the screen to each word (similar to a spoke of a wheel), which were used to indicate the intensity of each emotion by moving the cursor about the screen. During four runs of scanning, participants completed 40 trials (10 per run) with an inter-stimulus interval (ISI) of 30 seconds plus pseudo-random jitter (Poisson distribution, $\lambda = 4$ sec).

4.2.2.3 Behavioral scoring

Self-report data were transformed from two-dimensional cursor locations to categorical labels. Polygonal masks were created by hand corresponding to each emotion term on the response screen. A circular mask in the center of the screen was created for neutral responses. Because terms in the standard response screen did not perfectly match those in the neural models, the item ‘relief’ was scored as ‘content’ whereas ‘joy’ and ‘satisfaction’ were scored as ‘amusement’. The items ‘surprise’, ‘fear’, ‘anger’, ‘sadness’, and ‘neutral’ were scored as normal.
4.2.2.4 Data acquisition and pre-processing.

Scanning was performed on a 3 Tesla General Electric MR 750 system with 50-mT/m gradients and an eight channel head coil for parallel imaging (General Electric, Waukesha, Wisconsin, USA). High-resolution images were acquired using a 3D fast SPGR BRAVO pulse sequence (TR = 7.58 ms; TE = 2.936 ms; image matrix = 256\(^2\); \(\alpha = 12^\circ\); voxel size = 1 × 1 × 1 mm; 206 contiguous slices) for coregistration with the functional data. These structural images were aligned in the near-axial plane defined by the anterior and posterior commissures. Whole-brain functional images were acquired using a spiral-in pulse sequence with sensitivity encoding along the axial plane (TR = 2000 ms; TE = 30 ms; image matrix = 64 x 128; \(\alpha = 70^\circ\); voxel size = 3.8 x 3.8 x 3.8 mm; 34 contiguous slices). Four initial radiofrequency excitations were performed (and discarded) to achieve steady-state equilibrium.

Processing of MR data was performed using SPM8 (Wellcome Department of Imaging Neuroscience). Functional images were slice-time-corrected, spatially realigned to correct for motion artifacts, coregistered to high resolution anatomical scans, normalized to Montreal Neurologic Institute (MNI) space using high-dimensional warping implemented in the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm.html). Low-frequency noise was attenuated by high-pass filtering with a 0.0078 Hz cut-off.
4.2.3 General methods

4.2.3.1 Pattern classification

To rescale data for classification, preprocessed time-series were standardized by subtracting their mean and dividing by their standard deviation. Maps of partial least squares (PLS) regression coefficients (Wold et al., 2001) from stimulus-evoked decoding models (P. A. Kragel & LaBar, 2015) were resliced to match the voxel size of functional data. Classifier scores were computed by taking the scalar product of functional data at each timepoint and regression coefficients from content, amusement, surprise, fear, anger, sad, and neutral models. Individual time-points were assigned categorical labels by identifying the model with the maximal score. The similarity of classification models and the resting-state networks was assessed by computing the maximal Jaccard index for each emotion model (thresholded at the 95th percentile) and the seven networks identified in Yeo et al. (Yeo et al., 2011).

4.2.3.2 Statistical inference

When conducting inferential tests on classification frequency (count data), non-parametric tests were conducted. To test whether classifications were uniformly distributed across the emotion categories, a Friedman test was performed ($n = 499$ subjects, $k = 7$ emotions). Wilcoxon signed-rank test were performed to test for differences in frequency relative to chance rates (14.3%). To test whether classifier scores
changed over time, Friedman tests were conducted on the output of the emotion models separately (as classifier scores were found to violate assumptions of normality). Follow-up tests on the direction of these changes (either as increases or decreases) were conducted using general linear models with one constant regressor and another for linearly increasing time. Inference on the parameter estimate for changes over time was made using a $t$-distribution with 497 degrees of freedom.

The influence of individual differences in mood and personality were assessed using generalized linear models with a binomial distribution and a logit link function. Multiple models were constructed, using measures from the CESD, STAI, or facets from the NEO-PI-R to predict the frequency of classifications for the seven emotion categories. Inference on parameter estimates was made using a $t$-distribution with 497 degrees of freedom.

For the experience sampling study, classifier scores were averaged for trials congruent and incongruent with self-report for each subject. Classifier scores were rescaled by a factor of $3.8^{3/2}$ to accommodate for differences in voxel size between the resting-state and experience sampling experiments. Because these scores for congruent ($P = 0.0186$, Lilliefors test) and incongruent ($P = 0.0453$) trials exhibited non-normal distributions, Wilcoxon signed rank tests were used to test each sample against zero mean rank. The correspondence between the frequencies of classification labels from
self-report and neural decoding was assessed by computing the Pearson correlation for each subject. The correlation coefficients were Fisher transformed and tested against zero using a one-sample $t$-test.

### 4.3 Results

We applied these models to neural activation acquired during resting-state fMRI as part of the Duke Neurogenetics Study ($n = 499$). Two consecutive runs of resting-state scans were acquired, spanning a total duration of 8.53 minutes. Following preprocessing of data, we computed the scalar product of the resting-state signal and model weights at every time point of data acquisition (Figure 12). This procedure yielded scores that reflect the evidence for each of seven emotional states across the full scanning period.

If emotional brain states occur spontaneously, a proportion of classification scores from our decoding models should be greater than would be expected by chance. To estimate scores that would be observed by an uninformative, random pattern, null models were created by spatially permuting the true model weights. The distribution of scores from the true models (colored histograms, Figure 12b) was generally more extreme than those for null models (gray histograms, Figure 12b). Averaging across the seven distributions, the mean maximal score from null models was $0.502 \pm .293$ (s.d.), compared to $3.49 \pm .54$ (s.d.) for the stimulus-evoked models. Such a difference provides an initial indication that spontaneous patterns of activation during rest resemble
categorical brain states evoked by external emotional stimuli. To ensure that our emotion-specific brain states are not proxies for more general resting-state networks that subserve other functions, we examined the spatial overlap between our models and those commonly derived by connectivity based analysis of resting-state fMRI data (Yeo et al., 2011) (see Methods). On average, we observed little overlap (Jaccard index = 4.81 ± 0.77% (s.d.); range 4.0% - 6.0%) with the 7 most prominent networks found in resting-state data, implicating a high degree of independence.

Given this evidence, we next sought to identify whether the total time (or absolute frequency) in each state differed across emotion categories. Such an analysis informs the degree to which discrete emotional brain states may spontaneously occur and, by extension, could contribute to the identification of individual differences that map onto the likelihood of experiencing specific spontaneous states. To perform this comparison, we identified the model with the maximum score at each time point and summed the number of time points assigned to each category. The frequency of emotional states clearly differed across categories (Figure 12c, $\chi^2 = 1491.52$, $P < .0001$, Friedman test), with neutral states occurring more frequently than chance rates ($20.1 \pm 3.59\%$ (s.d.), $z = 20.50$, $P = 2.03E-93$), followed by states of surprise ($18.37 \pm 3.87\%$ (s.d.), $z = 16.38$, $P = 2.47E-60$) and amusement ($14.71 \pm 3.78\%$ (s.d.), $z = 1.25$, $P = 0.21$). States of sadness ($13.49 \pm 3.76\%$ (s.d.), $z = -3.31$, $P = 9.24E-4$), fear ($13.26 \pm 3.42\%$ (s.d.), $z = -5.28$, $P =$
Figure 12: Emotional states emerge spontaneously during resting-state scans.

(A) Procedure for classification of resting-state data. Scores are computed by taking the scalar product of preprocessed data and regression weights from decoding models. (B) Distribution of classification scores, with model output plotted in color and null maps in gray. Distributions are wider for classification models, indicating evidence for emotional states greatly exceeds that which would be expected by chance. (C) Frequency distributions for the classification of all seven emotional states ($n = 499$). The mean, 25th and 75th percentiles are indicated by black lines. The solid gray line indicates the number of trials which would occur from random guessing.
1.28E-7), and anger (11.31 ± 3.62% (s.d.), z = -13.07, P = 4.78E-39) occurred with lower frequency, while states of contentment occurred the least often (8.74% ± 3.42% (s.d.), z = -19.61, P = 1.33E-85).

Although patterns of neural activation were most often classified as neutral as a whole, it is possible that consistent fluctuations in the time-course of emotional states occur against this background. Research on MRI scanner-related anxiety has shown that self-report (Chapman, Bernier, & Rusak, 2010; van Minde, Klaming, & Weda, 2014) and peripheral physiological (Muehlhan, Luken, Wittchen, & Kirschbaum, 2011) measures of anxiety peak at the beginning of scanning when subjects first enter the scanner bore. This literature would predict that brain states indicative of fear should be most prevalent at the beginning of resting-state runs, and that neutral states should emerge over time, given their overall high prevalence. To determine whether emotional states consistently varied across subjects over the course of the scanning period, we first performed Friedman tests separately for each emotion category, all of which revealed significant effects of time (Appendix B). Next, we quantified the direction of these effects using general linear models to predict classifier scores using scan time as an input. We found the scores for fear exhibited a peak at the beginning of scanning and dropped off over time (β = -0.001, t_{497} = -4.92, P = 1.20E-006, see Fig. 2), whereas neutral states exhibited an
increasing trend throughout the scanning period ($\beta = 0.0017$, $t_{497} = 7.36$, $P = 7.66 \times 10^{-13}$), consistent with predictions.

Figure 13: Coherent fluctuations in emotional state during resting-state scans.

Time-series indicate the sample mean classification scores for all seven emotions ($n = 499$, errorbars reflect standard error at each acquisition). Data are concatenated across two sessions of 256 seconds (dashed line). Note the initial peak for fear scores at the beginning of each session and general increases in neutral scores across sessions.

To further establish the construct validity of the spontaneous emotional brain states, we reasoned that their incidence should vary with individual differences in self-reported mood and personality traits associated with specific emotions. We assayed depressive mood with the Center for Epidemiologic Studies Depression Scale (Radloff, 1977) (CESD) and state anxiety using the State-Trait Anxiety Inventory State Version (Spielberger, Gorsuch, & Lushene, 1970) (STAI-S), instructing participants to indicate how they felt during the resting-state scan itself. Binomial regression models revealed that higher depression scores were associated with increases in the frequency of sadness.
(β = 0.0025, t_{497} = 2.673, P = .0075, Fig. 3a) and no other emotional state (all P > .24). State anxiety was associated with increasing classifications of fear (β = 0.0033, t_{497} = 2.608, P = .0091) and decreasing frequency of contentment (β = -0.0031, t_{497} = -2.015, P = .0439).

Drawing from the Revised NEO Personality Inventory (NEO-PI-R) (20), we focused personality trait assessment on the specific Neuroticism subfacets of Anxiety, Angry Hostility, and Depression, due to their discriminant validity (P. T. Costa & McCrae, 1995; McCrae & Costa, 1992), heritability (Jang, Livesley, & Vernon, 1996; Jang, McCrae, Angleitner, Riemann, & Livesley, 1998), universality (McCrae & Costa, 1997), and close theoretical ties to the experience of fear, anger, and sadness. We found that increasing Anxiety scores were associated with more frequent classification of fear (β = 0.003, t_{497} = 1.978, P = 0.0479, Fig. 3b) and fewer classifications of anger (β = -0.004, t_{497} = -2.407, P = 0.0161). Angry Hostility scores were positively associated with the number of anger classifications (β = 0.0042, t_{497} = 2.400, P = 0.0164). Depression scores were positively associated with the frequency of fear (β = 0.003, t_{497} = 2.058, P = 0.0396) and sadness (β = 0.0037, t_{497} = 2.546, P = 0.0109). These results provide converging evidence across both state and trait markers that individual differences uniquely and differentially bias the spontaneous occurrence of brain states indicative of fear, anger, and sadness.
Figure 14: Individual differences in mood and personality modulate the occurrence of spontaneous emotional brain states.

(A) Differences in depressive and anxious mood are associated with increases in the frequency of sad and fear classifications. (B) Emotional traits of Anxiety, Angry Hostility, and Depression predict differences in the frequency of fear, anger, and sad classifications ($n = 499$, error bars reflect standard error, * indicates effects significant at $P < .05$).
Finally, we examined whether the predictions of our decoding models were consistent with self-report of emotional experience during periods of unconstrained rest. We conducted a separate fMRI experiment in which an independent sample of young adult participants ($n = 21$) performed an experience sampling task in the absence of external stimulation (Fig. 15a). Participants were instructed to rest and let their mind wander freely with their eyes open during scanning. Following intervals of rest of at least 30 seconds, a rating screen appeared during which participants moved a cursor to the location on the screen which best indicated how they currently felt.

If spontaneous emotional states are accessible to conscious awareness, then scores should be greater for emotion models congruent with self-report relative to scores for models incongruent with self-report. To test this claim, we extracted resting-state fMRI data from the 10-sec interval preceding each self-report query and applied multivariate models to determine the extent to which evidence for the emotional brain states in this window predicted the participants’ conscious emotional experience. Consistent with our hypothesis, we found that scores for models congruent with self-report were positive ($0.016 \pm 0.0093$ (s.e.m.), $z = 2.068$, $P = 0.0386$; Wilcoxon signed rank test), whereas scores for incongruent models were negative ($-0.0048 \pm 0.0017$ (s.e.m.), $z = -3.041$, $P = 0.0024$). Not only do these results demonstrate that classification models are sensitive to changes in emotional state reported by participants, but also that there is
Figure 15: Spontaneous emotional brain states exhibit correspondence with self-report.

(A) Participants \((n = 21)\) participated in an experience sampling task in which they reported their current emotional state at random intervals exceeding 30 sec during fMRI scanning. The five samples of data (lasting 10 sec, TR = 2 sec) preceding each rating were used to compute predictions of emotional state using multivariate decoding models. (B) Scores for classification models congruent with self-report are greater than incongruent models \((P = 0.0208;\) Wilcoxon signed rank test). (C) The frequency of classifications from multivariate models significantly correlates with those made by participant self-report \((P = .0196;\) One sample t-test). Gray line indicates best fitting least-squares line for group mean. In all plots, error bars reflect standard error of the mean.

selectivity in their predictions, as negative scores indicate evidence against emotion labels that are incongruent with self-report.

As an additional validation of our decoding models, we next examined the correspondence between the prevalence of individual emotional brain states as detected
via pattern classification and participant self-report. Classifications based on self-report and multivariate decoding yielded similar frequency distributions (Fig. 15c), in which neutral and amusement were the most frequent. We found a positive correlation between the frequency of classifications based on participant ratings and multivariate decoding ($r = .3876 \pm 0.102$ (s.e.m.), $t_{20} = 2.537$, $P = .0196$; One sample t-test), further demonstrating a link between patterning of brain states and subjective ratings of emotional experience – in the absence of stimuli or contextual cues.

### 4.4 Discussion

In summary, converging findings from our experiments provide evidence that brain states associated with distinct emotional experiences emerge during unconstrained rest. Whereas prior work has decoded stimulus-evoked responses to emotional events, our study demonstrates that spontaneous neural activity dynamically fluctuates between multiple emotional states in a reliable manner over time. Furthermore, individual differences in mood states and personality traits are associated with the relative incidence of brain states associated with fear, anger, and sadness. In combination, these findings have important implications for using unconstrained rest as a baseline for emotion activation tasks, as emotional states dynamically shift across rest and the incidence of specific emotions varies across individuals.
Our findings corroborate theories that posit emotions emerge from coordinated patterns of activity in separable brain systems (A. Damasio & Carvalho, 2013b; Ekman, 1992; J. Panksepp, 1982) that act to bias behavior in unique ways. Our results indicate that emotions are not exclusively driven by stimulus-bound or contextual factors but instead may also reflect the distributed activity of intrinsic brain states that play a central role in guiding behavior (D. J. Anderson & Adolphs, 2014). Given their sensitivity to individual differences, spontaneous emotional brain states may serve as a novel diagnostic tool to determine susceptibility to affective illness or as an outcome measure for clinical interventions aimed at reducing the spontaneous elicitation of specific emotions.
5. Somatosensory representations provide a neural basis for emotional contagion

This chapter describes a functional magnetic resonance imaging (fMRI) experiment examining the neural mechanism that mediates the spread of emotions between individuals, termed emotional contagion (Hatfield, Cacioppo, & Rapson, 1994). Embodied theories of cognition propose the emergence of this phenomenon, wherein perceiving emotions in others elicits the somatosensory re-experience of the emotion in oneself. To test the whether somatosensory cortex mediates such a process, neural activation was measured using fMRI while participants viewed facial and vocal expressions of emotion and subsequently rated their emotional experience. The extent to which patterns of somatosensory activation discriminate among perceived emotions, exhibit somatotopic organization, and are consistent with subjectively experienced emotion (i.e. feeling emotions congruent with those conveyed by stimuli) can be used as criteria for evaluating the role of this brain region in the social transmission of emotion.

5.1 Introduction

During social interactions, humans recognize emotions from facial and vocal cues with seemingly little effort. Often, the perception of emotions in others leads to contagious emotional behaviors such as crying and laughter. Simulationist models of emotion recognition (Adolphs, 2002; Goldman & Sripada, 2005; Niedenthal, 2007) propose that these phenomena result from neural processing in somatosensory cortex.
(Keysers, Kaas, & Gazzola, 2010). Although this region is primarily involved in tactile sensation, it has been argued that somatosensory representations also facilitate emotion recognition by linking non-tactile perceptual cues to bodily states associated with each emotional category (A. R. Damasio, 1996).

Damage to or inactivation of right somatosensory cortex disrupts the recognition of emotion from facial (Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000; Pitcher, Garrido, Walsh, & Duchaine, 2008) and vocal (Adolphs, Damasio, & Tranel, 2002; Banissy et al., 2010) expressions. However, it is unknown if this behavioral impairment is due to an experiential mirroring mechanism, as suggested by embodied cognition perspectives. Although emotional expressions can be decoded from patterns of activation within unimodal (Ethofer et al., 2009; Harry, Williams, Davis, & Kim, 2013) or multimodal association cortices (Peelen et al., 2010; Wegrzyn et al., 2015) it is unknown whether neural activity within primary somatosensory cortex codes categorical information from perceived emotions in the non-tactile domain, and whether such activity is related to subjective experience in terms of its separability and topographic organization.

To explore the nature of neural activation underlying emotion recognition, we conducted a functional magnetic resonance imaging (fMRI) experiment in which participants were presented with facial and vocal expressions of discrete emotions and
made on-line ratings of their own emotional experience in response to these percepts. Given that expressions of emotion lead to the convergence of facial configuration and shared mood (Hess & Blairy, 2001), we expected behavioral self-report to mirror the emotional content of stimuli. Further, if somatosensory representations reflect how one would feel when making an emotional expression, then it should be possible to decode emotion-specific patterns of neural activation within right somatosensory cortex, and the spatial configuration of these patterns should be consistent with known somatotopy.

5.2 Methods

5.2.1 Participants

Twenty-one healthy, right-handed individuals (M_{age} = 26 years, age range = 19 to 39 years, 11 males) completed the study. One additional participant was run in the experiment, but was excluded from analysis due to excessive head-motion during scanning (displacement exceeding 1 cm). Although no formal power analysis was performed, the final sample size is similar to those of related fMRI studies of emotion recognition (Harry et al., 2013; Peelen et al., 2010; Said et al., 2010; Skerry & Saxe, 2014). All participants provided written informed consent to participate in accordance with the Duke University Institutional Review Board and received $20 per hour as monetary compensation for completion of the study.
5.2.2 Experimental paradigm

During scanning, participants were presented with facial and vocal expressions of emotion, followed by a self-report. To isolate neural responses related to stimuli, the period between stimulus presentation and motor response was jittered following a Poisson distribution (\(\lambda = 4\) sec).

The stimuli used included standardized images of faces (Langner et al., 2010) and audio recordings of pseudo-utterances (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009), which clearly convey emotions of happiness, surprise, fear, anger, and sadness, in addition to neutral control expressions. Twelve expressions were presented in each modality for each emotion, resulting in a total of 144 unique stimuli. Participants viewed stimuli in one of four pseudorandom counterbalanced orderings, which alternated between blocks of facial or vocal expressions. Each block consisted of one male and one female presentation of each emotion, for a total of 12 trials. Facial stimuli were presented for 1.5 sec, while auditory stimuli lasted 1.65 ± .32 sec (mean ± s.d.). Each experimental session consisted of three runs of data acquisition, comprising four blocks and lasting on average approximately 10.26 min.

During the self-report phase, the Geneva Emotion Wheel (Scherer, 2005) was presented on the screen for 6 sec. This self-report assay consists of 16 emotion words organized radially about the center of the screen. Four circles emanate from the center of
the screen to each word (similar to a spoke of a wheel), which can be used to indicate the intensity of each emotion. Participants were instructed to use a joystick to move the cursor to the location on the screen which best indicated how they currently feel. Participants were additionally instructed to move the cursor to the center of the screen if they do not feel any of the emotions listed. Prior to functional scanning, participants completed a set of practice trials in which they moved the cursor to each emotion term to ensure proper function of the joystick and comprehension of the response format.

Presentation of stimuli and acquisition of behavioral responses was controlled using Cogent 2000 software (Wellcome Department of Imaging Neuroscience, http://www.vislab.ucl.ac.uk/cogent.php). Participants viewed stimuli on mirrors aligned with a LCD screen upon which images were projected from a stimulus control computer. Audio stimulation was presented using MR-compatible headphones (Resonance Technology, NorthRidge California).

5.2.3 Image acquisition

Scanning was performed on a 3 Tesla General Electric MR 750 system with 50-mT/m gradients and an eight channel head coil for parallel imaging (General Electric, Waukesha, Wisconsin, USA). High-resolution images were acquired using a 3D fast SPGR BRAVO pulse sequence: repetition time (TR) = 7.58 ms; echo time (TE) = 2.936 ms; image matrix = 256²; α = 12°; voxel size = 1 × 1 × 1 mm; 206 contiguous slices) for
coregistration with the functional data. These structural images were aligned in the near-axial plane defined by the anterior and posterior commissures. Whole-brain functional images were acquired using a spiral-in pulse sequence with sensitivity encoding along the axial plane (TR = 2000 ms; TE = 30 ms; image matrix = 64 × 128; α = 70°; voxel size = 3.8 × 3.8 × 3.8 mm; 34 contiguous slices). The first five images of each run were excluded from analyses to ensure the magnet had reached steady state.

5.2.4 Preprocessing and estimating neural activation.

Processing of MR data was performed using Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience). Functional images were slice-time-corrected, spatially realigned to correct for motion artifacts (Friston et al., 1995), coregistered to high resolution anatomical scans (Collignon et al., 1995), and normalized to Montreal Neurologic Institute (MNI) space using high-dimensional warping implemented in the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm.html).

Whole-brain patterns of neural activation were estimated using the general linear model approach implemented in SPM8. For each subject, BOLD responses were modeled by convolving box-car functions with a canonical hemodynamic response function separately for each trial. One additional regressor modelling the self-report phase was included in each run. To model nuisance effects, six motion parameters (roll,
pitch, yaw, in addition to translation in x, y, and z dimensions) and session constants, were incorporated into the model.

5.2.5 Regions of interest

Anatomical masks were created for brain regions implicated in a neural network model of emotion recognition (Adolphs, 2002; Adolphs et al., 2000) using the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). In particular, masks were created for right postcentral gyrus and bilaterally for posterior superior temporal sulcus (pSTS; defined as voxels within superior temporal gyrus posterior to y = -32 mm in MNI space), medial orbitofrontal cortex (mOFC), inferior frontal operculum (IFO), fusiform gyrus (FG), amygdala, and insula.

5.2.6 Classification of self-report

Cursor locations (two-dimensional data centered on the center pixel on the presentation computer) were used to predict the emotional content of stimuli. Because the self-report format was circular in nature, classification was performed using support vector machines (SVMs) with a radial basis function as implemented in LIBSVM (Chang & Lin, 2011). As cursor locations were assigned one of six labels, the default “one-against-one” multiclass algorithm was used (Hsu & Lin, 2002). Nested 5-fold cross validation was performed separately for each subject. The inner folds were implemented for selection of parameters C and γ, and the outer folds were used to provide cross-
validated measures of accuracy. Because distributions of classification accuracy typically violate the assumptions of parametric tests, one-tailed Wilcoxon sign-rank tests were performed for group inference. To examine which cursor locations led to the prediction of each expression, one-sample t-tests were performed across subjects on SVM decision values for all coordinates in the grid.

5.2.7 Multi-voxel pattern classification

Decoding of neural activity was performed via PLS-DA (Wold et al., 2001) using the NIPALS algorithm (Martens & Næs, 1989) as implemented in the libPLS toolbox (http://www.libpls.net). This method was selected because it effectively reduces the dimensionality of data, decreasing chances of overfitting. Classification of multiple categories was performed using a winner-takes-all approach, wherein one class is classified against all others. Because this approach creates an uneven proportion of classes (1:5), a weighted approach was taken for discriminant analysis to minimize bias due to class imbalance. Input data (144 trials) were mean centered before conducting the analysis. The number of latent variables was fixed at 1, to reduce the complexity of the model and maximize the amount of data available for training and testing. Classification was performed separately for each subject, using 2-fold cross validation. Group inference on accuracy was performed using one-tailed Wilcoxon sign-rank tests (with
chance rates of 1/6), with FDR correction (Benjamini & Hochberg, 1995) for multiple comparisons when appropriate.

To assess the relationship between experienced emotion and the information content of neural activation patterns, the accuracy of classifying self-report was correlated with the accuracy of classifying fMRI data across subjects for each region of interest, using Pearson’s coefficient. Inference was performed for each region using the Student’s t-distribution (two-tailed), with FDR correction for multiple comparisons.

Comparisons of PLS regression coefficients within the postcentral gyrus were assessed using a one-way ANOVA. Contrasts were made between emotions associated with movement of facial muscles in the lower (happiness, surprise) versus upper (sadness, anger) portions of the face (Bassili, 1979; M. L. Smith, Cottrell, Gosselin, & Schyns, 2005) at the group level. An AR(1) correction was applied to adjust for departures from sphericity (independence and homogeneity of variance). Statistical maps were thresholded using a voxel-wise threshold of \( P < .05 \) and extent of 21 voxels, which was shown to control the false-positive rate at \( \alpha < .05 \) in Monte Carlo simulations (Forman et al., 1995). Renderings of classification weights for lower vs upper face expressions were mapped to flattened and inflated models of the cortical surface (Van Essen, 2005) in the right hemisphere using Caret (Van Essen et al., 2001).
5.3 Results

Participants made on-line ratings in response to happy, sad, surprised, angry, fearful, and neutral expressions. The response format included multiple emotion terms arranged in a wheel shape (Fig. 16a), with four levels of intensity indicated by circles of varying size. Participants made ratings by moving the cursor to a location on the screen that best indicated their current feelings.

We first tested whether participants experienced emotions consistent with those conveyed by facial and vocal expressions by constructing classification models to predict the emotional content of stimuli using cursor locations on every trial. Consistent with our hypothesis of behavioral mirroring, cursor positions spanning both facial and vocal trials demonstrated significant discrimination, with an accuracy of 40.1 ± 3.79% (mean ± s.e.m.), compared to chance levels of 16.67% (Wilcoxon sign-rank test, $z = 3.84$ $P = 1.22 \times 10^{-4}$).

Examination of decision values from the classifiers revealed that emotions were best predicted by ratings within focal regions of the self-report inventory (Fig. 16c), indicating participants experienced relatively discrete feelings in response to the facial and vocal stimuli. Furthermore, the location of these regions exhibited a one-to-one correspondence with the emotions conveyed by the stimuli (Appendix C.1). Together, these findings provide clear evidence that participants experienced emotions congruent with those perceived from the facial and vocal cues.
Figure 16: Experimental paradigm and behavioral results.

(a) Graphical depiction of a single trial in which participants are first presented a facial or vocal expression of emotion, followed by a fixation cross, and a response screen, which subjects used to indicate emotional experience by moving a cursor. (b) Scatterplot of behavioral responses for all participants, with each point corresponding to a single trial. Axes reflect cursor positions along horizontal and vertical dimensions of the screen, standardized within subjects. (c) Parametric maps (one sample t-test, n = 21) of SVM decision values for each emotion category, showing which coordinates lead to the prediction of each emotion. Cursors located in blue regions are evidence against the labeled category, whereas red regions indicate positively predictive regions.

To examine whether regional patterns of fMRI response discriminated among perceived emotions, we conducted multi-voxel pattern classification on data from brain regions implicated in a neural network hypothesized to be critical for the recognition of
emotion (Adolphs, 2002): postcentral gyrus in the right hemisphere (corresponding to primary somatosensory cortex), posterior superior temporal sulcus (pSTS), medial orbitofrontal cortex (mOFC), inferior frontal operculum (IFO), fusiform gyrus (FG), amygdala, and insula. Among these regions, decoding of emotional categories from perceptual cues was successful from patterns of activation in postcentral gyrus, mOFC, IFO, FG, and insula at accuracy levels significantly above chance (all $P_{\text{adj}} < .05$; Fig. 17).

Figure 17: Multi-voxel pattern classification of BOLD response to facial and vocal expressions of emotion.

(a) Regions of interest rendered on the group mean anatomical image (n = 21). (b) Patterns of response within right postcentral gyrus ($z = 3.21$, $P_{\text{adj}} = 0.0047$), insula ($z = 2.66$, $P_{\text{adj}} = 0.0136$), medial orbitofrontal cortex (mOFC; $z = 1.92$, $P_{\text{adj}} = 0.0384$), inferior frontal operculum (IFO; $z = 1.93$, $P_{\text{adj}} = 0.0384$), and fusiform gyrus (FG; $z = 2.43$, $P_{\text{adj}} = 0.0175$) were classified at levels greater than chance (Wilcoxon sign-rank test). Dashed line reflects chance levels of accuracy. * $P < .05$, FDR corrected. Error bars reflect s.e.m. pSTS = posterior Superior Temporal Sulcus.
Having established that patterns of fMRI activity within right somatosensory cortex predict the emotional content of facial and vocal expressions in a manner consistent with self-reported emotional experience, we next tested whether classification weights within this region followed somatotopic organization consistent with those of perceived emotions. Although the spatial resolution of fMRI is too coarse to directly sample neural activity sensitive to individual facial muscles, the overrepresentation of the lip and mouth regions in somatosensory cortex can be utilized to compare emotional expressions that differentially engage lower versus upper regions of the face.

Because prior research has shown that happiness and surprise contain more distinctive information in lower regions of the face, we contrasted expressions of happiness and surprise against those of fear and anger, which contain more distinguishing information in upper portions of the face (Bassili, 1979; M. L. Smith et al., 2005). This analysis revealed two clusters in lateral postcentral gyrus (Fig. 18); one cluster spanned Brodmann areas (BAs) 3, 1, and 2 adjacent to parietal operculum (MNI center of mass = 57, -6, 28; peak t = 3.15) while the other was restricted to BA2 (MNI center of mass = 40, -30, 46; peak t = 3.65). The localization of these clusters is consistent with studies localizing oral and facial (Eickhoff, Grefkes, Fink, & Zilles, 2008; Miyamoto et al., 2006) representation in somatosensory cortex.
The perception of expressions associated with lower portions of the face was predicted by greater activation in inferior regions of the postcentral gyrus, consistent with known somatotopic organization. Solid lines demarcate borders of Brodmann areas 3, 1 and 2. Inset of facial images convey portions of the face which are diagnostic of each expression.

Given that pattern classification is opportunistic in discriminating among brain states and may have been driven by factors other than experienced emotion *per se* (e.g. low-level stimulus properties or physiological arousal), we next tested whether individual differences in the accuracy of neural classification correlated with those of
self-report. We found that the degree to which individuals experienced distinct emotional states was specifically associated with the information content of patterns spanning the full extent of postcentral gyrus ($r = .5932, P_{adj} < 0.0321; P_{adj} > .2$ for all other regions; Appendix C.2). Such a correspondence establishes a direct link between the information content of somatosensory activity and self-reported feelings during the perception of facial and vocal expressions of emotion.

5.4 Discussion

Our results demonstrate that patterned activation within somatosensory cortex contains information sufficient for the decoding of perceived emotional categories. Such refined discrimination of non-tactile stimulation within somatosensory cortex runs contrary to the classic view that the region is a unimodal sensory area and suggests that visual and auditory signals modulate neural activity at early stages of cortical processing (Ghazanfar & Schroeder, 2006). Taken together, our findings expand the functional role of the somatosensory cortex and provide novel evidence that emotions are reflected partly in the brain’s representation of the body (Adolphs, 2002; A. R. Damasio, 1996; Niedenthal, 2007).

We found that emotion-predictive patterns within postcentral gyrus exhibited somatotopic organization, suggesting that information related to body states contributed to the decoding of emotional expressions. Further, this result is concordant with
evidence that emotions are associated with categorically distinct bodily sensations (Nummenmaa et al., 2014) and refutes the notion that dimensions of valence and arousal were the principal basis of perception (Russell, Bachorowski, & Fernandez-Dols, 2003). Topographically organized somatosensory activation has been documented during the observation of touch (Blakemore, Bristow, Bird, Frith, & Ward, 2005; Ebisch et al., 2008; Schaefer, Xu, Flor, & Cohen, 2009), during the observation of actions (Gazzola & Keysers, 2009), and during the perception of sounds (Gazzola, Aziz-Zadeh, & Keysers, 2006). Given that BA 2 has generally been implicated in processing proprioceptive information whereas areas 3 and 1 are more closely tied to tactile stimulation (Keysers et al., 2010), our localization of emotion-predictive patterns in all three areas suggests that a combination of tactile and proprioceptive information is simulated during the perception of emotional expressions.

Beyond predicting the emotional content of stimuli, we found that somatic representations of perceived emotions uniquely correlated with the extent of experiential mirroring across individuals. Our observation of experiential mirroring supports accounts that posit emotion-related knowledge is embodied in somatosensory cortices (Goldman & Sripada, 2005; Niedenthal, 2007) and is consonant with behavioral studies showing the perception of emotional expressions leads to facial mimicry and congruent emotional experience (Hess & Blairy, 2001) – collectively referred to as
emotional contagion. By linking such behavior to distinct patterns of somatic activity, we provide a mechanistic interpretation for studies showing that the region plays an essential role in emotion recognition (Adolphs et al., 2002; Adolphs et al., 2000; Banissy et al., 2010; Pitcher et al., 2008) that is consistent with the somatic marker hypothesis (A. R. Damasio, 1996), wherein representations of body states associated with distinct emotions contribute to cognitive processing.

In sum, our findings provide novel functional evidence that somatosensory cortex plays a critical role in linking feeling states and perceived emotions in others, forming an embodied basis for the social spread of emotions. Our methodological approach serves as a template for objectively quantifying somatic states associated with categorical emotions. Assaying somatic states during the disruption of facial muscle activity (e.g., Hennenlotter et al., 2009) could establish whether peripheral feedback is essential in producing the observed effects, or whether centrally generated representations of body states are sufficient. The frequency and separability of somatic states could further be quantified during live social interactions (Redcay et al., 2010), to characterize their occurrence in more ecological settings. Future studies in these areas are necessary to characterize the role embodied emotions play in social interactions.
6. Discussion

The series of experiments comprising this dissertation have explored how emotional states are represented in terms of subjective self-report, peripheral indices of autonomic activation, and central nervous system activity sampled through blood oxygen-level-dependent fMRI. As reviewed in Chapter 1, these studies integrate recent methodological advances in machine-learning (Pereira, Mitchell, & Botvinick, 2009) and representational analysis (Kriegeskorte, Mur, & Bandettini, 2008) to adjudicate between dimensional and categorical models. This chapter assesses how research utilizing multivariate pattern classification has advanced our understanding of emotion, synthesizes this evidence into a neural model of emotion representation, and examines its theoretical and practical implications.

6.1 Converging evidence on the neural basis of emotion

An emerging body of research is beginning to reveal the correspondence between nervous system activity and distinct emotion categories using pattern classification based methods. Multivariate classification of autonomic system activity can successfully decode multiple distinct emotional states, yet the information content of autonomic activity is not principally organized by valence and arousal (Chapters 1 and 2). In studies examining central nervous system activity, researchers have demonstrated
that distinct patterns of fMRI activation can be classified along multiple emotion
categories, concordant with the findings presented in Chapter 3.

In one such study (Saarimäki et al., 2015), states of disgust, fear, happiness,
anger, surprise, and sadness were elicited using both movies and mental imagery during
fMRI scanning. Patterns of BOLD response were classified along the six categories using
a linear neural network with levels of accuracy significantly above chance (34% versus
20% and 23% versus 16.7% for movies and imagery, when using subject-independent
classification). Resultant importance maps showed that lateral prefrontal cortices, frontal
pole, precentral and postcentral gyri, precuneus, and posterior cingulate cortex
contributed to classification, although these results should be considered preliminary as
no inferential tests were performed to substantiate their significance.

Other work by Wager and colleagues (2015) used Bayesian spatial point process
models to classify brain states in a meta-analysis of 148 studies eliciting happiness, fear,
anger, sadness, and disgust. The model classified brain states at levels well above chance
(66% versus 20%), and produced generative models which estimate brain regions are
likely to be active (termed population centers) during the elicitation of the five emotions.
Generally, the models predicted activation spanning much of the brain, in both cortical
and subcortical structures; for instance, several brain regions predicted multiple emotion
categories, including amygdala, ventral striatum, orbitofrontal cortex, insula, and anterior cingulate.

Although these studies differ in their induction procedures, data acquisition, preprocessing and analytical methods, as well as the type of emotions being classified, they exhibit some striking similarities. The findings reported in Chapter 3 and by Wager and colleagues (2015) demonstrate a correspondence between predictions of fear and activation within subgenual anterior cingulate, insula, amygdala, and posterior midline regions (bordering posterior cingulate and precuneus). Further, states of happiness (or amusement) were classified on the basis of activation in temporal pole, middle and superior temporal gyrus, insula, precentral gyrus, and occipital gyrus. Sad states were predicted by activity within middle and superior temporal gyrus, and postcentral gyrus. Classification of anger was similarly predicted by activation within supplementary motor area, precuneus, superior temporal gyrus, and superior temporal lobe near the temporal-parietal junction. Generally, these commonalities suggest that distinct sets of brain regions likely reflect generalizable aspects of information processing that are linked to specific emotion categories.

One shared finding that is particularly relevant to models of emotion representation is that classification errors were not organized along dimensions of valence and arousal in any of the studies. In Wager et al. (2015), the majority of
classification errors were false-alarms to fear, most likely due to the overrepresentation of that emotion within their meta-analytic database. Confusions between emotions which were similar in terms of valence and arousal were relatively low compared to those that differed along such dimensions (e.g., misclassification between disgust and anger occurred with comparable frequency to that of happy and sad brain states). In Saarimäki et al., classification errors were distributed uniformly across emotions, suggestive of categorical representation. These findings are consonant with recent evidence demonstrating that the valence of emotional stimuli cannot be clearly classified on the basis of neural activity (Lindquist, Satpute, Wager, Weber, & Barrett, 2015). Thus, valence and arousal are not likely the primary determinants of organizing brain activity during emotional events, despite playing a central role in subjective experience.

Thus, of the many factors contributing to emotional experience (for a review, see Lambie & Marcel, 2002), it appears that categorical and dimensional aspects have separable neural bases. Discrete emotional experiences do not emerge solely from systems dedicated to processing valence and arousal (Lang & Bradley, 2010), but are rather represented in the activity of multiple separable neural systems. The remainder of the chapter integrates these findings into a model of emotion representation and discusses the predictions and implications of the model along with directions for future research.
6.2 A neural model of emotion representation

The affective dimensions of valence and arousal are ubiquitous in semantic and experiential measures of emotion, yet contribute little to the differentiation of discrete emotional states as evident in nervous system activity. If these constructs do not principally organize nervous system activity underlying emotional experiences, how do they emerge as the basis of cognitive representations of emotion?

Following sensory and perceptual processing, environmental stimuli elicit distributed responses in central and autonomic branches of the nervous system. These states are probabilistically linked to categorical aspects of emotion experience. Through valuation, these categorically distinct states can be related and experienced in terms of valence and arousal.

The solution presented here centers on two proposals. First, emotional experiences result from cognitive processing of neural ‘raw materials’ (LeDoux, 2014), which are reflected in the distributed activity of multiple brain systems. It is the combined nature of these materials that makes different types of emotional experience
unique and different from non-emotional experiences. Second, valence and arousal serve
to organize different kinds of emotional experiences when they are assigned value
(Dolan, 2002), a contextually dependent process known to be mediated by the
orbitofrontal cortex and amygdala (Dolan, 2007; Kable & Glimcher, 2007; D. V. Smith et
al., 2010; Winecoff et al., 2013). Thus, emotional events that are categorically distinct can
be valued quite similarly, and consequently be experienced similarly in terms of valence
and arousal, while instances of the same emotion can differ considerably in their
emotional value, depending on the context and learning history.

6.2.1 Distributed neural representations of emotion categories

This model proposes that activity within multiple brain systems contains
information which contributes to the categorical nature of emotional experience,
independent of valence and arousal (see Figure 19). When activity across these multiple
systems is configured in a prototypical way (e.g., reflecting the output of an affect
program (Tomkins, 1962), or spontaneously entering such a state), there is a high
probability of experiencing emotion in a discrete, categorical manner. Findings from the
empirical studies presented in this dissertation provide evidence supporting this aspect
of the model.

The examination of peripheral responding presented in Chapter 2 revealed that
distinct autonomic patterns are associated with categorical, but not dimensional, aspects
of emotional experience. In agreement with the burgeoning field of research on the emotion specificity of autonomic patterning (Christie & Friedman, 2004; Kolodyazhniy et al., 2011; Kreibig et al., 2007; Nyklicek et al., 1997; Rainville et al., 2006; Stephens et al., 2010), this finding supports the model’s proposal that patterns of bodily responses to emotional events contain information that is probabilistically linked to the experience of specific emotions.

The fMRI experiment presented in Chapter 3 formally assessed the representational content of central neural activation underlying emotional experiences using Bayesian model comparison. This analysis provided strong evidence that activity in separable brain systems is associated with categorically distinct emotional experiences, a finding consistent with recent studies classifying the categorical content of emotional stimuli to multi-voxel patterns of BOLD response (Saarimäki et al., 2015; T. D. Wager et al., 2015). Related work presented in Chapter 4 demonstrated that similar brain states emerge during wakeful rest, that their spontaneous occurrence tracks individual differences in mood and personality, and that their expression is congruent with self-report assayed using online experience sampling. Together, these findings support the proposal that brain states associated with categorically distinct emotional are an intrinsic property of the brain and are relatively independent of contextual effects.
Additional support for categorical representation of emotions comes from the study on emotional contagion presented in Chapter 5. This study revealed that the perception of facial and vocal expressions of emotion yields congruent emotional experience that is linked to distinct patterns of somatosensory activation. In linking category-specific emotional experience to the brain’s representation of the body, this finding illustrates one way in which sensory representations can drive categorical experience through mechanisms independent of valence and arousal.

6.2.2 The emergence of valence and arousal

Beyond quantifying categorical aspects of emotional experience, the studies comprising this dissertation were designed to additionally sample more basic affective dimensions such as valence and arousal. Congruent with a large literature on the meaning of emotion terms and the subjective experience of emotion (Lisa Feldman Barrett & Russell, 1999; Bush, 1973), behavioral data presented in Chapters 2 and 3 show that discrete emotional events are well-characterized in terms of valence and arousal. Further, self-report data presented in Chapters 4 and 5 acquired using the Geneva Emotion Wheel (Scherer, 2005) exhibited categorical clustering organized along dimensions of valence and control. Broadly, these results demonstrate that different kinds of emotional experiences can be organized along dimensions that span multiple
categories, in accordance with theories proposing cognitive representations of emotion (e.g., Rolls, 1990; Russell, 1980; Scherer, 1984).

In support of the proposal that the experience of valence and arousal are driven in part via evaluative processing, univariate analysis of fMRI data presented in Chapter 3 additionally showed that amygdala activity was positively associated with subjective ratings of arousal. This finding is in agreement with a large body of fMRI research (A. K. Anderson et al., 2003; Lewis et al., 2007; Small et al., 2003) and is consistent with the notion that the region is generally more active during states with high emotional value, either highly positive or negative. Although contrasts of activity of high versus low valence did not survive whole-brain correction for multiple comparison within orbitofrontal cortex, a number of clusters exhibited positive associations ($P < .05, k = 5$ voxels). Taken together with evidence from the literature on motivation and reward learning (for a review, see Berridge & Kringelbach, 2008), these findings corroborate the proposal that valence and arousal are linked to local processing within orbitofrontal cortex and amygdala.

**6.2.3 Predictions and implications**

An important prediction made by the model is that representations of discrete emotions can acquire value (in terms of valence and arousal, and possibly other dimensions) through learning. Over the long term, neural states that are bound to the
experience of a specific emotion will become associated with a particular value through interactions with the amygdala and orbitofrontal cortex. For instance, threats in the environment elicit neural states similar to one another and are generally experienced as fear. Generally, these encounters are unpleasant (low valence) and elicit behavioral responses and prompt actions that require considerable energy (high arousal). Thus, the model predicts the common finding that emotion words can be efficiently organized in terms of valence and arousal because they reliably associated with certain values of valence and arousal.

Importantly, however, the model does not predict that valence and arousal are the principal basis of emotional experience. While these factors do serve as an efficient means of organizing diverse emotional states, they are not essential for experiencing any particular emotion. Disrupting function of brain regions implicated in value representation would thus have relatively little impact on the experience of emotion (A. K. Anderson & Phelps, 2002; but see Feinstein, Adolphs, Damasio, & Tranel, 2011).

An implication of the model is that discrete emotions do not consistently map onto either positive or negative affect. This assertion is congruent with evidence showing that winning gambles can be simultaneously experienced as both good and bad (Larsen, McGraw, Mellers, & Cacioppo, 2004). Other examples of atypical emotional experiences (Wilson-Mendenhall, Barrett, & Barsalou, 2015) include instances of pleasant
fear (e.g., riding a roller-coaster or performing before a crowd) and unpleasant happiness (e.g., exhaustion after completing a challenging task). While some accounts predict these experiences occur when states of valence and arousal are conceptualized as an emotion category (L. F. Barrett, 2006b), the present model suggests that all instances of each emotion category share common features, but their experience as being pleasant or unpleasant can vary depending on the context.

In sum, the proposed model offers a conceptual framework for understanding how emotions are represented in nervous system activity, and experienced in terms of specific categories and broad dimensions. The multivariate methods used in this body of work serve as a blueprint for understanding the types of emotional experiences that can be identified and the accuracy with which they can be related to subjective experience – potentially advancing theoretical models of emotion in addition to serving as a practical tool for clinical research.

6.3 Conclusion and future directions

In conclusion, emotions are categorically represented in central and peripheral nervous system activity. The diversity of events that individuals experience as discrete emotions are not clearly differentiated by simple models of valence and arousal, highlighting the need to integrate categorical and dimensional frameworks. Further, machine-learning based approaches provide an objective way to quantify personal,
subjective mental states in biological data, offering new insight into decades old debates on the nature of emotion and providing a tool for understanding and treating mood and affective disorders.

While the studies presented in this dissertation provide clear evidence regarding categorical and dimensional organization of emotional states in nervous system activity, a number of important theoretical questions remain unresolved. Given that the experiments presented examined the surface features or family resemblance (Rosch & Mervis, 1975) of emotional responses and were correlational in nature, additional work using causal manipulations (e.g., transcranial magnetic stimulation, pharmacological intervention, or emotion regulation) will be necessary to determine whether emotional experiences depend on patterned nervous system activity. Because some theories propose certain emotions are evolutionary in origin (D. J. Anderson & Adolphs, 2014; A. Damasio & Carvalho, 2013b; Ekman & Cordaro, 2011; J. Panksepp, 1982; Robert Plutchik & Kellerman, 1980), comparative work examining emotional responding in primates and other mammals (see, e.g., Kriegeskorte, Mur, Ruff, et al., 2008 for a comparison of object representation in inferior temporal cortex) would provide further leverage regarding the nature of emotion. Finally, the majority of work in this dissertation focused on the most prominent dimensions of emotion, namely valence and arousal; given evidence that emotional experience can reliably be characterized using additional factors (Fontaine et
future work examining additional dimensions, such as approach-withdrawal or potency-control, may provide more accurate characterization of subjective experience. Answering these questions will further elucidate how emotions are reflected in nervous system activity.

In addition to motivating basic research on the nature of emotions, the findings presented in this dissertation have practical implications. Given the observation that emotional states occur spontaneously, the incidence of emotional states during resting state scans could serve as, or complement, biomarkers for mood and affective disorders (Fu et al., 2008; Zeng et al., 2012). As an example, if individuals with PTSD or depression are found to have reliable differences in the frequency of spontaneous fearful and or sad brain states, these neural biomarkers of emotional states could serve as a tool for monitoring treatment outcomes, with the goal of ultimately leading to better patient care. Furthermore, novel biomarkers for mood disorders with more specific experiential symptoms (e.g., social anxiety disorder or specific phobias) could be developed using a similar approach. Thus, objectively quantifying emotional states in nervous system activity can both advance our theoretical understanding of what emotions are and practically improve treatment of mental health disorders.
Figure A.1: Confusion matrices for seven-way classification using different model selection and cross-validation methods. Columns indicate the number of latent variables used in classification, and rows correspond to different cross validation schema (either standard 5-fold or subject independent 8-fold cross validation). The colormap indicates the number of trials in each cell (each emotion totaled 128 trials), and is equivalent to a proportion of 0% to 46.875%. The mean accuracy and 95% confidence interval are indicated above each confusion matrix, and parameter estimates (and 95% confidence intervals) for dimensional and categorical terms are indicated below each matrix. Note that average performance and the structure of errors are generally robust to changes in analysis methods.
Figure A.2: Scatterplot and histograms depicting the intensity range of single-trial parameters (β estimates) used as inputs for classification versus partial least squares (PLS) regression coefficients for the classification of content, amusement, surprise, and neutral trials. Each point corresponds to an individual voxel. The majority of input β estimates are within one order of magnitude from each other (suggesting additional scaling is not needed). With the exception of a small number of very low amplitude voxels, the amplitude of PLS regression coefficients does not depend on the range of input β estimates.
Figure A.3: Scatterplot and histograms depicting the intensity range of single-trial parameters (β estimates) used as inputs for classification versus partial least squares (PLS) regression coefficients for the classification of fear, anger, and sad trials. Each point corresponds to an individual voxel. The majority of input β estimates are within one order of magnitude from each other (suggesting additional scaling is not needed). With the exception of a small number of very low amplitude voxels, the amplitude of PLS regression coefficients does not depend on the range of input β estimates.
Table A.4: Peak regression coefficients for activation patterns predicting contentment.

<table>
<thead>
<tr>
<th>Extent (ml)</th>
<th>Z</th>
<th>Z X MNI X</th>
<th>MNI Y</th>
<th>MNI Z</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.96</td>
<td>3.72</td>
<td>21</td>
<td>-40</td>
<td>-16</td>
<td>R</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td></td>
</tr>
<tr>
<td>3.43</td>
<td>25</td>
<td>-44</td>
<td>-8</td>
<td></td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>Brodmann area 36</td>
</tr>
<tr>
<td>3.16</td>
<td>25</td>
<td>-36</td>
<td>-4</td>
<td></td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>Brodmann area 27</td>
</tr>
<tr>
<td>1.54</td>
<td>3.72</td>
<td>-28</td>
<td>-44</td>
<td>-12</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>Brodmann area 36</td>
</tr>
<tr>
<td>3.29</td>
<td>-25</td>
<td>-36</td>
<td>-16</td>
<td></td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>Brodmann area 36</td>
</tr>
<tr>
<td>2.74</td>
<td>-28</td>
<td>-36</td>
<td>-8</td>
<td></td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>Hippocampus</td>
</tr>
<tr>
<td>5.60</td>
<td>3.72</td>
<td>-13</td>
<td>-51</td>
<td>11</td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 30</td>
</tr>
<tr>
<td>3.72</td>
<td>17</td>
<td>-63</td>
<td>18</td>
<td></td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 30</td>
</tr>
<tr>
<td>3.72</td>
<td>6</td>
<td>-47</td>
<td>18</td>
<td></td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 30</td>
</tr>
<tr>
<td>6.97</td>
<td>3.72</td>
<td>48</td>
<td>-21</td>
<td>34</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Postcentral Gyrus</td>
<td>Brodmann area 2</td>
</tr>
<tr>
<td>3.72</td>
<td>59</td>
<td>-13</td>
<td>34</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Postcentral Gyrus</td>
<td>Brodmann area 2</td>
</tr>
<tr>
<td>3.54</td>
<td>40</td>
<td>-28</td>
<td>41</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Inferior Parietal Lobule</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td>4.44</td>
<td>3.72</td>
<td>-47</td>
<td>-17</td>
<td>41</td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Postcentral Gyrus</td>
<td>Brodmann area 3</td>
</tr>
<tr>
<td>3.24</td>
<td>-44</td>
<td>-9</td>
<td>30</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td>3.06</td>
<td>-55</td>
<td>-25</td>
<td>41</td>
<td></td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 2</td>
</tr>
<tr>
<td>2.52</td>
<td>3.72</td>
<td>6</td>
<td>-66</td>
<td>56</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td>3.72</td>
<td>13</td>
<td>-63</td>
<td>56</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td>3.24</td>
<td>17</td>
<td>-55</td>
<td>60</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td>0.99</td>
<td>3.54</td>
<td>-32</td>
<td>-9</td>
<td>15</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Claustrum</td>
<td>*</td>
</tr>
<tr>
<td>3.19</td>
<td>-36</td>
<td>-6</td>
<td>7</td>
<td></td>
<td>L</td>
<td>Sub-lobar</td>
<td>Claustrum</td>
<td>*</td>
</tr>
<tr>
<td>2.57</td>
<td>-44</td>
<td>-6</td>
<td>15</td>
<td></td>
<td>L</td>
<td>Sub-lobar</td>
<td>Insula</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td>1.59</td>
<td>3.54</td>
<td>40</td>
<td>-78</td>
<td>18</td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>3.35</td>
<td>32</td>
<td>-82</td>
<td>15</td>
<td></td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>1.81</td>
<td>3.54</td>
<td>2</td>
<td>-6</td>
<td>41</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 24</td>
</tr>
<tr>
<td>3.12</td>
<td>-6</td>
<td>6</td>
<td>34</td>
<td></td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 24</td>
</tr>
<tr>
<td>2.72</td>
<td>-6</td>
<td>13</td>
<td>30</td>
<td></td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 24</td>
</tr>
<tr>
<td>0.99</td>
<td>3.43</td>
<td>-40</td>
<td>-78</td>
<td>18</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Middle Temporal Gyrus</td>
<td>Brodmann area 39</td>
</tr>
<tr>
<td>2.56</td>
<td>-40</td>
<td>-78</td>
<td>30</td>
<td></td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Middle Temporal Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>2.35</td>
<td>-44</td>
<td>-82</td>
<td>11</td>
<td></td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>0.82</td>
<td>3.43</td>
<td>21</td>
<td>-63</td>
<td>41</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td>3.35</td>
<td>25</td>
<td>-66</td>
<td>34</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td>1.76</td>
<td>3.35</td>
<td>13</td>
<td>-70</td>
<td>-16</td>
<td>R</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
</tr>
<tr>
<td>3.29</td>
<td>13</td>
<td>-74</td>
<td>-8</td>
<td></td>
<td>R</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td>*</td>
</tr>
<tr>
<td>3.19</td>
<td>21</td>
<td>-74</td>
<td>-16</td>
<td></td>
<td>R</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
</tr>
<tr>
<td>3.40</td>
<td>3.35</td>
<td>6</td>
<td>-40</td>
<td>49</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td>3.24</td>
<td>6</td>
<td>-25</td>
<td>45</td>
<td></td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 31</td>
</tr>
<tr>
<td>3.04</td>
<td>-2</td>
<td>-36</td>
<td>45</td>
<td></td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 31</td>
</tr>
<tr>
<td>0.88</td>
<td>3.16</td>
<td>-17</td>
<td>-59</td>
<td>22</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 31</td>
</tr>
<tr>
<td>2.97</td>
<td>-17</td>
<td>-63</td>
<td>15</td>
<td></td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 30</td>
</tr>
<tr>
<td>2.51</td>
<td>-13</td>
<td>-70</td>
<td>22</td>
<td></td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Cuneus</td>
<td>Brodmann area 18</td>
</tr>
<tr>
<td>1.54</td>
<td>3.12</td>
<td>6</td>
<td>-89</td>
<td>-1</td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Lingual Gyrus</td>
<td>Brodmann area 18</td>
</tr>
<tr>
<td>Extent (ml)</td>
<td>Z</td>
<td>MNI&lt;sub&gt;x&lt;/sub&gt;</td>
<td>MNI&lt;sub&gt;y&lt;/sub&gt;</td>
<td>MNI&lt;sub&gt;z&lt;/sub&gt;</td>
<td>Hemisphere</td>
<td>Lobe</td>
<td>Gyrus</td>
<td>Region</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>13.77</td>
<td>3.72</td>
<td>-51</td>
<td>6</td>
<td>-23</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Middle Temporal Gyrus</td>
<td>Brodmann area 38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
<td></td>
</tr>
<tr>
<td>13.61</td>
<td>3.72</td>
<td>63</td>
<td>-25</td>
<td>-4</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Middle Temporal Gyrus</td>
<td>Brodmann area 22</td>
<td></td>
</tr>
<tr>
<td>1.26</td>
<td>3.72</td>
<td>-2</td>
<td>-21</td>
<td>15</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Thalamus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td>3.16</td>
<td>3.72</td>
<td>-6</td>
<td>-9</td>
<td>11</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Thalamus</td>
<td>Med. Dorsal Nucleus</td>
</tr>
<tr>
<td>2.14</td>
<td>3.72</td>
<td>51</td>
<td>-2</td>
<td>45</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Middle Frontal Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>1.32</td>
<td>3.72</td>
<td>-40</td>
<td>2</td>
<td>41</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Occipital Lobe</td>
<td>Cuneus</td>
<td>Brodmann area 17</td>
</tr>
<tr>
<td>5.38</td>
<td>3.54</td>
<td>-6</td>
<td>82</td>
<td>11</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Cuneus</td>
<td>Brodmann area 18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gyrus</td>
<td>Brodmann area 18</td>
<td></td>
</tr>
<tr>
<td>3.06</td>
<td>3.43</td>
<td>-6</td>
<td>-85</td>
<td>3</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Lingual Gyrus</td>
<td>Brodmann area 21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gyrus</td>
<td>Brodmann area 38</td>
<td></td>
</tr>
<tr>
<td>2.95</td>
<td>3.43</td>
<td>-6</td>
<td>85</td>
<td>22</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Cuneus</td>
<td>Brodmann area 38</td>
</tr>
<tr>
<td>0.82</td>
<td>3.43</td>
<td>51</td>
<td>2</td>
<td>-39</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Middle Temporal Gyrus</td>
<td>Brodmann area 21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 38</td>
<td></td>
</tr>
<tr>
<td>3.43</td>
<td>3.43</td>
<td>48</td>
<td>10</td>
<td>-39</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 38</td>
</tr>
<tr>
<td>2.33</td>
<td>3.43</td>
<td>51</td>
<td>13</td>
<td>-23</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 38</td>
</tr>
</tbody>
</table>

Table A.5: Peak regression coefficients for activation patterns predicting amusement
<table>
<thead>
<tr>
<th>Extent (ml)</th>
<th>Z</th>
<th>MNI_x</th>
<th>MNI_y</th>
<th>MNI_z</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.30</td>
<td>3.72</td>
<td>-17</td>
<td>-47</td>
<td>-12</td>
<td>L</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td>* Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.72</td>
<td>-32</td>
<td>-47</td>
<td>-12</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Fusiform Gyrus</td>
<td>37 Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.72</td>
<td>29</td>
<td>-40</td>
<td>-12</td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>36 Brodmann area</td>
</tr>
<tr>
<td>1.32</td>
<td>3.72</td>
<td>-25</td>
<td>36</td>
<td>-12</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Inferior Frontal Gyrus</td>
<td>47 Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.09</td>
<td>-25</td>
<td>29</td>
<td>-20</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Middle Frontal Gyrus</td>
<td>11 Brodmann area</td>
</tr>
<tr>
<td></td>
<td>2.56</td>
<td>-17</td>
<td>44</td>
<td>-16</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Anterior Cingulate</td>
<td>32 Brodmann area</td>
</tr>
<tr>
<td>4.12</td>
<td>3.72</td>
<td>36</td>
<td>25</td>
<td>-4</td>
<td>R</td>
<td>Sub-lobar Insula</td>
<td></td>
<td>13 Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.54</td>
<td>29</td>
<td>17</td>
<td>-8</td>
<td>R</td>
<td>Sub-lobar Lentiform Nucleus</td>
<td>Putamen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.54</td>
<td>32</td>
<td>21</td>
<td>3</td>
<td>R</td>
<td>Sub-lobar Claustrium</td>
<td></td>
<td>* Brodmann area</td>
</tr>
<tr>
<td>4.39</td>
<td>3.72</td>
<td>-32</td>
<td>21</td>
<td>7</td>
<td>L</td>
<td>Sub-lobar Insula</td>
<td></td>
<td>13 Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.54</td>
<td>-32</td>
<td>17</td>
<td>-8</td>
<td>L</td>
<td>Sub-lobar Claustrium</td>
<td></td>
<td>* Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.24</td>
<td>-40</td>
<td>10</td>
<td>3</td>
<td>L</td>
<td>Sub-lobar Insula</td>
<td></td>
<td>13 Brodmann area</td>
</tr>
<tr>
<td>21.95</td>
<td>3.72</td>
<td>6</td>
<td>32</td>
<td>26</td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>32 Brodmann area</td>
</tr>
<tr>
<td></td>
<td>3.72</td>
<td>-6</td>
<td>-9</td>
<td>30</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>23 Brodmann area</td>
</tr>
</tbody>
</table>

Table A.6: Peak regression coefficients for activation patterns predicting surprise.
<table>
<thead>
<tr>
<th>Extent (ml)</th>
<th>Z</th>
<th>MNIx</th>
<th>MNIy</th>
<th>MNIz</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.72</td>
<td>2</td>
<td>21</td>
<td>30</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area</td>
<td></td>
</tr>
<tr>
<td>2.19</td>
<td>3.72</td>
<td>44</td>
<td>-70</td>
<td>18</td>
<td>R</td>
<td>Occipital Gyrus</td>
<td>Brodmann area</td>
<td></td>
</tr>
<tr>
<td>3.35</td>
<td>29</td>
<td>-74</td>
<td>22</td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Precuneus</td>
<td>Brodmann area</td>
<td></td>
</tr>
<tr>
<td>2.72</td>
<td>36</td>
<td>-78</td>
<td>22</td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area</td>
<td></td>
</tr>
<tr>
<td>2.69</td>
<td>3.72</td>
<td>44</td>
<td>10</td>
<td>22</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Inferior Frontal Gyrus</td>
<td>Brodmann area 9</td>
</tr>
<tr>
<td>3.35</td>
<td>36</td>
<td>10</td>
<td>26</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>3.16</td>
<td>48</td>
<td>13</td>
<td>15</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Parahippocampal Gyrus</td>
<td>Brodmann area 36</td>
<td></td>
</tr>
<tr>
<td>1.54</td>
<td>3.72</td>
<td>36</td>
<td>-2</td>
<td>45</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Middle Frontal Gyrus</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td>3.72</td>
<td>40</td>
<td>-2</td>
<td>53</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>1.98</td>
<td>3.54</td>
<td>-6</td>
<td>-70</td>
<td>-27</td>
<td>L</td>
<td>Anterior Lobe Nodule</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>2.95</td>
<td>10</td>
<td>-74</td>
<td>-23</td>
<td>R</td>
<td>Lobe</td>
<td>Declive</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>2.55</td>
<td>6</td>
<td>-66</td>
<td>-20</td>
<td>R</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>5.10</td>
<td>3.54</td>
<td>-28</td>
<td>-74</td>
<td>26</td>
<td>L</td>
<td>Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 31</td>
</tr>
<tr>
<td>3.29</td>
<td>-44</td>
<td>-82</td>
<td>15</td>
<td>L</td>
<td>Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 19</td>
<td></td>
</tr>
<tr>
<td>3.29</td>
<td>-36</td>
<td>-85</td>
<td>22</td>
<td>L</td>
<td>Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 19</td>
<td></td>
</tr>
<tr>
<td>3.43</td>
<td>-40</td>
<td>-6</td>
<td>45</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>3.29</td>
<td>-40</td>
<td>6</td>
<td>30</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>2.97</td>
<td>-44</td>
<td>-6</td>
<td>53</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 4</td>
<td></td>
</tr>
<tr>
<td>1.21</td>
<td>3.35</td>
<td>51</td>
<td>-47</td>
<td>18</td>
<td>R</td>
<td>Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td>2.74</td>
<td>55</td>
<td>-47</td>
<td>11</td>
<td>R</td>
<td>Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
<td></td>
</tr>
<tr>
<td>1.98</td>
<td>3.19</td>
<td>6</td>
<td>-17</td>
<td>3</td>
<td>R</td>
<td>Sub-lobar Thalamus</td>
<td>Nucleus</td>
<td></td>
</tr>
<tr>
<td>2.71</td>
<td>-6</td>
<td>-6</td>
<td>3</td>
<td>L</td>
<td>Sub-lobar Thalamus</td>
<td>*</td>
<td>Med. Dorsal Nucleus</td>
<td></td>
</tr>
<tr>
<td>2.65</td>
<td>-6</td>
<td>-17</td>
<td>7</td>
<td>L</td>
<td>Sub-lobar Thalamus</td>
<td>*</td>
<td>Med. Dorsal Nucleus</td>
<td></td>
</tr>
<tr>
<td>0.82</td>
<td>3.06</td>
<td>25</td>
<td>10</td>
<td>-4</td>
<td>R</td>
<td>Sub-lobar Lentiform Nucleus</td>
<td>Putamen</td>
<td></td>
</tr>
<tr>
<td>2.69</td>
<td>32</td>
<td>-6</td>
<td>-8</td>
<td>R</td>
<td>Sub-lobar</td>
<td>Lentiform Nucleus</td>
<td>Putamen</td>
<td></td>
</tr>
</tbody>
</table>

Table A.7: Peak regression coefficients for activation patterns predicting fear.
Table A.8: Peak regression coefficients for activation patterns predicting anger.

<table>
<thead>
<tr>
<th>Extent (ml)</th>
<th>Peak Z</th>
<th>MNI_x</th>
<th>MNI_y</th>
<th>MNI_z</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.74</td>
<td>3.72</td>
<td>44</td>
<td>-13</td>
<td>-4</td>
<td>R</td>
<td>Sub-lobar</td>
<td>Insula</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td>3.72</td>
<td>44</td>
<td>-21</td>
<td>-1</td>
<td>R</td>
<td>Sub-lobar</td>
<td>Temporal</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td>3.72</td>
<td>51</td>
<td>-17</td>
<td>3</td>
<td>R</td>
<td>Temporal</td>
<td>Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td>5.60</td>
<td>3.72</td>
<td>-51</td>
<td>-21</td>
<td>3</td>
<td>L</td>
<td>Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td>3.43</td>
<td>-40</td>
<td>-21</td>
<td>-1</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Claustrum</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>3.35</td>
<td>-47</td>
<td>-6</td>
<td>-4</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Insula</td>
<td>Brodmann area 13</td>
<td></td>
</tr>
<tr>
<td>0.93</td>
<td>3.54</td>
<td>25</td>
<td>21</td>
<td>41</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Superior Frontal Gyrus</td>
<td>Brodmann area 8</td>
</tr>
<tr>
<td>2.30</td>
<td>21</td>
<td>25</td>
<td>49</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Superior Frontal Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>1.04</td>
<td>3.54</td>
<td>-25</td>
<td>2</td>
<td>53</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Sub-Gysral</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td>3.12</td>
<td>-13</td>
<td>10</td>
<td>56</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Medial Frontal Gyrus</td>
<td>Brodmann area 6</td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>3.43</td>
<td>-25</td>
<td>-82</td>
<td>-35</td>
<td>L</td>
<td>Posterior Lobe</td>
<td>Pyramis</td>
<td>*</td>
</tr>
<tr>
<td>2.78</td>
<td>-36</td>
<td>-74</td>
<td>-35</td>
<td>L</td>
<td>Posterior Lobe</td>
<td>Pyramis</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
Table A.9: Peak regression coefficients for activation patterns predicting sadness.

<table>
<thead>
<tr>
<th>Extent (ml)</th>
<th>Peak Z</th>
<th>MNI X</th>
<th>MNI Y</th>
<th>MNI Z</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.57</td>
<td>3.72</td>
<td>-17</td>
<td>-59</td>
<td>-16</td>
<td>L</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>3.35</td>
<td>-36</td>
<td>-59</td>
<td>-20</td>
<td>L</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>3.29</td>
<td>-13</td>
<td>-55</td>
<td>-23</td>
<td>L</td>
<td>Anterior Lobe</td>
<td></td>
<td>Dentate</td>
</tr>
<tr>
<td>1.65</td>
<td>3.72</td>
<td>-44</td>
<td>-78</td>
<td>-12</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Fusiform Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td></td>
<td>2.97</td>
<td>-44</td>
<td>-85</td>
<td>-4</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Inferior Occipital Gyrus</td>
<td>Brodmann area 18</td>
</tr>
<tr>
<td></td>
<td>2.86</td>
<td>-47</td>
<td>-78</td>
<td>-4</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Inferior Occipital Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>6.69</td>
<td>3.72</td>
<td>55</td>
<td>-6</td>
<td>-1</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td></td>
<td>3.43</td>
<td>59</td>
<td>-21</td>
<td>-1</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 21</td>
</tr>
<tr>
<td></td>
<td>3.04</td>
<td>51</td>
<td>-9</td>
<td>11</td>
<td>R</td>
<td>Sub-lobar</td>
<td>Insula</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td>4.61</td>
<td>3.54</td>
<td>47</td>
<td>-2</td>
<td>-8</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Insula</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td></td>
<td>3.54</td>
<td>-51</td>
<td>2</td>
<td>-1</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td></td>
<td>3.35</td>
<td>-55</td>
<td>-9</td>
<td>-1</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td>2.19</td>
<td>3.54</td>
<td>10</td>
<td>-63</td>
<td>-4</td>
<td>R</td>
<td>Anterior Lobe</td>
<td>Culmen</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>3.04</td>
<td>6</td>
<td>-63</td>
<td>7</td>
<td>R</td>
<td>Anterior Lobe</td>
<td>Culmen of Vermis</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>2.69</td>
<td>13</td>
<td>-66</td>
<td>11</td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 30</td>
</tr>
<tr>
<td>1.98</td>
<td>3.54</td>
<td>-2</td>
<td>-59</td>
<td>45</td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td></td>
<td>2.59</td>
<td>-9</td>
<td>-70</td>
<td>18</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 30</td>
</tr>
</tbody>
</table>
Table A.10: Peak regression coefficients for activation patterns predicting neutral.

<table>
<thead>
<tr>
<th>Extent (ml)</th>
<th>Peak Z</th>
<th>MNI_x</th>
<th>MNI_y</th>
<th>MNI_z</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.65</td>
<td>3.72</td>
<td>-2</td>
<td>-82</td>
<td>-4</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Lingual Gyrus</td>
<td>Brodmann area 18</td>
</tr>
<tr>
<td>3.43</td>
<td>17</td>
<td>-74</td>
<td>-8</td>
<td></td>
<td>R</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
</tr>
<tr>
<td>3.24</td>
<td>21</td>
<td>-66</td>
<td>-4</td>
<td></td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Lingual Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>0.82</td>
<td>3.72</td>
<td>-21</td>
<td>-89</td>
<td>15</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Cuneus</td>
<td>Brodmann area 17</td>
</tr>
<tr>
<td>2.85</td>
<td>-21</td>
<td>-97</td>
<td>22</td>
<td></td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 18</td>
</tr>
<tr>
<td>5.49</td>
<td>3.72</td>
<td>-51</td>
<td>-55</td>
<td>30</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 39</td>
</tr>
<tr>
<td>3.54</td>
<td>-36</td>
<td>-66</td>
<td>41</td>
<td></td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 39</td>
</tr>
<tr>
<td>3.12</td>
<td>-51</td>
<td>-40</td>
<td>45</td>
<td></td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Inferior Parietal Lobule</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td>1.98</td>
<td>3.72</td>
<td>-32</td>
<td>36</td>
<td>26</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Middle Frontal Gyrus</td>
<td>Brodmann area 9</td>
</tr>
<tr>
<td>3.54</td>
<td>-40</td>
<td>29</td>
<td>26</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Middle Frontal Gyrus</td>
<td>Brodmann area 9</td>
</tr>
<tr>
<td>2.83</td>
<td>-25</td>
<td>32</td>
<td>34</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Middle Frontal Gyrus</td>
<td>Brodmann area 8</td>
</tr>
<tr>
<td>11.52</td>
<td>3.72</td>
<td>44</td>
<td>-40</td>
<td>37</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Supramarginal Gyrus</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td>3.72</td>
<td>40</td>
<td>-40</td>
<td>45</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Inferior Parietal Lobule</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td>3.72</td>
<td>51</td>
<td>-32</td>
<td>45</td>
<td></td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Inferior Parietal Lobule</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td>1.59</td>
<td>3.72</td>
<td>-51</td>
<td>-9</td>
<td>49</td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Postcentral Gyrus</td>
<td>Brodmann area 3</td>
</tr>
<tr>
<td>3.54</td>
<td>-55</td>
<td>-13</td>
<td>37</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Postcentral Gyrus</td>
<td>Brodmann area 4</td>
</tr>
<tr>
<td>2.25</td>
<td>-44</td>
<td>-17</td>
<td>49</td>
<td></td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Postcentral Gyrus</td>
<td>Brodmann area 3</td>
</tr>
<tr>
<td>2.96</td>
<td>3.54</td>
<td>-55</td>
<td>10</td>
<td>3</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 44</td>
</tr>
<tr>
<td>3.43</td>
<td>-47</td>
<td>21</td>
<td>-8</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Inferior Frontal Gyrus</td>
<td>Brodmann area 47</td>
</tr>
<tr>
<td>3.35</td>
<td>-47</td>
<td>13</td>
<td>7</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 44</td>
</tr>
<tr>
<td>1.92</td>
<td>3.35</td>
<td>-25</td>
<td>-70</td>
<td>-12</td>
<td>L</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
</tr>
<tr>
<td>3.35</td>
<td>-17</td>
<td>-74</td>
<td>-8</td>
<td></td>
<td>L</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
</tr>
<tr>
<td>2.83</td>
<td>-17</td>
<td>-66</td>
<td>-20</td>
<td></td>
<td>L</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
</tr>
<tr>
<td>0.82</td>
<td>3.35</td>
<td>51</td>
<td>-59</td>
<td>-12</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Fusiform Gyrus</td>
<td>Brodmann area 37</td>
</tr>
<tr>
<td>1.98</td>
<td>55</td>
<td>-63</td>
<td>-1</td>
<td></td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Inferior Temporal Gyrus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td>1.21</td>
<td>3.35</td>
<td>-28</td>
<td>-25</td>
<td>60</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 4</td>
</tr>
<tr>
<td>2.51</td>
<td>-25</td>
<td>-21</td>
<td>72</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 4</td>
</tr>
<tr>
<td>2.43</td>
<td>-32</td>
<td>-21</td>
<td>68</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 4</td>
</tr>
<tr>
<td>1.37</td>
<td>3.16</td>
<td>-6</td>
<td>-28</td>
<td>60</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Paracentral Lobule</td>
<td>Brodmann area 5</td>
</tr>
<tr>
<td>2.72</td>
<td>-2</td>
<td>-28</td>
<td>49</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Paracentral Lobule</td>
<td>Brodmann area 31</td>
</tr>
<tr>
<td>2.46</td>
<td>-2</td>
<td>-25</td>
<td>68</td>
<td></td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Paracentral Lobule</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td>Z-Score</td>
<td>MNI Coordinates</td>
<td>Lobe</td>
<td>Gyrus</td>
<td>Brodmann Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------------</td>
<td>------</td>
<td>-------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.93</td>
<td>2.99 -55 -51 -12</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Inferior Temporal Gyrus</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.49</td>
<td>-55 -55 -20</td>
<td>L</td>
<td>Posterior Lobe</td>
<td>Declive</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.04</td>
<td>2.97 10 40 -4</td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Anterior Cingulate</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.95</td>
<td>13 51 3</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Medial Frontal Gyrus</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure A.11: Similarity of dimensional and categorical affective spaces based on self-report. (A) Scatterplot of group-average distances between pairwise combinations of emotions in dimensional and categorical spaces ($r = .778$, $P < .0001$). The x-axis reflects distance in an affective space organized by dimensions of valence and arousal, and the y-axis reflects distance in a space organized along seven orthogonal emotion categories. The gray line indicates the best fitting least-squares line. (B) Parameter estimates for models predicting errors in classifying neural data using distances in dimensional and categorical models individually (left and center panels) and when combined (right panel). Increasing distance in dimensional affective space is associated with increasing errors when considered alone and when modeled with categorical distances. Increasing distance in categorical affective space was associated with fewer classification errors, but only when variance in the dimensional model is considered. Thus, inclusion of the dimensional term suppressed error in the model and allowed categorical and dimensional effects to be teased apart, increasing the overall model fit. Error bars reflect 95% confidence intervals.
Table A.12: Peak contrasts of parameter estimates for valence and arousal.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Extent (ml)</th>
<th>Z</th>
<th>MNIx</th>
<th>MNIy</th>
<th>MNIz</th>
<th>Hemisphere</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive &gt; Negative Valence</td>
<td>5.93</td>
<td>4.08</td>
<td>40</td>
<td>-44</td>
<td>45</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Inferior Parietal Lobule</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td></td>
<td>3.09</td>
<td>36</td>
<td>59</td>
<td>45</td>
<td>45</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 19</td>
</tr>
<tr>
<td></td>
<td>2.87</td>
<td>55</td>
<td>-28</td>
<td>45</td>
<td>45</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Inferior Parietal Lobule</td>
<td>Brodmann area 40</td>
</tr>
<tr>
<td></td>
<td>9.22</td>
<td>3.58</td>
<td>-51</td>
<td>-9</td>
<td>41</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 4</td>
</tr>
<tr>
<td></td>
<td>3.46</td>
<td>-44</td>
<td>-59</td>
<td>37</td>
<td>37</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 39</td>
</tr>
<tr>
<td></td>
<td>3.15</td>
<td>-32</td>
<td>-66</td>
<td>41</td>
<td>41</td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 39</td>
</tr>
<tr>
<td></td>
<td>4.83</td>
<td>3.51</td>
<td>-6</td>
<td>32</td>
<td>18</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Anterior Cingulate</td>
<td>Brodmann area 32</td>
</tr>
<tr>
<td></td>
<td>3.17</td>
<td>-6</td>
<td>13</td>
<td>34</td>
<td>34</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 24</td>
</tr>
<tr>
<td></td>
<td>3.05</td>
<td>6</td>
<td>40</td>
<td>22</td>
<td>22</td>
<td>R</td>
<td>Limbic Lobe</td>
<td>Cingulate Gyrus</td>
<td>Brodmann area 32</td>
</tr>
<tr>
<td>Negative &gt; Positive Valence</td>
<td>60.91</td>
<td>5.95</td>
<td>48</td>
<td>-70</td>
<td>7</td>
<td>R</td>
<td>Occipital Lobe</td>
<td>Middle Occipital Gyrus</td>
<td>Brodmann area 37</td>
</tr>
<tr>
<td></td>
<td>5.76</td>
<td>-47</td>
<td>-74</td>
<td>3</td>
<td>3</td>
<td>L</td>
<td>Occipital Lobe</td>
<td>Inferior Temporal Gyrus</td>
<td>Brodmann area 37</td>
</tr>
<tr>
<td></td>
<td>5.74</td>
<td>48</td>
<td>-44</td>
<td>18</td>
<td>18</td>
<td>R</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 13</td>
</tr>
<tr>
<td></td>
<td>7.08</td>
<td>5.33</td>
<td>36</td>
<td>-2</td>
<td>49</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>40</td>
<td>10</td>
<td>26</td>
<td>26</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Precentral Gyrus</td>
<td>Brodmann area 6</td>
</tr>
<tr>
<td></td>
<td>3.74</td>
<td>40</td>
<td>21</td>
<td>22</td>
<td>22</td>
<td>R</td>
<td>Frontal Lobe</td>
<td>Middle Frontal Gyrus</td>
<td>Brodmann area 9</td>
</tr>
<tr>
<td></td>
<td>7.41</td>
<td>4.25</td>
<td>-9</td>
<td>-51</td>
<td>45</td>
<td>L</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td></td>
<td>4.12</td>
<td>6</td>
<td>-55</td>
<td>49</td>
<td>49</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Precuneus</td>
<td>Brodmann area 7</td>
</tr>
<tr>
<td></td>
<td>2.73</td>
<td>2</td>
<td>-55</td>
<td>26</td>
<td>26</td>
<td>L</td>
<td>Limbic Lobe</td>
<td>Posterior Cingulate</td>
<td>Brodmann area 23</td>
</tr>
<tr>
<td>Positive &gt; Negative Arousal</td>
<td>41.81</td>
<td>Inf</td>
<td>-51</td>
<td>-17</td>
<td>3</td>
<td>L</td>
<td>Temporal Lobe</td>
<td>Superior Temporal Gyrus</td>
<td>Brodmann area 22</td>
</tr>
<tr>
<td>Z-score</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>Hemisphere</td>
<td>Lobe/Region</td>
<td>Brodmann</td>
<td>Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>------------</td>
<td>-------------------------------------------------</td>
<td>-----------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.30</td>
<td>-51</td>
<td>25</td>
<td>11</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Brodmann</td>
<td>area 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.20</td>
<td>-40</td>
<td>29</td>
<td>-8</td>
<td>L</td>
<td>Frontal Lobe</td>
<td>Brodmann</td>
<td>area 45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.46</td>
<td>4.52</td>
<td>-13</td>
<td>6</td>
<td>15</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Brodmann</td>
<td>area 47</td>
<td></td>
</tr>
<tr>
<td>4.43</td>
<td>-9</td>
<td>-28</td>
<td>-4</td>
<td>L</td>
<td>Sub-lobar</td>
<td>Caudate</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.43</td>
<td>2</td>
<td>-28</td>
<td>-4</td>
<td>L</td>
<td>Midbrain</td>
<td>Red</td>
<td>Nucleus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.19</td>
<td>3.79</td>
<td>36</td>
<td>-47</td>
<td>45</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Brodmann</td>
<td>area 7</td>
<td></td>
</tr>
<tr>
<td>3.47</td>
<td>29</td>
<td>-59</td>
<td>49</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Brodmann</td>
<td>area 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.32</td>
<td>36</td>
<td>-32</td>
<td>41</td>
<td>R</td>
<td>Parietal Lobe</td>
<td>Brodmann</td>
<td>area 40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Negative > Positive Arousal**
Appendix B

Table B.1: Friedman’s ANOVAs for changes in classification scores over time

<table>
<thead>
<tr>
<th>Classification Model</th>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$\chi^2$</th>
<th>Prob &gt; $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Time</td>
<td>2.07E+07</td>
<td>255</td>
<td>80980.6</td>
<td>3766.43</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.77E+08</td>
<td>126990</td>
<td>5331.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amusement</td>
<td>Time</td>
<td>2.31E+07</td>
<td>255</td>
<td>90494.6</td>
<td>4208.92</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.75E+08</td>
<td>126990</td>
<td>5312</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>Time</td>
<td>6.06E+06</td>
<td>255</td>
<td>23774.5</td>
<td>1105.76</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.92E+08</td>
<td>126990</td>
<td>5445.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>Time</td>
<td>2.01E+07</td>
<td>255</td>
<td>78655.96</td>
<td>3658.31</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.78E+08</td>
<td>126990</td>
<td>5335.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>Time</td>
<td>9.89E+06</td>
<td>255</td>
<td>38802.6</td>
<td>1804.72</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.88E+08</td>
<td>126990</td>
<td>5415.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sad</td>
<td>Time</td>
<td>4.68E+07</td>
<td>255</td>
<td>183491.7</td>
<td>8534.24</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.51E+08</td>
<td>126990</td>
<td>5125.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>Time</td>
<td>1.93E+07</td>
<td>255</td>
<td>75639.1</td>
<td>3517.99</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.78E+08</td>
<td>126990</td>
<td>5341.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6.98E+08</td>
<td>127743</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SS = sum of squares, df = degrees of freedom, MS = mean square.
Appendix C

Figure C.1: Experiential self-report of emotion mirrors that conveyed in facial and vocal expressions. (a) Response screen used for self-report. (b) Thresholded parametric maps of decision values from nonlinear SVMs (one sample t-test, \( n = 21 \)). Colored areas indicate the regions that predict each type of expression (\( P < .0001 \) uncorrected).
Figure C.2: The information content of response patterns within right postcentral gyrus increases with the separability of self-report. (a) Scatterplot depicts cross-validated estimates of accuracy for classification of self-report and neural data, with each point corresponding to a single subject (n = 21). Dashed lines indicate chance accuracy levels. Solid black line indicates the best least-squares fit to the data. (b) Histogram of bootstrap distribution of Pearson’s correlation coefficient, with dashed lines indicating 95% confidence interval computed using the bias corrected and accelerated percentile method.
References


Damasio, A., & Carvalho, G. B. (2013a). The nature of feelings: evolutionary and neurobiological origins. *Nat Rev Neurosci, 14*(2), 143-152. doi: 10.1038/nrn3403


Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci, 8*(9), 700-711. doi: 10.1038/nrn2201


doi: 10.1109/TIT.1968.1054102


189


195


Biography

Philip Kragel was born in Baltimore, Maryland in October of 1984. He received a Bachelor of Science in Engineering in Biomedical Engineering in 2006, and a Master of Engineering Management in 2007, both from Duke University. Subsequently, he worked as a computer analyst and programmer at Duke University in the laboratories of Dr. David Rubin and Dr. Kevin LaBar. In 2010, Philip began his graduate studies in the department of Psychology & Neuroscience at Duke University under the supervision of Dr. Kevin LaBar. Following completion of his doctoral research, Philip will work as a postdoctoral researcher in the laboratory of Dr. Tor Wager at the University of Colorado Boulder starting in the fall of 2015.

Peer Reviewed Publications


**Honors and Awards**


Claire Hamilton Conference Travel Award, 2011