Event-related potential and functional MRI studies of emotion and self-relevance

A dissertation submitted by

Eric C. Fields

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Psychology

Tufts University

May 2017

Adviser: Gina Kuperberg, M.D., Ph.D.

Abstract

This dissertation is comprised of three related, but distinct projects. Chapters 1 and 2 explore the processing of emotional stimuli and how emotional properties of stimuli interact with context, goals, schemas, and basic cognitive processes. Chapter 1 presents a functional MRI study showing that self-related processing in the medial prefrontal cortex (a brain region long associated with the self) is modulated by the emotional valence of social scenarios. Specifically, self-relevance only increased activation in this region for positive scenarios, suggesting the mPFC may be sensitive to self-positivity biases and may even play a role in maintaining these biases. Chapter 2 reviews the late positive potential of the ERP commonly observed to emotional stimuli, proposes a functional theory of this component, and presents an initial study testing this theory. Chapter 3 is a methodological chapter exploring issues in the statistical analysis of ERP data, particularly the problem of multiple comparisons presented by the large amount of data across time and space. I present new software to implement permutation-based mass univariate statistics for factorial designs, and simulation data on the relative power and flexibility of these statistical tests.

Acknowledgements

A doctoral dissertation in cognitive neuroscience lists a single author but is in fact the work of many people.

I first want to thank my direct collaborators for the work presented here. My advisor, Gina Kuperberg, was a close collaborator and mentor for all of this work and the resources and funding provided by Dr. Kuperberg supported this work. The fMRI study presented in Chapter 1 was done in collaboration with Nate Delaney-Busch, Kirsten Weber, and Ben Stillerman. Candida Ustine and Ellen Lau also helped with getting the study off the ground and technical issues. The ERP study described in Chapter 2 was done in collaboration with Nate Delaney-Busch. Hannah Levin also helped with many aspects of this study as an undergraduate research assistant, and Ayanna Thomas and Elizabeth Race gave valuable advice on the memory aspects of that study.

More generally I want to thank everyone in the NeuroCognition Lab who has helped me over the years. Nate Delaney-Busch has been my co-thinker on nearly all the work I've done and I have benefited from innumerable conversations with him about all the topics and ideas covered in this dissertation. Phil Holcomb was an invaluable source of knowledge when I was first learning the complexities of event-related potentials. I first learned the fMRI technique working with Ellen Lau and Kirsten Weber. And I have benefited from the work of many undergraduate and full-time research assistants over the years: Abby Swain, Arim Choi, Kristina Fanucci, Margarita Zeitlin, Candidate Ustine, Scott Burns, Connie Choi, Ben Stillerman, Minjae Kim, and Lena Warnke.

I thank my undergraduate research mentors, William Langston and John Pennington, without whom I would have never ended up in a doctoral program in the first place. And I thank family and friends—particularly Allison Fogel and my parents, Merle and Doris Fields—for their support over the years.

Finally, I thank my dissertation committee, Gina Kuperberg, Phil Holcomb, Elizabeth Race, and Jon Freeman, for their advice, support, and service.

Table of Contents

List of Tables

List of Figures

Event-related potential and functional MRI studies of emotion and self-relevance

Eric C. Fields

General Introduction

This dissertation is comprised of three related, but distinct, projects. Chapters 1 and 2 explore the processing of emotional stimuli and how emotional properties of stimuli interact with context, goals, schemas, and basic cognitive processes. Chapter 3 is a methodological chapter exploring issues in the statistical analysis of ERP data, presenting new software to implement permutation-based mass univariate statistics, and presenting novel simulation data on the relative power and flexibility of these statistical tests.

In Chapter 1, I present a functional MRI study examining the processing of self-relevant and non-self-relevant scenarios with different emotional content. This work builds on two key literatures. First, a large literature examining the neural correlates of self has identified the medial prefrontal cortex as the brain region most consistently associated with self-related processing. Second, a large behavioral literature suggests that most people have an (unrealistically) positively biased self-schema, and we have recently shown effects of this positivity bias on the N400 component of the ERP (Fields & Kuperberg, 2015). Following form this, we were interested in whether self-related activity in the mPFC would be modulated by emotional valence in the social vignettes that had previously shown self-relevance x emotion interactions in ERPs (Fields & Kuperberg, 2012, 2015, 2016). Indeed, we found that the mPFC showed greater activation for self-relevant scenarios only for positive vignettes. In Chapter 1, I discuss the implications of these findings for the function of the mPFC and future work on the neural correlates of self.

Chapter 2 explores the processing of emotional stimuli, and the interaction of emotional properties of stimuli with many other factors (including self-relevance), in ERPs. In particular, I focus on the component of the ERP most consistently modulated by the emotion: the late positive component (LPC). I present a systematic review of the literature on this component to emotional words and language. From this, I propose that the LPC may be related to or synonymous with the

P300 component of the ERP and that it may reflect a process whereby the current model of the environment is updated in response to informative and important stimuli. I then present an ERP study intended as an initial investigation of this theory and discuss the future of research on the function of the LPC.

Chapter 3 provides an in depth critique of traditional methods for analyzing ERP data. In particular, the ERP literature too often does not sufficiently address the multiple comparisons problem presented by a technique that generates large amounts of data across time and space (electrodes), thereby inflating the Type I error rate in ERP studies. This problem is often combined with low power, leading to studies and statistical tests that provide little evidence despite reporting statistically significant effects. This is likely one factor contributing to the complicated and sometimes contradictory literature reviewed in Chapter 2. I discuss permutation-based mass univariate statistics as a solution to this problem and address two key barriers to the widespread adoption of these techniques. First, current implementations of these techniques do not allow for factorial designs and the testing of interaction effects. I introduce new software to implement factorial versions of these statistics. Second, mass univariate techniques are often seen as sacrificing power to maintain flexibility and Type I error rate. I show via simulation studies that, in fact, these techniques can have greater power than traditional analysis approaches when constrained via the same spatial and temporal assumptions, while also maintaining reasonable power even when these assumptions are significantly relaxed.

Chapter 1: An fMRI investigation of valence effects on self-related activation in mPFC

Introduction

The relationship between emotion and the self-concept lies at the core of human wellbeing. Understanding this complex relationship is critical for understanding motivation, learning, decision making, and mood disorders such as depression and anxiety (Beck, Rush, Shaw, & Emery, 1979; Dunning, Heath, & Suls, 2004; Goldin et al., 2013; Sharot & Garrett, 2016; Shestyuk & Deldin, 2010; Taylor & Brown, 1988). Consequently, it is important that we understand the cognitive and neural mechanisms by which the self-concept and self-esteem are constructed and maintained (Beer, 2014; Beer & Flagan, 2015; Chavez & Heatherton, 2015; Sharot & Garrett, 2016). Here we report a functional MRI study examining the effects of differently valenced self-relevant and non-self-relevant information on processing in the medial prefrontal cortex (mPFC), a region that has regularly been associated with the self. We begin by reviewing the social cognitive neuroscience literature on the self, and then discuss interactions with the self-concept and associated motivated processes.

Medial prefrontal cortex and the social cognitive neuroscience of self

The self has been long been a topic of investigation in philosophy and psychology (Swann & Bosson, 2010). As functional MRI and other neuroimaging methods have become widely available to study the intact, healthy brain, researchers have investigated the neural basis of the self. Social and cognitive neuroscientists have asked whether the self is processed uniquely in the brain and whether there might be regions of the brain specialized for processing self-related information. In the neuroimaging literature, a network of brain regions has been associated with self-related processing including the temporal poles, temporal-parietal junction, and much of the cortical midline, including medial prefrontal cortex (mPFC), precuneus, and the posterior cingulate

cortex (Legrand & Ruby, 2009; Northoff et al., 2006; Qin, Duncan, & Northoff, 2013). Within this network, the region most consistently associated with self-related processing is the mPFC. However, there has been much debate about whether this region (or any subsection of it) is specialized for self-related processing and what its function might be.

Empirically, it is clear that the mPFC is regularly activated by self versus other contrasts (Araujo, Kaplan, & Damasio, 2013; Denny, Kober, Wager, & Ochsner, 2012; Northoff et al., 2006; D. D. Wagner, Haxby, & Heatherton, 2012). But it is also activated when thinking about other people (Amodio & Frith, 2006; Saxe, 2009; Van Overwalle, 2009; D. D. Wagner, Kelley, Haxby, & Heatherton, 2016; Welborn & Lieberman, 2015), as well as in non-social domains (Legrand & Ruby, 2009; Spreng, Mar, & Kim, 2009). Interpreting this pattern of results has proven difficult for a number of reasons. First, the mPFC is a large region and it is possible that only a sub-region or subset of neurons in the mPFC is specialized for the self (Denny et al., 2012). Second, it is possible that mPFC activations that don't initially appear self-related are nevertheless due to processing of information in relation to the self. For example, it may be that mPFC is activated by evaluations and mental state attributions of others because we understand others via simulation or comparison to ourselves (e.g., Mitchell, 2009; Uddin, Iacoboni, Lange, & Keenan, 2007). Finally, the opposite could also be true: activations that appear to be self-related may actually be due to confounds across experimental conditions. There are a number of challenges in studying the self, and it is especially difficult to develop an appropriate control condition since self and non-self conditions almost always differ on properties other than self-relevance such as familiarity, emotional associations, depth of knowledge, etc. (Gillihan & Farah, 2005; Legrand & Ruby, 2009). It may be that it is these factors, rather than the self per se, that is driving mPFC activations. Thus, it is difficult to establish that any given activation either is *or* isn't related to the self specifically.

Given these challenges, it is not surprising that theories of the function of the mPFC are diverse. One perspective is that mPFC activation reflects a domain general process that is simply

often increased during self-related processing. For example, the mPFC has been argued to be part of a network supporting domain general inferential processing (Legrand & Ruby, 2009) or allowing people to consider situations other than the immediate present reality (i.e., "self-projection"; Buckner & Carroll, 2007). Others have argued that the mPFC is specialized for processing of social information and/or mental states and is engaged whether the target of processing is the self or others (Amodio & Frith, 2006; Saxe, 2009; Zaki & Ochsner, 2011). Finally, some researchers have argued that some part of the mPFC is specialized for self-related processing. The dorsal mPFC (dmPFC) has been proposed to compute inferences about or evaluation of the self (Northoff & Bermpohl, 2004; Uddin et al., 2007), while the ventral mPFC (vmPFC) is argued to compute or represent the value or meaning of a stimulus to the self (Abraham, 2013; D'Argembeau, 2013; Flagan & Beer, 2013).

Whatever its precise function, the consistent activation of the mPFC by self-related stimuli and conditions shows that it plays an important role in processing information about the self. It is therefore worth asking how factors known to be important in self-related processing impact mPFC activation. Indeed, investigating such factors may lead to a better understanding of the function of the mPFC.

Positive self-evaluations, motivated information processing, and the social psychology of self

Interestingly, there is some discrepancy between the study of the self in the social psychological and cognitive neuroscience literatures (Chavez & Heatherton, 2015; Flagan & Beer, 2013). While the neuroscience literature has often focused on the search for self-specific neural regions or sought to understand how the self is represented in the brain, the social and cognitive literature on the self has emphasized the way in which the self as a schema organizes information. One particular focus has been the role of motivated reasoning and positive biases with regard to the self.

It is well established that people tend to view themselves in an unrealistically positive light when compared to others or objective standards. We see ourselves as having more positive (and less negative) traits and abilities than others and we expect more positive outcomes for ourselves across many domains (Alicke & Govorun, 2005; Armor & Taylor, 2002; Dunning et al., 2004; Taylor & Brown, 1988). We are able to maintain these positive self-evaluations via motivated reasoning and asymmetric treatment of positive and negative self-related information. In response to negative information about the self, we employ a variety of strategies such as re-interpreting outcomes, shifting standards of comparison, and attributing negative outcomes to external, situation-specific factors (Armor & Taylor, 2002; Mezulis, Abramson, Hyde, & Hankin, 2004). The result is that beliefs are more likely to be updated in response to positive information than negative information about the self (Sharot & Garrett, 2016). In fact, work with Bayesian learning models has shown that models best explain behavior when separate learning rates are assumed for positive and negative information (reviewed in Sharot & Garrett, 2016).

This "self-positivity bias" has important real-world consequences. Positive self-views are often seen as key for self-esteem and motivation (Sharot & Garrett, 2016; Taylor & Brown, 1988) and lack of a self-positivity bias is associated with mood disorders (Beck et al., 1979; Garrett et al., 2014; Goldin et al., 2013; Shestyuk & Deldin, 2010). In addition, modelling work shows that unrealistic positivity is adaptive in many circumstances (D. D. P. Johnson & Fowler, 2011). On the other hand, there are negative consequences to positive illusions such as failure to adjust behavior in response to knowledge of disease risk factors and inadequate studying by students who have an unrealistic perception of their comprehension (Dunning et al., 2004; D. D. P. Johnson & Fowler, 2011). It is therefore important to understand the neurocognitive mechanisms that underlie unrealistic self-positivity (Chavez & Heatherton, 2015).

In previous work, we have used ERPs to examine effects of the self-positivity bias during processing of social vignettes (Fields & Kuperberg, 2015). Specifically, we examined whether the

N400 component of the ERP could be used as a measure of positively biased self-views. The N400 is a component that is reduced to the extent that the semantic features of a word (or other meaningful stimuli) match expectations generated by the preceding context (Kutas & Federmeier, 2011). We therefore hypothesized that positive words would elicit a smaller N400 in self-relevant (versus other-relevant) contexts and that this may provide a more accurate and less biased measure of self-views than many self-report and behavioral measures (see Fields & Kuperberg, 2015). We presented two-sentence scenarios with a neutral, positive, or negative critical word such as "A man knocks on Sandra's hotel room door. She sees that he has a tray/gift/gun in his hand". These scenarios were made self-relevant by simply changing the subject: "A man knocks on *your* hotel room door. *You* see that he has a tray/gift/gun in his hand" (see Brunyé, Ditman, Mahoney, Augustyn, & Taylor, 2009). As predicted, the N400 was reduced to positive words in self-relevant contexts (with no effect of self-relevance in neutral or negative scenarios).

The present study

Given the social psychological literature on positive self-evaluations and our ERP results, we were interested in how valence in the paradigm just described would modulate the effects of self-relevance in the mPFC. That question was the focus of the present work.

Although previous work has examined the interaction of self-relevance and emotion in fMRI, most of these studies have established self-relevance via the task or participant's ratings of stimuli (e.g., Chavez & Heatherton, 2015; Chavez, Heatherton, & Wagner, in press; Fossati et al., 2003; Fossati et al., 2004; Moran, Macrae, Heatherton, Wyland, & Kelley, 2006; Ochsner et al., 2004; Phan et al., 2004; see also, Lee & Siegle, 2012). This means that self-relevance was established differently than emotion and that in most cases participants themselves determined which stimuli were self-related, not the experimenters.

Herbert, Herbert, and Pauli (2011) report an initial study that used bottom-up stimulus based manipulations for both emotion and self-relevance. They presented neutral, positive, and

negative words preceded by “my”, “his”, or “the”. In this study, self-relevance (i.e., “my”) increased effects of emotion in the vmPFC, an area often associated with emotion processing. This was true for both positive and negative words versus neutral words. However, when this paradigm was used in an ERP study (Herbert, Herbert, Ethofer, & Pauli, 2011), results did not show the interaction we previously observed on the N400. This may be due to the limited context provided by short noun phrases and the relatively weaker self-relevance of text in first-person compared to second-person (see Brunyé et al., 2009). We therefore wanted to extend these results to investigate the effects of the self-positivity bias in the mPFC using longer, more naturalistic scenarios.

We presented two-sentence social vignettes that were either self-relevant or other relevant and that had a neutral, positive, or negative critical word in the second sentence (see example above). Our previous ERP results indicated that the processing of such scenarios interacts with the positively biased self-concept in a fast and relatively automatic manner. Here we were interested in whether such positive biases and motivated processing would affect self versus other activations in mPFC.

To the extent that mPFC activation reflects (or is a result of) the evaluation of stimuli as self-relevant or reflects activation of self-representations, it may be that activation will be greater for scenarios that participants find more plausible and congruent with their self-concept. Our ERP results and the literature reviewed above suggest this would be the positive self-relevant scenarios. This would predict a pattern of activation similar to that observed on the N400 in our ERP study, with greater self versus other effects for positive scenarios specifically. On the other hand, it is possible that scenarios *incongruent* with participants’ expectations will elicit greater activity in self-related regions as they attempt the more difficult process of integrating negative, self-relevant information with their (generally positive) self-concept. Either of these results is consistent with the idea that the self-positivity bias interacts with self-related processing in the

mPFC. In contrast, it is possible that mPFC activations will not interact with the emotion condition, which would suggest that the processing represented by mPFC activations is relatively independent of the biases and motivated processing studied by social psychologists.

Methods

Participants

Seventeen female subjects were recruited through an advertisement on a Tufts University community website (tuftslife.com). Self-reported race and ethnicity was non-Hispanic White for 12 participants, Hispanic for one participant, Asian for one participant, mixed Asian/White for two participants, and unreported for one participant. All participants were right-handed native English speakers (having learned no other language before the age of 5) between the ages of 18 and 23 (M = 20.7, SD = 1.3), who reported no history of psychiatric or neurological disorders. Participants were paid for their participation and provided informed consent in accordance with the procedures of the Institutional Review Board of Massachusetts General Hospital.

Stimuli

Stimuli were a modified version of those used in our previous ERP work (Fields & Kuperberg, 2012, 2015, 2016).[1] 216 sets of two-sentence scenarios were developed, each with three Emotion conditions (positive, neutral, and negative) and two Self-Relevance conditions (self and other). These were crossed in a 3 x 2 factorial design so that there were six versions for each scenario: self-positive, self-neutral, self-negative, other-positive, other-neutral, and other-negative (see Table 1.1 for examples).

[1] The stimuli used in Fields and Kuperberg (2012, 2015, 2016) were modified so that length of the second sentence and CW position were more consistent (see below). We also eliminated 6 vignettes so that the scenarios would divide evenly across six runs. The stimuli were otherwise as similar as possible to our previous studies. For easy comparison, the same scenarios are reported in Table 1.1 as we have reported in our previous papers.

Other			Self		
Positive	**Neutral**	**Negative**	**Positive**	**Neutral**	**Negative**
A man knocks on Sandra's hotel room door. She sees that he has a gift in his hand.	A man knocks on Sandra's hotel room door. She sees that he has a tray in his hand.	A man knocks on Sandra's hotel room door. She sees that he has a gun in his hand.	A man knocks on your hotel room door. You see that he has a gift in his hand.	A man knocks on your hotel room door. You see that he has a tray in his hand.	A man knocks on your hotel room door. You see that he has a gun in his hand.
Fletcher writes a poem for a class. His friends think it's a very beautiful composition.	Fletcher writes a poem for a class. His friends think it's a very intricate composition.	Fletcher writes a poem for a class. His friends think it's a very boring composition.	You write a poem for a class. Your friends think it's a very beautiful composition.	You write a poem for a class. Your friends think it's a very intricate composition.	You write a poem for a class. Your friends think it's a very boring composition.
Vince spends time with relatives over the break. This turns out to be a wonderful experience for him.	Vince spends time with relatives over the break. This turns out to be a characteristic experience for him.	Vince spends time with relatives over the break. This turns out to be a disastrous experience for him.	You spend time with relatives over the break. This turns out to be a wonderful experience for you.	You spend time with relatives over the break. This turns out to be a characteristic experience for you.	You spend time with relatives over the break. This turns out to be a disastrous experience for you.
After dinner, Lydia is involved in a discussion. She makes a few remarks that impress her friends.	After dinner, Lydia is involved in a discussion. She makes a few remarks that surprise her friends.	After dinner, Lydia is involved in a discussion. She makes a few remarks that hurt her friends.	After dinner, you are involved in a discussion. You make a few remarks that impress your friends.	After dinner, you are involved in a discussion. You make a few remarks that surprise your friends.	After dinner, you are involved in a discussion. You make a few remarks that hurt your friends.
Carmelo has been in his current job for over a year. He learns he is getting a bonus this December.	Carmelo has been in his current job for over a year. He learns he is getting a transfer this December.	Carmelo has been in his current job for over a year. He learns he is getting a pay-cut this December.	You have been in your current job for over a year. You learn you are getting a bonus this December.	You have been in your current job for over a year. You learn you are getting a transfer this December.	You have been in your current job for over a year. You learn you are getting a pay-cut this December.

Table 1.1: *Examples of two-sentence scenarios in each of the six conditions.* The critical word is underlined (but did not appear underlined in the actual stimulus lists).

All scenarios were written in the present tense. The first sentence (4-13 words long) always introduced a situation involving one or more people, only one of which was specifically named (the protagonist), and it was always neutral or ambiguous in valence: e.g., *A man knocks on Sandra's hotel room door*. The named protagonist was male half the time and female the other half. To create the self conditions, the named person was changed to "you", e.g., *A man knocks*

on your hotel room door. In some scenarios this necessitated changing the conjugation of the verb, but the first sentence was otherwise identical across the self and other conditions. The second sentence (8-10 words) continued the scenario and was the same across all emotion conditions except for one word, the critical word, which was positive, neutral, or negative, e.g.: *She/You see(s) that he has a gift/tray/gun in his hand*. To obtain a roughly equivalent time course of neural response, the critical word was always the either the sixth word (48 scenarios) or seventh word (168 scenarios) of the sentence. The part of speech of the critical word was the same across the three Emotion conditions for each scenario: adjective for 130 scenarios, verb for 50 scenarios, and noun for 36 scenarios. The second sentence was otherwise emotionally neutral or ambiguous. All scenarios were written so that they would be plausible for most college students, both male and female.

	Other			Self		
	Neutral	**Positive**	**Negative**	**Neutral**	**Positive**	**Negative**
(log) HAL Frequency*	8.19(2.53)	8.28(2.44)	8.05(2.22)	--	--	--
CW length (letters)	7.47(2.22)	7.64(2.39)	7.16(2.49)	--	--	--
Scen. Length (char w/spaces)	94.1(11.4)	94.2(11.3)	93.7(11.5)	92(11.6)	92.1(11.4)	91.5(11.6)
Valence (CW)	4.31(0.56)	5.68(0.55)	2.32(0.56)	--	--	--
Arousal (CW)	3.39(0.64)	4.48(0.82)	3.83(0.63)	--	--	--

Table 1.2: *Stimuli ratings and characteristics*. Means are shown with standard deviations in parentheses. Valence and arousal were rated on seven point scales from very unpleasant and least arousing to very pleasant and most arousing respectively. "--" indicates that, for ratings conducted on the words in isolation from the scenario contexts, the values were the same in the self conditions as in the other conditions since the identical critical words were used (except for in six scenarios in which the verb was conjugated differently). *Some words did not exist in the HAL database and these were represented as null values in our calculations

Procedure

Stimulus presentation and task

Scenarios were divided into six lists, which were counterbalanced such that the same critical word did not appear more than once in any list and a given scenario's context did not appear more than once per list. Each of the six lists included 216 sentence pairs (36 in each condition), which were broken into six blocks of 36 sentence pairs. Participants were randomly

assigned to one of the lists. Stimuli were presented on a projector in white font centered on a black background. Each trial began with a fixation cross of variable duration (most commonly 2 seconds but ranging up to 20 seconds) to introduce jitter. Fixation timings were determined using Optseq (https://surfer.nmr.mgh.harvard.edu/optseq) to optimize the ability to deconvolve the BOLD response. Each of the two sentences of the scenario were then presented in full for 4 seconds.

Six comprehension questions were randomly interspersed in each block and appeared for 4 seconds directly after the second sentence of the scenario. For example, the scenario "Casper is/You are new on campus. His/Your classmates think he is/you are quite idiosyncratic/clever/dumb compared to others." was followed by the question "Did Casper/you go to this school last year?" with the correct answer being "no". Participants were instructed to a press a button corresponding to the index finger and middle finger for yes and no respectively before the question left the screen.

MRI Acquisition

Structural and functional magnetic resonance images were acquired using a 3T Siemens Trio scanner using a 32-channel head coil. FMRI data were acquired over six runs, each lasting for approximately 7 minutes and 38 seconds. In each run, 230 functional volumes (36 axial slices (AC-PC aligned), 3.2 mm slice thickness, .64 mm skip, 200 mm field of view, in-plane resolution of 3.125 mm) were acquired with a gradient-echo sequence (TR = 2s, TE = 25ms, flip angle = 77º, ascending acquisition order). In addition, at the beginning and end of the scanning session, we acquired two T1-weighted high-resolution structural images (1 mm isotropic multi-echo MPRAGE: TR = 2.53s, flip angle = 7º, four echoes with TE = 1.64ms, 3.5ms, 5.36ms, 7.22ms). We used the higher quality of the two structural scans from each subject (based on visual inspection) for the subsequent analysis.

MRI processing and analysis

Pre-processing as well as the first and second level analyses of the fMRI data were conducted in SPM8 (www.fil.ion.ucl.ac.uk/spm), supplemented by additional add-on toolboxes (ArtRepair: cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html; Mazaika, Hoeft, Glover, & Reiss, 2009).

The first four images in each run were discarded to ensure that transient non-saturation effects did not affect the analysis. The next step was to detect spikes and interpolate these bad slices from surrounding images (using the ArtRepair toolbox). On average 0.3% of slices (range 0 to 4.0%) were removed and interpolated in each run. Then, images were slice-time corrected and the volumes were realigned to the first images of each run and then to each other. The functional images were co-registered to the structural image by co-registering the mean functional image to the structural MPRAGE. The anatomical images were segmented into grey and white matter, and the spatial normalization parameters acquired during this step were used to normalize the functional images to the ICBM template for European brains. Finally, the images were smoothed with an 8 mm FWHM Gaussian kernel.

We modeled the data using a general linear model over 6 runs (following the experimental design described above). Each run had the following regressors: one for fixation, one for the first sentence of each scenario, six for the second sentence of each scenario (one for each condition created by the 3x2 design: Self-Positive, Self-Neutral, Self-Negative, Other-Positive, Other-Neutral, and Other-Negative), and one for the comprehension questions. The trials were modeled from the start of the second sentence and the duration was the presentation time of the second sentence (4 s). All regressors were convolved with a canonical hemodynamic response function. The realignment parameters for movement correction were also included in the model.

We defined the following contrasts to take to the second level for a random effects group analysis: the Self vs. Other contrast at each level of Emotion and each of the six conditions

created by the 3x2 design vs. Fixation. To test the interaction of Emotion and Self-Relevance, we used a within subjects ANOVA design matrix that consisted of one regressor for each individual subject and one regressor for the Self vs. Other contrast at each level of Emotion. To follow up significant interactions, we used separate one-sample t-test design matrices that consisted of a single regressor for the Self vs. Other contrast for each of the Emotion condition. In addition, we used six one-sample t-test design matrices to test each of the six conditions against baseline (fixation).

Based on a priori hypotheses concerning the mPFC, we defined a region of interest (ROI) for this entire region. To create the ROI, we used Denny et al.'s (2012) anatomical definition of the mPFC in MNI space ($|x|<25$, $y>15$, $z>-5$). To exclude coordinates outside the brain, we intersected this with the MNI template used in the normalization preprocessing step. We report whole-brain effects at a voxel-level threshold of $p < 0.001$ and a small volume correction FWE-corrected at the peak of a priori regions of interest described above. All reported coordinates are in MNI space.

Results

Behavioral data

Accuracy for the comprehension questions ranged from 72% to 100% with an average of 88%. Participants failed to provide a response on only 3.1% of comprehension questions (an average of 1.1 of the 36 questions). If these trials are excluded, accuracy ranged from 81% to 100% with an average of 91%. Thus, the behavioral data indicated that participants were reading and comprehending the scenarios.

Whole brain analyses

When compared to baseline all six conditions activated visual areas in the occipital and temporal lobe as well as a large network of regions generally associated with sentence processing

including lateral temporal lobe and ventrolateral prefrontal cortex. These activations are reported in the Appendix.

The 3 x 2 ANOVA comparing the six conditions directly to each other did not reveal significant effects of Emotion, Self-Relevance, or an Emotion x Self-Relevance interaction in any brain region.

Medial prefrontal small volume analysis

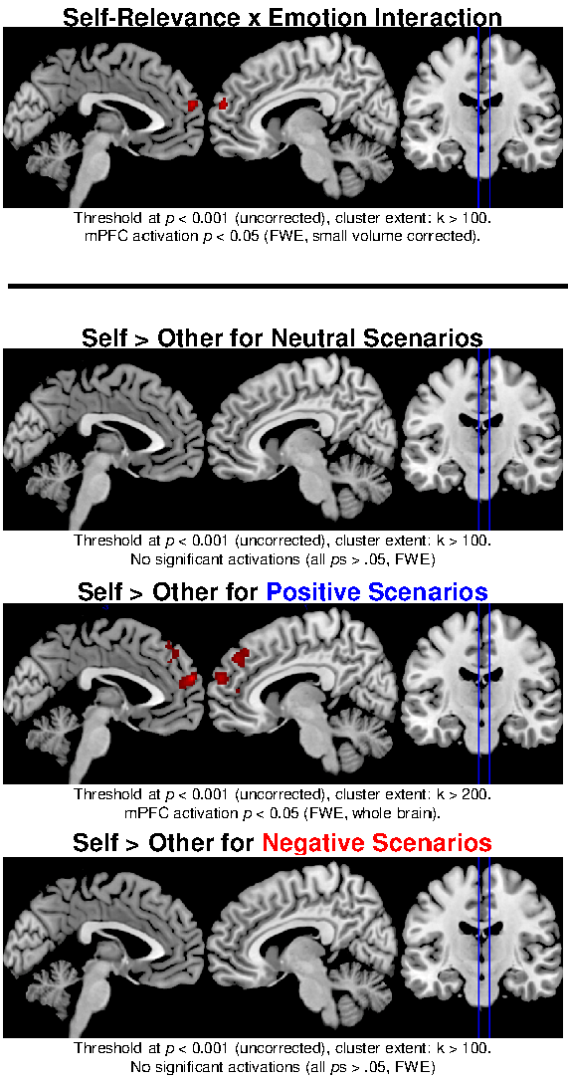

Figure 1.1 *Emotion x Self-Relevance interaction and follow-ups*. The mPFC small volume analysis revealed a significant Self-Relevance x Emotion interaction centered at MNI 0, 60, 22. We followed-up this interaction by looking at self versus other activation within each emotion condition. Only the positive scenarios showed significant activations.

For the ROI analysis within the anatomically defined mPFC, there were no significant main effects of Emotion or Self-Relevance. However, there was a significant Emotion x Self-Relevance interaction in a cluster centered at MNI coordinates 0, 60, 22 (cluster-level *p*(FWE) = .014, k = 116, peak voxel level *p* = .047, z-score = 4.26). Follow-ups revealed that there was a significant effect of Self-Relevance in this area as well as more dorsal parts of the mPFC for positive scenarios (see Table 1.3, Figure 1.1), but not the negative or neutral.

R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level
L	0.003	5.10	-2, 60, 22	*p*(FWE) < .001, k = 506
L	0.009	4.80	-6, 62, 24	
R	0.066	4.24	8, 38, 46	*p*(FWE) = .001, k = 235
R	0.090	4.14	6, 46, 40	

Table 1.3: *Self-positive vs. other-positive in the mPFC ROI*.

Discussion

Following from our previous work with ERPs (Fields & Kuperberg, 2015), we had participants undergo fMRI while reading two-sentence social vignettes that were either contextually self-relevant or other-relevant and that contained a neutral, positive, or negative critical word in the second sentence. We observed an interaction of self-relevance and valence in the medial prefrontal cortex driven by increased activation for self-relevant versus other-relevant contexts specifically for the positive scenarios.

Notably, this is the same general pattern of interaction that we observed on the N400 component of the ERP (i.e., self-other effects for positive, but not neutral or negative scenarios; Fields & Kuperberg, 2015). Given what we know about the N400 (Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011; Lau, Phillips, & Poeppel, 2008), the mPFC activation we observed is probably not the generator of the N400 effect (but see “Open Questions” later in this discussion). However, as we discuss below, these effects may be generated by the same underlying psychological mechanisms.

The social cognitive neuroscience of self-positivity

As noted in the Introduction, motivation and positive biases have been at the core of the social psychological understanding of the self, but these factors have received less attention in the social neuroscience literature, which has focused more on investigating the core representation and processing of self (Chavez & Heatherton, 2015; Flagan & Beer, 2013). However, this is beginning to change. In a series of studies, Jennifer Beer and colleagues have examined the role of multiple frontal regions and their connections with subcortical regions in self-insight and unrealistic self-positivity. Interestingly, the portions of mPFC most often associated with the self have not emerged as a significant part of the network underlying self-positivity in the explicit social comparison and judgement tasks they used. Instead they have identified a key role for the medial orbitofrontal cortex (mOFC) operating in two different networks of regions depending on whether self-esteem is under threat (reviewed in Beer, 2014), as well as the ventral anterior cingulate (vACC) as a region that is important for identifying opportunities for self-enhancement (reviewed in Beer & Flagan, 2015).

In another line of work, Chavez et al. (in press) have used multi-voxel pattern (MVPA) analysis to suggest that the self and positive valence may be represented by overlapping populations of neurons in the ventral medial prefrontal cortex. Specifically, they showed representational similarity at the voxel level for thinking about the self and viewing positive pictures and found that a MVPA pattern classifier trained to recognize positive valence can distinguish self-relevant from non-self-relevant conditions at above chance performance. They interpret these results as showing “that positive affect is a key component to self-referential thought”. However, like the findings reviewed in the previous paragraph, these effects were in an area of mPFC significantly ventral to the effects we observed.

Our work adds to these findings by suggesting that the “classic” self-other mPFC effect in more dorsal regions of mPFC (cf. Denny et al., 2012) can also be modulated by positive biases

and associated motivated processes. As discussed in the Introduction, there is disagreement about the function of the mPFC and the extent to which it is specialized for self-related (or social) processing (Denny et al., 2012; Legrand & Ruby, 2009; Northoff & Bermpohl, 2004; Saxe, 2009; Uddin et al., 2007; Zaki & Ochsner, 2011). However, its consistent activation by self-related experimental conditions (Denny et al., 2012; Legrand & Ruby, 2009; Northoff et al., 2006; Qin et al., 2013) suggests that it plays an important role in processing information about the self. Our results therefore suggest that a core aspect of self-related processing is engaged to a greater degree when information matches positive self-views.

In fact, this greater activation for positive self-information in more dorsal regions of mFPC may be part of the process by which unrealistic self-positivity is maintained. As noted in the Introduction, we are more likely to update our beliefs in response to positive than negative information about the self (Sharot & Garrett, 2016). Interestingly, some previous fMRI studies examining how unrealistic optimism is maintained have linked mPFC activity specifically to belief updating in response to positive self-related information. Sharot, Korn, and Dolan (2011) had participants estimate their likelihood of experiencing various adverse events and then presented the actual average probability of that event. After this task, they re-assessed participants' estimates of the likelihood of each event. They replicated findings that participants were unrealistically optimistic and that they updated their beliefs less in response to unexpectedly negative information than unexpectedly positive. In addition, they found that the same region of mPFC that showed the interaction observed in the present study was related to tracking prediction errors and belief updating specifically for unexpectedly *positive* (but not negative) feedback. Garrett et al. (2014) replicated these results and extended them to depression patients (see also Sharot & Garrett, 2016 for general discussion).

Further support for the importance of the mPFC to self-positivity comes from work examining the role of mPFC in self-esteem. Chavez and Heatherton (2015) have shown that

structural connectivity between the mPFC and the ventral striatum is associated with trait self-esteem while functional connectivity between these regions is associated with state self-esteem. In follow-up work, they have shown that structural connective between mPFC and ventral striatum predicts self-esteem 8 months later (Chavez & Heatherton, 2017). Thus, the self versus other activation seen specifically to positive scenarios in our study may not just reflect the fact that these scenarios were perceived as more self-relevant, it may also be related to the cognitive processes whereby the self-positivity bias is constructed and maintained.[2]

More generally, our findings suggest that simply comparing self-relevant versus non-self-relevant conditions may not reveal how the self is processed or represented in the brain per se. Instead, neural activity in response to self-relevant information is likely to reflect how the specific nature of that information interacts with aspects of the self-schema. Valence may be particularly important in this regard, and our results suggest that the positive biases and forms of motivated reasoning that social psychologists have studied for decades can also influence self-related mPFC activity. Future research should continue to take these processes into account and explore their implications for a social neuroscientific understanding of the self (see also Beer, 2014; Beer & Flagan, 2015; Chavez & Heatherton, 2015; Chavez et al., in press).

Open questions

Above we noted that while the pattern of interaction in mPFC here and on the N400 in our previous work (Fields & Kuperberg, 2015) was the same, the mPFC is not a likely generator for the N400 (see Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011; Lau et al., 2008). However, an alternative interpretation is that the two effects do reflect the same underlying activation, and that the previously observed ERP effect was not actually a modulation of the N400. Under this

[2] Of course, in the present paradigm subjects knew that the scenarios weren't actually about them, and thus would not necessarily have a reason to update their beliefs in response to them. However, it is likely that aspects of the way in which positive and negative self-related information is processed asymmetrically is relatively automatic rather than fully conscious and strategic (Beer, 2014).

interpretation, what appeared to be a reduced negativity, was in fact an increased positivity overlapping closely with the timing of the N400. Testing this possibility would likely require fMRI and joint EEG/MEG recordings in the same set of subjects (e.g., Lau, Gramfort, Hamalainen, & Kuperberg, 2013). If confirmed, it would provide interesting information about the timing of self-related mPFC activity and suggest that ERPs could be used in the future to study such activity.

In the Introduction, we suggested that the opposite pattern of results may also make sense: that is, increased activity in response to scenarios *incongruent* with the positive self-concept could be expected to increase activity in the mPFC as participants attempt to integrate negative, self-relevant information with their positive self-concept. However, we did not observe mPFC activation to the self-other contrast for negative scenarios. Notably, this is broadly consistent with our ERP results, where we also saw no effect associated with the cost of processing unexpected or incongruent negative information about the self. This may simply be due to the nature of our scenarios and paradigm. Although second person leads to reading text as being about the self all else being equal (Brunyé et al., 2009; Brunyé, Ditman, Mahoney, & Taylor, 2011), participants could easily simply reject our scenarios as being about the self once it was clear they were incongruent with what they believed about themselves, and therefore not engage much self-related processing.

On the other hand, the research discussed above (Garrett et al., 2014; Sharot et al., 2011) suggests that the mPFC is associated with belief updating in response to specifically positive, self-relevant information (see also Chavez & Heatherton, 2015, 2017), and that other regions (including ventrolateral PFC) are associated with the processing of negative self-relevant information. Thus, while the present results suggest that mPFC is modulated by the self-positivity bias, it will be important for future research to further examine the specific functional role of the mPFC and whether it is simply modulated by the self-positivity bias or plays a role in maintaining it.

Chapter 2: Toward an understanding of the late positive component of the ERP to emotional stimuli

Part I: Empirical and Theoretical Review of the Late Positive Component (LPC)

The last several decades have seen an increasing interest in the study of emotion. Why do we have emotional responses to certain stimuli? How do those responses affect the processing and encoding of stimuli? In cognitive neuroscience, the most prominent research in this area has been the use of fMRI to identify the neural networks involved in processing emotional stimuli. Initial work focused on identifying neural regions specialized for processing emotion (Phan, Wager, Taylor, & Liberzon, 2002), but more recent work has developed more sophisticated models of the interactions of cognition and emotion (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Parallel to this work, over the last decade or so, there has been a dramatic increase in the study of the event-related potential (ERP) response to emotional stimuli (Hajcak, Weinberg, MacNamara, & Foti, 2012).

ERPs are a measure of the electrical activity of the brain as recorded from the scalp. Although not well suited to identifying *where* in the brain processing is occurring like fMRI, ERPs are a much more direct assessment of brain activity than fMRI, because they measure the change in electrical potentials cause by synaptic activity. This measure is essentially instantaneous and highly temporally precise. The ERP is the average of the electrophysiological response across many trials and consists of a series of identifiable components that represent various aspects of the cognitive response to the eliciting event. ERPs are therefore very useful for understanding the mechanisms of cognitive processing and how they unfold over time (Luck, 2014).

Not surprisingly, emotional properties of stimuli have been shown to affect many components of the ERP, starting with early components that reflect sensory and attentional

processes (Citron, 2012; Hajcak et al., 2012). However, many of these effects have been inconsistent. By far the most consistent ERP effect to emotional stimuli has been the modulation of a positive component that peaks rather late in the stimulus-locked response, usually after 500 ms. As will be reviewed in detail below, this late positive component (LPC) is usually larger to emotional (positive and negative) stimuli than neutral stimuli, but is also modulated by many other factors, often in complex interactions.

A rather complex and often contradictory series of findings on this component has emerged. What is clear is that the LPC represents some kind of additional processing to stimuli that are socially and emotionally relevant. But many open questions remain. What exactly is the LPC sensitive to? And what is its functional role? In other words, what cognitive process within the brain is reflected by the scalp recorded LPC?

These questions are important. Given the large number of studies that have now reported an effect on the LPC and the large number of circumstances in which it has been elicited, it would appear to be a fundamental part of the cognitive response to emotional stimuli. Understanding this component would therefore tell us not just about the LPC itself and its use as a dependent measure, it would also tell us something much more general about how the brain responds to emotional stimuli and, as will be argued, motivationally relevant stimuli more generally. As a result, the meaning of the LPC is relevant not just to ERP researchers or even cognitive neuroscientists, but to anyone interested in social psychology, emotion, and cognition.

Here I present a detailed review of the LPC via examination of the ERP literature on emotional language. I then discuss proposals for the function of the LPC and suggest a new way forward for building a functional theory of this component by taking seriously its connections to the extensively studied P300 component. In Part II of this chapter, I describe an ERP study deriving from this framework.

Review of the LPC

The LPC is modulated by many types of emotional stimuli and has been widely examined in response to both pictures and words.[3] Although pictures are often seen as eliciting stronger emotional responses, they also present challenges. As the saying goes, "a picture is worth a thousand words": pictures are more complex and more difficult to control for a variety of factors from physical differences and visual complexity to semantic and social content. In addition, by manipulating sentence and discourse context, language offers a natural, powerful, and easy means to examine the effects of the emotional properties of stimuli in context, which, as will be seen, is crucial to understanding the LPC. The present review therefore focuses on studies that have examined the late positive component in response to linguistic stimuli (words, phrases, sentences, and discourse). This review represents a qualitative, but relatively systematic, review of this literature,[4] while also referring to and drawing insights from the picture literature where it is particularly relevant (for reviews, see Hajcak, MacNamara, & Olvet, 2010; Hajcak et al., 2012; Olofsson, Nordin, Sequeira, & Polich, 2008).

Basic characteristics and descriptions

Temporal and morphological characteristics

The emotional LPC is a positive-going deflection of the ERP elicited by words, pictures, and other stimuli. Effects on the LPC are nearly always largest at central and/or parietal electrodes, and this has held true for a variety of experimental designs as well as for mastoid, earlobe, nose, and average references. There is variability in the temporal characteristics of the component. In the word literature reviewed here, the LPC usually begins between 400 and 500 ms after the

[3] For studies directly comparing the LPC to words and to other types of stimuli see Frühholz, Jellinghaus, and Herrmann (2011); Hinojosa, Carretie, Valcarcel, Mendez-Bertolo, and Pozo (2009); B. L. Liu, Jin, Wang, and Hu (2010); Rellecke, Palazova, Sommer, and Schacht (2011); Tempel et al. (2013); and Vanderploeg, Brown, and Marsh (1987).

[4] Overall, 97 studies reporting an emotional LPC elicited by linguistic stimuli, reported in 80 separate papers, were examined for this review. Studies published or in press after mid-2014 are generally not included.

onset of an eliciting word (although effects have been reported beginning earlier) and lasts a few hundred milliseconds (but can last longer). A study in the picture literature has shown that the duration of the LPC is affected by stimulus duration (Gable, Adams, & Proudfit, 2015). This relationship may be less pronounced in the word literature, but in the studies examined for the present review there also appeared to be a positive relationship between SOA and LPC duration.

Component terminology

Several different terms have been used to describe late positive responses to emotional stimuli: P3 (or P300 or P3b), late positive component (LPC) or late positive complex, late positive potential (LPP), slow wave or positive slow wave, or simply late positivity. LPC and LPP have been the most commonly used of these. Which term is used seems to primarily reflect the authors' preference, interpretation, and/or theoretical assumptions: there is no clear relationship between the term used to describe a late positive component to emotional stimuli and any (physical or experimental) properties of the component itself. For consistency and clarity, I will use the term late positive component (LPC) regardless of the terminology used by the authors of the studies being discussed. However, I use this term descriptively and do not mean to imply any functional or theoretical interpretation by its use. Neither do I mean to imply that all findings were on exactly the same component or cognitive process, as this is an open question.

How many late positivities are there?

A potentially important open question is whether there is a single emotion-sensitive late positivity or multiple later positive-going components that can be responsive to emotional manipulations. This is an inherently difficulty question to answer. It can be very difficult to distinguish components both within and across ERP studies for a variety of reasons. A given cognitive process represented by a particular component can have a different latency, appearance, and scalp distribution depending on differences in other cognitive processes and components. On the other hand, because of the diffuse nature of ERP scalp distributions, two distinct

neurocognitive processes can evoke similar potentials (for discussion of distinguishing and interpreting ERP components, see: Coulson, King, & Kutas, 1998; Donchin & Coles, 1988; Kappenman & Luck, 2012; Luck, 2014, Ch. 2).

There is some evidence for the possibility of multiple emotion-sensitive late positivities. Some studies have reported emotion effects to more than one later positivity in the same waveform (Bernat, Bunce, & Shevrin, 2001; Dillon, Cooper, Grent-'t-Jong, Woldorff, & LaBar, 2006; Herbert, Kissler, Junghofer, Peyk, & Rockstroh, 2006; Knost, Flor, Braun, & Birbaumer, 1997; B. L. Liu et al., 2010; Naumann, Bartussek, Diedrich, & Laufer, 1992; Naumann, Maier, Diedrich, Becker, & Bartussek, 1997; Schapkin, Gusev, & Kuhl, 2000; Tempel et al., 2013; Vanderploeg et al., 1987), but in some of these cases visual inspection indicates that these responses could have been measured as a single component. It is also worth noting that the presence of more than one positivity to the same stimuli does not necessarily imply different processes/components: it is quite possible that the same cognitive process could be engaged or peak more than once in response to a stimulus (e.g., R. Johnson, Jr. & Donchin, 1985). Some work using principle components analysis (PCA) also suggests that the LPC might have two or more subcomponents (Delplanque, Silvert, Hot, Rigoulot, & Sequeira, 2006; Foti, Hajcak, & Dien, 2009; Gable et al., 2015; González-Villar, Triñanes, Zurrón, & Carrillo-de-la-Peña, 2014; Hajcak et al., 2010; MacNamara, Foti, & Hajcak, 2009; Vanderploeg et al., 1987; Weinberg, Hilgard, Bartholow, & Hajcak, 2012).

It should be noted that the current lack of clarity on distinct late positive responses does not make research on emotion-sensitive late positivities impossible or useless and it does not necessarily prevent generalizations. In fact, questions of component identity and distinction are common in the ERP literature (e.g, Coulson et al., 1998; Polich, 2012). In the case of the LPC, even when researchers have argued for multiple late positivities, they have generally interpreted these to be related and to reflect similar processes (Gable et al., 2015; Hajcak et al., 2012;

Weinberg et al., 2012). And, as discussed in Luck (2014, Ch. 4), often the best ERP designs do not rely on knowing the exact identity of a component to draw inferences. In any case, it will only be through further theory-driven research into the sensitivity and function of the LPC (see later in this chapter) that these questions can ultimately be resolved.

Properties of stimuli affecting the LPC

Nearly all studies in the ERP literature have used a dimensional model of emotion to study emotional words (for a discussion of dimensional models in contrast with other models, see Scherer, 2000). In particular, researchers have primarily worked with the popular two-factor model of emotion that argues that emotion varies along two fundamental dimensions, valence and arousal, which account for a large portion of the variance in the psychology and neuroscience of emotion (e.g., Bradley & Lang, 2007; Osgood, Suci, & Tannenbaum, 1957; Russell, 1980).[5] Valence describes how positive/pleasant/good versus negative/unpleasant/bad a stimulus is,[6] while arousal describes how arousing/exciting versus calming a stimulus is. Valence and arousal are by far the most researched properties of words that affect the late positivity, although, as we will see, this has not necessarily brought clarity on the effects of these dimensions.

Arousal

Studies that have compared emotionally valenced stimuli (pleasant and unpleasant) to neutral stimuli have almost always used neutral stimuli that were less arousing than the pleasant and unpleasant stimuli. The general pleasant/unpleasant > neutral effect on the LPC (discussed in detail in the next section) could therefore be interpreted as an arousal effect. In fact, some

[5] A few studies have examined words relating to particular emotions or categories of emotional experience (Carretie et al., 2008; Fogel, Midgley, Delaney-Busch, & Holcomb, 2012; Hinojosa, Carretie, Valcarcel, et al., 2009; Knost et al., 1997; Metzger, Orr, Lasko, McNally, & Pitman, 1997; Pauli, Amrhein, Muhlberger, Dengler, & Wiedemann, 2005; Ponz et al., 2014; Thomas, Johnstone, & Gonsalvez, 2007; Weinstein, 1995). But with a few exceptions (Briesemeister, Kuchinke, & Jacobs, 2014; Zhang & Guo, 2014), they have not controlled for valence and arousal and thus it is not possible to determine the effects of these particular emotions or categories independently.

[6] I will primarily use the terms pleasant and unpleasant rather than positive and negative in this review to avoid confusion with positive and negative ERP amplitude, but in my usage they are interchangeable.

researchers have suggested that the LPC is primarily modulated by arousal (e.g., Olofsson et al., 2008). There is some evidence in line with this view. Across many experimental designs more highly arousing stimuli do elicit a larger LPC, a study in the picture literature has shown the LPC to covary with autonomic arousal (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000), and a particularly well controlled word study showed that when many word properties were matched there was an effect of arousal but not valence on the LPC (although only with a task that didn't draw attention to valence: Delaney-Busch, Wilkie, & Kuperberg, 2016; see also Hofmann, Kuchinke, Tamm, Võ, & Jacobs, 2009).

However, studies in both the picture (e.g., Briggs & Martin, 2009; Ito, Larsen, Smith, & Cacioppo, 1998; Weinberg & Hajcak, 2010) and word (see Table 2.1) literature have shown differential LPCs to stimuli matched in arousal. In addition, the few studies in the word literature that have explicitly examined valence-independent effects of arousal have shown mixed results, with some showing a larger LPC to high arousal words (Bayer, Sommer, & Schacht, 2012; Delaney-Busch et al., 2016; Hofmann et al., 2009) and some failing to find significant effects of arousal (Bayer, Sommer, & Schacht, 2010; Citron, Weekes, & Ferstl, 2013; Recio, Conrad, Hansen, & Jacobs, 2014). Finally, context and task effects (discussed below) suggest that the LPC cannot be accounted for by properties of the eliciting stimulus alone.

Taken together, the literature suggests that arousal alone cannot account for LPC amplitude, but all else being equal more arousing stimuli will generally elicit a larger LPC.

Valence

Generally speaking, both pleasant and unpleasant stimuli elicit a larger LPC than neutral stimuli, although many studies have found at least one or both of these to not differ from neutral. Only a few studies have found overall null effects of valence; in many cases, these can be shown to be task or context effects and in other cases such explanations seem likely or plausible (see discussion below). The comparison between pleasant and unpleasant has yielded variable results.

Some of these effects are hard to interpret because of a lack of controls for various stimulus properties, particularly ratings of arousal between pleasant and unpleasant words.[7] However, as is clear in Table 2.1, every possible pattern (between pleasant and unpleasant words) has also been found in well-controlled stimuli with arousal-matched words. Thus, the LPC is clearly responsive to both pleasantly and unpleasantly valenced words, but also distinguishes between them in many cases. A number of ideas have been advanced to explain these valence effects.[8]

Researchers finding a larger LPC for unpleasant words have often discussed their effects in terms of a "negativity bias", the idea that, all else being equal, negative information has stronger effects on attention and cognition than positive information, possibly for fundamental evolutionary reasons having to do with survival. Although well supported by the behavioral literature (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Ito et al., 1998; Rozin & Royzman, 2001; Taylor, 1991), findings LPC to words have not strongly supported the idea of a negativity bias. As can be seen in Table 2.1, when stimuli are matched for arousal nearly equal numbers of studies

[7] A number of other lexical and semantic factors are often controlled: length, frequency, concreteness, imageability, orthographic neighborhood, etc. Of course, if these factors interact with emotional properties, simply holding them constant across conditions will not reveal the full story. Only a handful of studies have explicitly examined the interaction of emotional characteristics and other lexico-semantic properties of words. Emotion effects on the LPC have been found for all parts of speech for which they have been examined: adjectives, nouns, and verbs. Kissler, Herbert, Winkler, and Junghofer (2009) compared emotion effects for nouns and adjectives and did not find significant effects of the part of speech on the effects of emotion (see also Palazova, Mantwill, Sommer, & Schacht, 2011). A majority of ERP studies looking at emotional words have controlled for word frequency across conditions. To date studies examining both frequency and emotional variables have not found significant interactions between these variables (Delaney-Busch et al., 2016 [personal communication]; Palazova et al., 2011). Finally, a few studies have examined the interaction of concreteness and emotional properties and found varying results (Delaney-Busch et al., 2016; Kaltwasser, Ries, Sommer, Knight, & Willems, 2013; Kotz & Paulmann, 2007).

[8] One possible explanation for a highly variable pattern of effects is that null hypothesis is true. In such a case, we would expect a null result 95% of the time and a significant effect (with random directionality) 5% of the time. Publication bias (the "file drawer problem") could then inflate the number of significant effects such that the literature would evince a more equal pattern like that seen here. However, that is less likely to be a significant problem in this literature. This is because finding a null effect in the pleasant-unpleasant comparison should not make a paper difficult to publish as long as one or both differ from neutral, as is the case in most studies. So the distribution shown in Table 2.1 (which represents the least confounded studies) probably represents something close to the true distribution of effects that have been found. This distribution is more consistent with a situation in which not all relevant variables have been taken into account than a situation of a true null effect, as discussed later in this paper.

have found larger LPC to pleasant words as to unpleasant words.[9] In one framework, the negativity bias is complemented by a "positivity offset" (Ito et al., 1998). The idea is that the positive motivational system responds more strongly at low levels of arousal while the negative motivational system responds more strongly at higher levels of arousal. Some researchers have proposed that words simply are not as arousing as pictures (even when they receive the same arousal ratings in norming studies) making the positivity offset more relevant for word stimuli (Kissler, Assadollahi, & Herbert, 2006; B. L. Liu et al., 2010). However, there is evidence pointing against this idea: many studies have seen a larger LPC to unpleasant words than pleasant words and studies finding a larger LPC to pleasant words have not necessarily used less arousing stimuli. In addition, B. L. Liu et al. (2010) compared pictures and words to explicitly examine the lower arousal for words hypothesis and did not find support for it.

One explanation for the variability in the effects of valence may be a difference between subjects' perception of the stimuli and the ratings used to categorize them. Valence (as well as arousal and other emotional properties) is clearly a subjective property of words. Differences between the subjects completing stimulus norming studies and subjects in ERP studies or differences between the contexts in which these two groups of subjects encounter the stimuli may lead to different subjective experiences of emotional properties. There is some evidence to support this. Delaney-Busch et al. (2016) presented pleasant, unpleasant, and neutral words matched for arousal and a number of other word properties and found that unpleasant words elicited a larger LPC but pleasant and neutral did not differ. However, when stimuli were recategorized based on individual participants' valence categorizations, both unpleasant and

[9] Interestingly, studies examining words in sentence or discourse context (discussed in following sections) seem particularly likely to show a larger LPC to unpleasant words (see Table 2.1). However, the total number of such studies is small, and a chi square test for independence show that this pattern cannot be distinguished from coincidence at $\alpha = .05$. No study has yet provided the relevant comparison by directly examining the same set of words (in the same participants) in isolation and in context.

pleasant stimuli elicited a larger LPC than neutral stimuli without differing from each other. In addition, a signal detection analysis of behavioral data indicated that participants had a harder time distinguishing between pleasant and neutral words than they did distinguishing unpleasant words from other words (N. Delaney-Busch, personal communication). Thus inexact measures of actually experienced valence, arousal, etc. may account for any number of contradictory findings in the literature, and this is an important reminder of the challenges of studying a subjective phenomenon (for additional evidence on the importance of idiosyncratic emotional interpretation, see Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009).

Another likely explanation for variability in emotion effects is that there are simply stimulus characteristics that differ between studies that have not been accounted for. For example, one property that may be important is ambiguity. Because we cannot know how to prioritize stimuli in the environment until we can place a value on them, it has been argued that ambiguous stimuli can be assessed as more motivationally significant than pleasant or unpleasant stimuli until they are disambiguated (Hirsh & Inzlicht, 2008; Sokolov, Spinks, Näätänen, & Lyytinen, 2002). This idea has received support in ERP studies examining the feedback-related negativity (FRN; Gu, Ge, Jiang, & Luo, 2010; Hirsh & Inzlicht, 2008) and the LPC (Tritt, Peterson, & Inzlicht, 2012). It is likely that neutral stimuli are often more ambiguous than pleasant or unpleasant stimuli, and this could account for a few studies that have seen a larger LPC to neutral stimuli than emotional stimuli (e.g., Citron et al., 2013; see also Fields & Kuperberg, 2012, 2016).

To summarize, it is clear that there is nearly always an effect of emotional valence on the LPC and that this effect is driven by emotionally valenced stimuli eliciting a larger LPC than neutral stimuli. However, whether pleasant or unpleasant stimuli elicit the larger LPC and whether only one or both valences differ from neutral is highly variable. This is true even in studies that are controlled for the most relevant factors given our current knowledge. There are a number of possible reasons for this, but they all come back to the fact that there are many interacting factors

that affect the amplitude of the LPC. In addition to stimulus factors other than valence and arousal, these include effects of context and task, which are discussed in the following sections.

Table 2.1: Studies comparing pleasant and unpleasant words matched in arousal

Experiment	Stimuli	Task	Neg-Pos Comparison
Bailey & Chapman (2012)	single words	memory	POS>NEG
Bayer, Sommer, & Schact (2012)	single words	lexical decision, silent reading	POS>NEG
Carretie et al. (2008)	single words	lexical decision	NEG=POS
Delaney-Busch & Kuperberg (2013)	discourse	comprehension questions	NEG>POS
Delaney-Busch (2016), Exp. 1	single words	semantic categorization	NEG=POS
Delaney-Busch (2016), Exp. 2	single words	valence categorization	NEG>POS
Dillon et al. (2006)	single words	arousal rating	NEG=POS
Fields & Kuperberg (2016), Exp. 2	discourse	comprehension questions	NEG=POS
Fischler & Bradley (2006), Exp. 1	single words	valence categorization	NEG=POS
Fischler & Bradley (2006), Exp. 2	single words	emotionality categorization	NEG=POS
Fischler & Bradley (2006), Exp. 3	single words	silent reading	NEG=POS
Fischler & Bradley (2006), Exp. 4	single words	semantic categorization	NEG>POS
Fischler & Bradley (2006), Exp. 6	noun phrases	coherence judgment	NEG=POS
Fischler & Bradley (2006), Exp. 7	noun phrases	semantic categorization	NEG=POS
Herbert et al. (2006)	single words	memory	POS>NEG
Herbert, Herbert, et al. (2011)	noun phrases	silent reading	POS>NEG
Herbert, Junghofer, & Kissler (2008)	single words	silent reading	POS>NEG
Herbert, Pauli, & Herbert (2011)	noun phrases	silent reading	NEG>POS
Hinojosa, Mendez-Bertolo, & Pozo (2010)	single words	lexical decision	NEG=POS
Hofmann et al. (2009)	single words	lexical decision	NEG=POS
Holt, Lynn, & Kuperberg (2009), Exp. 1	discourse	valence categorization	NEG>POS
Holt, Lynn, & Kuperberg (2009), Exp. 2	discourse	comprehension questions	NEG>POS
Kaltwasser et al. (2013)	single words	concreteness judgment	NEG=POS
Kanske & Kotz (2007), Exp. 1	single words	lexical decision	NEG=POS
Kanske & Kotz (2007), Exp. 2	single words	lexical decision	NEG>POS
Kissler et al. (2009)	single words	silent reading, count adjectives, count nouns	POS>NEG
Kuperberg et al. (2011)	discourse	valence categorization	NEG>POS
Schacht & Sommer (2009b), Exp. 2	noun-verb phrases	lexical decision, relatedness judgment, font judgment	NEG=POS
Shestyuk & Deldin (2010), Exp. 1	single words	social judgment (other or self)	POS>NEG
Zhang & Guo (2014)	single words	semantic categorization	NEG>POS

Task effects

There have been several studies that have compared effects of emotion on the LPC with the same stimuli under different tasks. A number of studies have shown that tasks that lead to shallower processing or draw less attention to emotional attributes elicit smaller or null effects, while tasks that make emotional properties of stimuli task-relevant lead to larger effects of emotion. In an early study, Naumann et al. (1997) presented unpleasant and neutral nouns and showed a larger LPC for unpleasant nouns when subjects made a valence judgment, but not in structural decision or concreteness decision conditions. Similarly, but with words in discourse context, Holt et al. (2009) found a smaller effect of emotion with a semantic task than a valence categorization task. Fischler and Bradley (2006) report a series of five studies using the same stimuli but different tasks. They found effects of emotion for valence categorization, emotion categorization, and semantic categorization, with the largest effects seen in the emotion-related tasks. Interestingly, they also showed an effect of emotion for silent reading (i.e., no additional task) but not during a lexical decision task, demonstrating that certain tasks may actually draw attention away from emotional and semantic aspects of stimuli (see also González-Villar et al.; Kissler et al., 2009; Sass et al., 2010; Scott, O'Donnell, Leuthold, & Sereno, 2009). However, it should be noted that several studies have shown emotion effects with lexical decision tasks (Carretie et al., 2008; Hofmann et al., 2009; Kanske & Kotz, 2007; Schacht & Sommer, 2009a; Williamson, Harpur, & Hare, 1991), including Schacht and Sommer (2009b) who showed an effect of emotion for a lexical decision task and a semantic relatedness task, but not for a structural (font judgment) task.

While the studies above found changes in the *magnitude* of effects with different tasks, other studies have shown changes in the *pattern* of effects as a function of task. The clearest demonstration of this comes from Delaney-Busch et al. (2016) who showed an effect of arousal (with no effect of valence) with a semantic categorization task but an effect of valence (with no effect of arousal) with a valence categorization task. Fields and Kuperberg (2016) have shown a

more complex effect of task on an interaction between self-relevance and emotion that will be discussed below. Taken together, these and other studies (e.g., Fischler & Bradley, 2006; Holt et al., 2009) show that specific task demands can lead to qualitatively different patterns of effects on the LPC.

Context effects

One of the strengths of using language as stimuli is the ability to examine the effects of processing emotional stimuli in context in ways that are comparable to real-world social situations. Sentences and broader discourse can be used to describe any variety of real-world contexts and differences in these contexts can be manipulated in arbitrarily subtle and complex ways. In the following sections, I review studies examining such manipulations as well as effects of the broader experimental context.

Local context (phrases, sentences, discourse)

Although most studies have examined single emotional words, a number of studies have now examined emotional words in the context of phrases, sentences, and discourse (i.e., more than one sentence). These function as a manipulation of the local context for the stimulus. Broadly speaking, the same effects seen in individual words are seen to words in context when the context is held constant. For example, in neutral, non-emotionally-constraining sentence/discourse contexts, Holt et al. (2009; see also Kuperberg et al., 2011) and Bayer et al. (2010) have each shown effects of emotion on the LPC that are indistinguishable from effects to single words (but see Footnote 9 on the possibility that context increases the likelihood of seeing a negativity bias).

What about manipulations of the context? In two-word noun phrases (e.g., “dead tyrant”, “dead puppy”) that were fully crossed for valence, Fischler and Bradley (2006) did not see emotion effects on the first word under task conditions that lead subjects to read the words as a phrase. When subjects evaluated each word individually, however, the first words elicited a standard emotion effect on the LPC. Fischler and Bradley argue that when word pairs are perceived as a

phrase, evaluative processes are withheld until the whole phrase is perceived. This may explain the results of Clegg et al. (2010), who did not find robust effects to emotional words in the first sentence of two-sentence vignettes: in this case there was still significant information about the scenario to come (see also Ding, Wang, & Yang, 2014).

Delaney-Busch and Kuperberg (2013) examined the effect of emotional context on the processing of emotional words. In their study, two-sentence vignettes were presented in which the first sentence contained a key pleasant or unpleasant word and the critical word in the second sentence either matched or opposed the first sentence in valence, e.g.: *Lucy was a(n) awful/great engineer. Her creations were big failures/successes every time.* Interestingly, they found a main effect of the valence of the critical word in the second sentence, but this did not interact with the valence of the first sentence. However, under some circumstances congruity effects between context and word valence can be observed. Bartholow, Fabiani, Gratton, and Bettencourt (2001) presented a paragraph context implying either a positive or negative trait for a fictional person. A final sentence ended with a critical word that was either congruent or incongruent in valence with the implied trait. In addition to a main effect of valence (negative>positive), the ERPs to the final word showed a larger LPC for the incongruent condition (see also Bartholow, Pearson, Gratton, & Fabiani, 2003; Van Duynslaeger, Sterken, Van Overwalle, & Verstraeten, 2008; Van Overwalle, Van den Eede, Baetens, & Vandekerckhove, 2009). Because the implied traits had semantics beyond their valence and because the valence of the context was clearly associated with a character in the situation model, Bartholow et al.'s contexts were likely more constraining (emotionally and otherwise) than Delaney-Busch & Kuperberg's and this may explain their differing results (see Delaney-Busch & Kuperberg, 2013 for further discussion of this point).

Global context

The LPC is also sensitive to more global aspects of context, such as the nature of the other stimuli and trials in an experiment. For example, the LPC is sensitive to a version of the

oddball effect. The oddball effect is classically studied with simple stimuli on the P300 component. When a stimulus stands out from the preceding stimuli (i.e., is an oddball), it elicits a larger P300 (Polich, 2012). The emotion-sensitive LPC shows a similar effect. For example, Crites, Cacioppo, Gardner, and Bernston (1995) showed that unpleasant stimuli showed the largest LPC when embedded in a stream of pleasant stimuli but pleasant stimuli showed a larger LPC when unpleasant stimuli were the majority (see also Cacioppo, Crites, Berntson, & Coles, 1993; Cacioppo, Crites, Gardner, & Berntson, 1994).

A particularly interesting example of stimulus context comes from Fogel et al. (2012). Using stimuli similar to Delaney-Busch et al. (2016), they showed an effect of emotion on the LPC. However, when highly charged taboo words were included in the stimulus set, the difference between the standard emotional words and neutral words was no longer significant; only the taboo words elicited a large LPC. Presumably stimuli that had once seemed engaging and motivationally significant became more like distractors when the taboo words were included (a type of anchoring effect), which shows the important influence of context on what will be marked as motivationally significant (see also the effects of self-relevance in the following section).

Self-relevance

Another important context variable that is likely to modulate emotional processing is social relevance. In single word studies, a word like “murder” is presented in isolation. But who is being murdered and why? The answer to this question could have significant effects on the activation of emotion circuits.

To test this idea, a few studies have examined effects of self-relevance on emotional word processing. In two studies Herbert and colleagues (Herbert, Herbert, Ethofer, et al., 2011; Herbert, Pauli, et al., 2011) have examined ERPs to pleasant, unpleasant, and neutral nouns preceded by “my” (self-relevant), “the” (control), or “he” (other relevant; only used in the Herbert, Herbert et al. study). They found effects of emotion on the LPC only for nouns preceded by “my” (although the

exact pattern of these effects differed across studies; see also, Li & Han, 2010; Schindler, Wegrzyn, Steppacher, & Kissler, 2014; Shestyuk & Deldin, 2010).

Fields and Kuperberg (2016) have conducted a study examining the influence of self-relevance on emotional word processing in larger discourse context. In this studies, we presented scenarios such as: *Someone knocks on **Sandra's/your** hotel room door. **She/you** see(s) that he has a gift/tray/gun in his hand.* ERPs were recorded to the emotional word (underlined above) in the second sentence. Pleasant and unpleasant words showed a larger LPC than neutral words (and did not differ from each other) in the self-relevant condition; however, there was no effect of emotion on the LPC in the other-relevant condition. This, along with the Herbert and colleagues studies and others (Li & Han, 2010; Schindler et al., 2014; Shestyuk & Deldin, 2010), show that self-relevance can have effects similar to those seen for taboo words in Fogel et al. (2012) (reviewed in the preceding section): when self-relevant emotional words are included in the experimental context, non-self-relevant words may become more like distractors among more motivationally relevant stimuli and lose their ability to draw special attention and processing. Other people's problems matter less when you have your own problems, so to speak.

However, task has a significant effect on this interaction. In another study using the same stimuli (Fields & Kuperberg, 2012), we used a production task (produce a third sentence continuing the scenario). This study showed an interaction between emotionality and self-relevance wherein the LPC for self-relevant neutral words was larger than non-self-relevant neutral words, but there was no effect of self-relevance for pleasant or unpleasant words. We argued that this effect might be due to attempts to assess the valence of the neutral words, many of which were more ambiguous than the pleasant or unpleasant words. This was likely to be particularly important with this particular task in the self-relevant condition because of participants' desire to produce a continuation consistent with their self-concept (Swann, 2011), whereas there was no motivation to go beyond the easiest or most salient interpretation in the other-relevant

condition (see discussion in Fields & Kuperberg, 2012, 2016). Thus, these studies show a complex interaction of factors that have been discussed thus far: emotional properties of words, local context, global context, and task (see also Fields, 2015).

Toward a Theory of the LPC

As will now be clear, the LPC is consistently modulated by the emotional properties of words, but the exact nature of these effects is variable. There are many discrepancies and conflicting findings, and these are not always easy to explain. How can we make sense of this literature?

As a useful staring point, we might ask: Is the LPC an emotion-specific component or is it sensitive to something broader that is, in turn, affected by or correlated with emotion? Although it has sometimes been described as representing a process of emotional evaluation, researchers have more often taken the latter view. The literature reviewed above clearly shows that there is no one-to-one relationship between any given emotional property of a stimulus and LPC amplitude. Rather, it is clear that the cognitive processes reflected by the LPC are modulated by multiple interacting factors that are likely to play important roles in real-world social contexts.

A common view of the LPC is therefore that it is increased to the extent that a stimulus is *motivationally significant*. Emotional stimuli are inherently motivationally significant due to their connection to enduring biological and social motivations. Indeed, emotions have been referred to as “relevance detectors” (Frijda, 1986) with the idea that a core function of emotions are to tell us which stimuli most likely to be relevant for our goals and motivations. However, the motivational significance of a stimulus will also depend on the context. The same word read in a self-relevant context will have a different motivational significance than in a non-self-relevant context. In the same vein, the word “puppy” will likely have different levels of motivational significance in the phrase “I saw a puppy” and “I adopted a puppy”. More broadly, the motivational significance of a given stimulus is relative and will depend on the other stimuli in a particular context: a spider

discovered in your living room may dominate your attention under normal circumstances, but if your house is on fire you may not notice it at all. An experimental analogue is Fogel et al.'s (2012) finding that there is no standard effect of emotion on the LPC in an experiment that included taboo words (see discussion above).

Finally, the motivational significance of a stimulus will depend on the motivations of the participant, which will differ depending on the task they are engaged in (as well as individual differences). For example, a task that requires an emotional judgment will make the emotional properties of words more relevant and thus it is not surprising that such tasks tend to increase the effect of emotion on the LPC. In addition, which emotional properties modulate the LPC can depend on which properties are made relevant by the task (Delaney-Busch et al., 2016) and some tasks may also make properties of neutral words particularly relevant (Fields & Kuperberg, 2016).

Importantly all these various factors can interact in complex ways to determine the motivational significance of a stimulus (e.g., Fields & Kuperberg, 2016). In addition, it is unlikely that we have uncovered or fully understand all the factors that are involved. It is therefore not surprising that the literature contains much variability and inconsistency. However, a purely empirical approach of examining every plausible factor and interaction that might affect the LPC may be a never-ending task. Instead, what are needed are better theories of the LPC, that can guide research and help us to understand *why* the various factors have the effect they do on the LPC.

What is likely to be particularly helpful in making progress in understanding the LPC, and therefore understanding the processing of emotional stimuli more generally, is understanding the *function* of the LPC. To say that the LPC is modulated by motivational significance is simply to summarize and organize of all the various factors that affect its amplitude. But the question remains: what cognitive process is being modulated by motivational significance and all the factors that contribute to it? That is the question I turn to in the following section.

What is the function of the LPC?

The corollary to the question raised above about whether the LPC is modulated specifically in response to emotional factors is the question of whether it represents emotion-specific neural processes. Some researchers have in fact linked the LPC to (emotional and motivational) evaluative or appraisal processes (Citron et al., 2013; Herbert, Herbert, Ethofer, et al., 2011; Herbert et al., 2008; Herbert, Pauli, et al., 2011; Hinojosa, Carretie, Mendez-Bertolo, Miguez, & Pozo, 2009; Kotz & Paulmann, 2011) or the activation of motivational circuits (Hinojosa, Carretie, Valcarcel, et al., 2009; Hinojosa et al., 2010).

However, researchers have more commonly regarded the LPC as reflecting more domain general cognitive processes that are modulated by the *output* of emotion evaluation circuits. In many cases, researchers have remained very broad and vague about the LPC's function, describing it as simply reflecting sustained or elaborated processing, engagement with the stimulus, or processing load (e.g., Crites et al., 1995; Fields & Kuperberg, 2012; Herbert et al., 2008; Herbert et al., 2006; Kaltwasser et al., 2013; Palazova et al., 2011; Schacht & Sommer, 2009b; Shestyuk & Deldin, 2010). Somewhat more specifically, perhaps the most common view of the LPC is that it represents increased or sustained attention to emotional stimuli (Carretie et al., 2008; Fields & Kuperberg, 2012; Fischler & Bradley, 2006; Gable et al., 2015; Herbert, Herbert, Ethofer, et al., 2011; Herbert et al., 2008; Herbert, Pauli, et al., 2011; Hinojosa, Carretie, Mendez-Bertolo, et al., 2009; Hinojosa, Carretie, Valcarcel, et al., 2009; Hinojosa et al., 2010; Holt et al., 2009; Kissler et al., 2009). It has also been linked to increased sensory and/or perceptual processing (Bayer et al., 2010; Bradley & Lang, 2007), mental imagery (Kanske & Kotz, 2007), and stimulus encoding and/or memory encoding (Herbert, Herbert, Ethofer, et al., 2011; Herbert et al., 2008; Herbert, Pauli, et al., 2011; Hinojosa, Carretie, Mendez-Bertolo, et al., 2009; Kissler et al., 2009; Shestyuk & Deldin, 2010).

Obviously, this is a wide range of processes, encompassing much of cognition. In addition, rarely have researchers been very specific. As Donchin (Donchin, 1981; Donchin & Coles, 1988) argued years ago, a high-level psychological construct cannot really explain the function of an ERP component. For example, there is no specific neural process or mechanism that corresponds to a concept like "attention". Instead, attention is a general psychological concept that refers to the purpose and outcomes of many different neural processes. ERP components are likely to reflect these more specific computational mechanisms (and, not surprisingly, there are many ERP components associated with attention: Luck, 2012; Luck & Kappenman, 2012).

Unfortunately, the specific functional meaning of the LPC has not generated much discussion or empirical investigation in the literature. However, a common framework for understanding the LPC has been to regard it as being related to and functionally similar to the P300 component of the ERP (e.g., Bartholow et al., 2001; Cacioppo et al., 1993; Cacioppo et al., 1994; Citron, 2012; Fields & Kuperberg, 2016; Hajcak et al., 2010; Hajcak et al., 2012; Kissler et al., 2009; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Olofsson et al., 2008; Schacht & Sommer, 2009a; Weinberg et al., 2012). And while this connection has often been made, its implications have scarcely been explored. In the following section, I argue that this relationship may point to a functional interpretation of the LPC.

The LPC and the P300

The P300 is perhaps the most studied component of the ERP.[10] The terminology of ERP components provides a potential point of confusion here. Despite its name, the P300 does not necessarily peak at 300 ms. In fact, it only peaks around 300 ms for simple, easy to discriminate stimuli; in response to more complex experimental manipulations, such as manipulations of the semantic content of words, it often peaks within the LPC time window (e.g., Fabiani, Karis, &

[10] The term P300 can be used to refer to multiple components (or subcomponents). When unqualified, it generally refers to the P3b component (see Polich, 2012), and that is how I use the term here.

Donchin, 1986; Kutas, McCarthy, & Donchin, 1977). In addition, like the LPC, the P300 appears as broadly distributed, parietally centered positivity. Thus, the P300 and LPC are temporally and morphologically very similar.

The P300 is commonly elicited by "oddball" paradigms where stimuli of two (or more) categories are presented with one category being rare and the other being common. In such a paradigm, the amplitude of the P300 increases as the probability of a stimulus category decreases. More generally, the amplitude of the P300 is inversely related to the expectancy (or surprise) of a stimulus, which is of course influenced by aspects of the context and sequence of stimuli, not simply global or objective probability (Donchin & Coles, 1988; R. Johnson, Jr., 1986; Squires, Wickens, Squires, & Donchin, 1976). The P300 is also larger to task-relevant stimuli than to stimuli that require no response or can be ignored. These two factors, stimulus probability (or expectancy) and task-relevance are the most well-known and studied determinants of P300 amplitude, but it is known to be elicited in many circumstances and influenced by many factors, particularly factors associated with attention to and encoding of a stimulus. For example, one attempt to summarize the various factors affecting P300 amplitude is Johnson's (1986) classic "Triarchic Model", which proposed that the amplitude is determined by the following formula.

P300 amplitude = ((1 – EQ x AT) x [(GP + SE) + (TC + SC +SV)])

Where: *EQ* is equivocation, *AT* is attention, *GP* is global probability, *SE* is sequence effects, *TC* is task complexity, *SC* is stimulus complexity, and *SV* is stimulus value.

Thus, both the P300 and LPC are highly dynamic components that are modulated by the particular goals imposed by a task, as well as several aspects of local and global context. The effect of emotion on the LPC has often been seen as analogous to the effect of task-relevance on the P300. Whereas the P300 is larger to "targets" (stimuli that require a response), emotional stimuli, have been argued to be "natural targets" due to their inherent motivational significance (Hajcak et al., 2010). That is, a target in a traditional P300 paradigm is made motivationally relevant by the task, but emotional stimuli are naturally motivationally relevant and do not need to

be rare or task-relevant to draw the additional processing reflected by the LPC. This fits well with thinking on the P300: for example, in Johnson's model task-relevance is related to the attention factor, and it is of course known that emotional stimuli draw attention (Compton, 2003). Other relationships are also possible. For example, emotional stimuli could also be related to the stimulus value factor. P300 studies supporting this aspect of the model have manipulated value via differing monetary reward for different stimuli (reviewed in R. Johnson, Jr., 1986); emotional stimuli are, presumably, naturally valuable.

In addition to both components being highly sensitive to task and context, the P300 and LPC may relate to similar outcomes. A key attribute of the P300 is that stimuli that elicit larger P300 amplitude tend to be remembered better (Fabiani & Donchin, 1995; Fabiani et al., 1986; Fabiani, Karis, & Donchin, 1990; Kamp, Brumback, & Donchin, 2013; Karis, Fabiani, & Donchin, 1984; A. D. Wagner, Koutstaal, & Schacter, 1999). Although memory effects have not been directly investigated with regard to the LPC, it is well known that emotional stimuli are remembered better than non-emotional stimuli (Hamann, 2001; Kamp, Potts, & Donchin, 2015; Laney, Campbell, Heuer, & Reisberg, 2004).

Thus, the P300 and LPC appear as very similar components on the scalp (i.e., they are temporally and morphologically similar), are sensitive to many of the same or related experimental variables, and share a relationship to key cognitive processes such as attention and memory. This suggests that the relationship between the LPC and the P300 should be taken more seriously. It seems plausible that the emotional LPC and the P300 share a common underlying computational mechanism, or at least that functional theories of the P300 may be useful guides for building a functional theory of the LPC.

Theories of P300 function

The classic functional theory of the P300 is Donchin's context updating model (Donchin, 1981; Donchin & Coles, 1988), which proposes that the P300 is related to building accurate and

useful models of a given context. Clearly, at any given time we must have some model or schema of our current environment. This model allows us to predict what is likely to happen, which actions are strategic for pursuing our goals, et cetera. Since our environment constantly changes, usually in ways that are not fully predictable, we must make use of incoming sensory information to update our model. In Donchin's own words, the context updating account proposes simply "that the P300 is elicited by processes associated with the maintenance of our model of the context of the environment" (Donchin & Coles, 1988, p. 370). It is therefore larger to stimuli that require greater updating of the context model. Unexpected events elicit greater context updating because they provide more new information than expected events and act as signals that our environment may be different than we thought it was. Task relevance modulates this effect because our model of the environment is tailored to our goals and motivations. That is, we don't attend to or encode all aspects of our environment equally; instead, we are trying to build a model of the context that helps to achieve our goals. The relationship of the other factors influencing P300 amplitude to context updating is often similarly intuitive.

It is also intuitive that emotional stimuli might be associated with this sort of context updating process. First, emotions are often elicited by stimuli that are either better or worse than expected, and thus emotions may serve as a signal that our environment has changed or is different than we thought. Second, in a complex, noisy and ever-changing environments, emotions act as "relevance detectors" (Frijda, 1986), telling us which information is relevant to our goals and motivations and thus which information is most important to integrate into our context model. Thus, whereas the P300 is often said to be modulated by task relevance, perhaps it is more accurate to say that it is modulated by motivational relevance, which could subsume task relevance and expectancy as subcomponents.

Although it remains the most prominent account of the P300, Donchin and colleagues' particular formulation of context updating was not further developed after the late 1980's. However,

in recent years there has been renewed interest in theories of the P300, many of which build in some way upon the context updating account. In some cases, these theories have already taken responses to emotional stimuli into account or are easily adapted to emotional stimuli. Nevertheless, the emotional ERP literature has not taken much note of them.

Nieuwenhuis et al. (2005) proposed a neurobiological theory of the P300 that fits nicely with the LPC literature. In fact, these authors do not clearly differentiate between the classic P300 and late positivities to emotional stimuli (which had been investigated much less at the time). They propose that the P300 is generated as a result of widespread release of norepinephrine in response to motivationally relevant stimuli. Much like the discussion above, Nieuwenhuis et al. see surprise, task relevance, and emotion as each contributing to motivational relevance. The effect of NE on cortical structures is to increase signal to noise ratio in the service of better processing of significant stimuli. The resulting increase in synaptic activity, primarily in temporal and parietal regions, generates the P300. Polich (2007) builds on the work of Nieuwenhuis et al. to propose that the P300 represents specifically inhibitory processes. Based on a biased competition framework, Polich argues that unexpected and important stimuli lead to inhibition of other processes via the effects of NE release. In this framework, the P3b (i.e., the classic P300) in particular is the result of inhibition in the service of memory updating processes in the temporal and parietal lobes, which is broadly consistent with the context updating account.[11]

Although Nieuwenhuis et al. (2005) initially contrasted their account to the context updating theory, in more recent work Nieuwenhuis (2011) has suggested this may have been a mistake. Instead, he has emphasized the similarities between these two accounts by examining the relationship of the NE system in learning and belief updating. He also links these theories to another literature that may be important to understanding both the P300 and LPC, the Bayesian

[11] This account is also compatible with Mather and Sutherland's (2011) arousal biased competition theory, which asserts the emotionally arousing stimuli sharpen biased competition. It may be that such effects are mediated by the norepinephrine system.

literature on predictive cognition (particularly the work of Dayan & Yu, 2006). Like the context updating theory, Bayesian theories are fundamentally concerned with the way in which incoming stimuli are used to update internal models that improve predictions of future events (Clark, 2013; Courville, Daw, & Touretzky, 2006; Feldman & Friston, 2010; Kuperberg & Jaeger, 2016; Perfors, Tenenbaum, Griffiths, & Xu, 2011; Qian, Jaeger, & Aslin, 2012; Yu & Dayan, 2005). However, these theories propose testable and implemented models, which make possible much more specific predictions about the representational format of the context model and the mechanisms involved in the context updating process. While the connection between the P300 and Bayesian theories of cognition has been noted in the literature (De Swart, Kok, & Das-Smaal, 1981; Kopp, 2008; Nieuwenhuis, 2011; Yu & Dayan, 2005), no detailed Bayesian theory of the P300 has been proposed and there have been no connections made between Bayesian theories and the LPC to my knowledge. Developing specific and implemented computational models that can account for the complex P300 and LPC literatures will be an important task for future work, and such models may lead to new theoretical insights and hypotheses.[12]

Summary

What I have provisionally suggested is that the P300 and LPC may be the same or closely related components of the ERP. If this is true, it would give us a much stronger idea of the neurobiology of the LPC. Specifically, it would suggest that when a stimulus is identified as motivationally relevant, it leads to widespread norepinephrine release from the locus coeruleus and the LPC represents subset of the activity generated or enhanced by this release, likely in the parietal and temporal cortices (Nieuwenhuis et al., 2005; Polich, 2007). Identifying the LPC as a form of the P300 would also give us a much clearer idea of the cognitive process represented by the LPC. It would suggest that the LPC is related to processes of updating internal models in

[12] Spratling (2008) has shown the mathematical equivalence of Bayesian predictive coding accounts and biased competition theories, which may represent a way to integrate Bayesian approaches with Polich's (2007) inhibition account of the P300.

order to improve our understanding of the situation and the processing of future stimuli (Donchin, 1981; Donchin & Coles, 1988; Kopp, 2008; Nieuwenhuis, 2011; Yu & Dayan, 2005). Note that this contrasts somewhat with most common account of the LPC as being related to attention. Although LPC amplitude and attention would be closely related in this account, it suggests that the LPC is more directly reflects processes associated with memory and learning than simply attention per se.

It is important to note that all of these claims rely on many assumptions and are relatively speculative at this point. It is also important to note that the lack of a clear functional theory does not mean that past and current research examining the LPC is not useful (for example, research using the P300 was and is able to answer many research questions without reference to a functional theory; see discussion in Donchin, 1981; Polich, 2012). However, such an account would greatly inform theories of how emotion affects cognitive processing and expand the usefulness of the LPC as a dependent measure.

Part II: An ERP study of stimulus probability and emotion

A first step to building a stronger combined theory of the P300 and LPC is to more clearly establish the relationship (and/or differences) between these components. Although they have often been seen as similar in the literature, and sometimes even as the same component, there has not been much attempt to test this idea empirically. Here I describe an initial study intended to test this theory in two ways.

First, to the extent that the LPC and P300 are the same component or share a closely related computational function, the LPC should be sensitive to the experimental factors that modulate the P300. In some cases, such as the effect of task-relevance, this has been confirmed in the literature (see above). But the most prominent factor affecting the P300, stimulus expectancy, has not been clearly examined in the LPC literature. By examining both emotion and

expectancy in the same experiment, with the same stimuli and subjects, the ERPs elicited by each factor could be directly compared. In addition, the nature of any interaction effect between these factors may provide insights into a unified theory of the P300 and LPC.

Theory and research on the P300 suggests a second hypothesis: the LPC should be related to later memory for a stimulus. The context updating account argues that the P300 is associated with memory updating processes. Thus one of the first major predictions of this theory was that stimuli that evoked a larger P300 would be better remembered (Donchin, 1981). This prediction was subsequently confirmed in numerous studies (Fabiani & Donchin, 1995; Kamp et al., 2013; Karis et al., 1984; A. D. Wagner et al., 1999), and must be accounted for by any theory of the P300. Indeed, recent theoretical work on the P300 has continued to emphasize its functional role in learning and memory (Kopp, 2008; Nieuwenhuis, 2011; Polich, 2007; Yu & Dayan, 2005). But while emotional stimuli are known to be better remembered than neutral stimuli (Hamann, 2001), the relationship between LPC amplitude and memory has not been tested directly (although at least one study has looked at the relationship between the P300 elicited by emotional isolates and memory: Kamp et al., 2015). Because it is possible to observe the relationship between the P300 and memory with an unexpected memory test after an oddball paradigm (Fabiani et al., 1986), hypotheses about memory can be tested within the emotional oddball paradigm intended to test the hypotheses described above.

The present study

The present study is based on classic oddball studies from the P300 literature (e.g., Duncan-Johnson & Donchin, 1977; Fabiani et al., 1986). Participants were presented with blocks of neutral and negative words with the proportion of each category varying across blocks. Their task was simply to categorize each word as negative or neutral.

Surprisingly, this simple study design has not yet been used. A number of studies described as “emotional oddball” experiments have been performed, but they have not

manipulated probability in the relevant way. For example, Weinberg et al. (2012) presented sequences of five images, one of which was a different valence category (neutral, positive, or negative) than the other five. However, this stimulus was always the 4th or 5th item of the sequence. In other words, it was an "oddball" in the sense that it was different from the preceding stimuli, but not in the sense that it was low probability or unexpected. At the 4th item of the sequence there was a 50% chance of the oddball and when the oddball was the last item it could be predicted with 100% accuracy. Thus, on average, the oddball had a probability of 75% (see also, e.g., Crites et al., 1995; Delplanque et al., 2006; Kamp et al., 2015). Other studies have compared across emotion conditions within a rare category. For example, Delplanque et al. (2006) report an oddball study with a simple geometric surfaces as the standards and pictures that were either neutral, positive, or negative as the rare targets (see also, e.g., Delplanque, Lavoie, Hot, Silvert, & Sequeira, 2004; Mardaga & Hansenne, 2009). The goal of the present work is to manipulate the actual predictability and expectancy of each emotional category.

To further compare the processing of emotional words to the P300, we also conducted an oddball study in which the gender of names was used as the relevant categories, but which was otherwise perfectly matched to the emotional words oddball. This is a design that has been used in several classic P300 studies (e.g., Fabiani et al., 1986; Kutas et al., 1977).

Finally, after participants completed both oddball studies, they were given a surprise memory recall test (cf. Fabiani et al., 1986). This enabled us to test the relationship of ERPs at encoding to later memory.

Methods

Participants

37 participants were recruited via a departmental paid studies system, flyers around campus, a lab call-back list, and word of mouth. All participants were right-handed, native English speakers with no history of neurological or psychiatric problems (including mood disorders).

Participants were consented in accordance with procedures approved by the Institutional Review Board of Tufts University and paid for their participation. We excluded from analysis any participant with more than 25% of trials rejected for EEG artifact in any condition; 7 participants met this criterion. An additional 5 participants were excluded due to experimenter error or technical problems during data collection and one participant was unable to finish the experiment due to sleepiness. The remaining 24 participants (12 female) were 18 to 32 years of age (*M* = 21.5, *SD* = 3.1). 22 participants identified as non-Hispanic white and 2 identified as black or African-American.

Stimuli

Emotional stimuli

Stimuli for the emotion blocks consisted of 300 negative words and 300 neutral words drawn from the Warriner, Kuperman, and Brysbaert (2013) database of emotional ratings for nearly 14,000 English lemmas. Stimuli were selected via the following process. To obtain candidate negative words, we selected all words from the Warriner et al. (2013) emotion norms with a valence rating of 3 or below (on a 9-point scale with 5 indicating completely neutral valence) and an arousal rating above the median in the database (4.11 on a 9-point scale). To obtain candidate neutral items, we selected all words with a valence rating between 4.5 and 5.5 and an arousal rating below the median. Some items in this database are phrases rather than single words, and these were discarded. Due to previous research suggesting taboo words may be treated specially and may interfere with other emotion effects on the LPC (Fogel et al., 2012), these words were eliminated as well.

We obtained norms for a number of factors: word length (letters), number of morphemes, number of phonemes, number of syllables, part of speech, log word frequency and log contextual diversity (from $SUBTLEX_{US}$: Brysbaert & New, 2009), orthographic neighborhood, phonological neighborhood, mean bigram frequency, bigram frequency by position, lexical decision reaction

times, and naming reaction times were acquired from the English Lexicon Project (elexicon.wustl.edu; Balota et al., 2007). We also obtained concreteness ratings (Brysbaert, Warriner, & Kuperman, 2014), age of acquisition ratings (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), and semantic neighborhood size (Buchanan, Westbury, & Burgess, 2001; Durda & Buchanan, 2006) for each word. A measure of baseline familiarity was obtained by calculating the percentage of participants who reported knowing each word in two large rating studies (Brysbaert et al., 2014; Kuperman et al., 2012). To reduce ambiguity and better match the emotion blocks to the name blocks (see below), we selected only the words that were listed solely as nouns. After eliminating repeated base words (e.g., "terrorism" and "terrorist") and emotionally ambiguous words (e.g., "abortion"), we used Match v2.3 (van Casteren & Davis, 2007) to select a set of words that were maximally distinct on emotional properties and best matched for all other properties.

Stimulus norms and ratings for the resulting set of 300 negative and 300 neutral words are reported in Table 2.2. As can be seen, the emotion conditions did not differ on length, morphemes, phonemes, syllables, orthographic neighborhood, phonological neighborhood, mean bigram frequency, bigram frequency by position, log word frequency, log contextual diversity, age of acquisition, concreteness, semantic neighborhood, lexical decision RT, or naming RT [all $|t|$s < 1.7, all ps > 0.10]. A Mann-Whitney U test showed that the conditions differed on baseline familiarity (percent of raters who knew the word) [U = 48,687.5, p = .034], but practically speaking this difference was very small (99.0% versus 99.9%) and all words had high baseline familiarity.

Because null hypothesis tests cannot provide direct evidence in favor of the null hypothesis, we also calculated Bayes' factor (BF) via the procedures described in Rouder, Speckman, Sun, Morey, and Iverson (2009) and as implemented in JASP (jasp-stats.org; JASP Team, 2017) with the default prior width of .707. Bayes factor is a ratio of the evidence in favor of the alternative versus the evidence in favor of the null, and can provide evidence for or against

either hypothesis (see discussion in Rouder et al., 2009). A BF of 1 indicates evidence is equal for the null and alternative hypothesis, a BF of less than 1 indicates evidence in favor of the null, and a BF of greater than 1 indicates evidence in favor of the alternative. As can be seen in Table 2.2, all BFs were less than 1 for the matched properties.[13]

Property	NEG Mean (SD)	NEU Mean (SD)	*p*	BF_{10}
length (letters)	7.97 (2.3)	7.92 (2.1)	0.770	0.095
morphemes	1.89 (0.8)	1.92 (0.8)	0.639	0.101
phonemes	6.75 (2.1)	6.70 (2.0)	0.762	0.095
syllables	2.7 (1.0)	2.68 (1.0)	0.777	0.095
orthographic neighborhood	1.12 (2.4)	1.03 (2.4)	0.647	0.101
phonological neighborhood	2.65 (5.9)	2.59 (6.1)	0.903	0.092
mean bigram frequency	3885 (1458)	4008 (1466)	0.306	0.152
bigram frequency by position	4041 (1954)	4318 (2197)	0.103	0.333
log word frequency ($SUBTLEX_{US}$)	2.13 (0.5)	2.08 (0.6)	0.230	0.184
log contextual diversity ($SUBTLEX_{US}$)	1.99 (0.5)	1.92 (0.5)	0.115	0.307
% known	0.99 (0.02)	0.99 (0.03)	0.034	--
age of acquisition	10.09 (2.3)	10.28 (2.1)	0.285	0.159
concreteness	3.18 (0.9)	3.29 (0.9)	0.149	0.252
semantic neighborhood	2.89 (9.9)	4.81 (17.8)	0.110	0.325
lexical decision RT	746 (99)	744 (99)	0.740	0.096
lexical decision RT (normalized)	-0.15 (0.3)	-0.14 (0.3)	0.888	0.092
naming RT	696 (80)	697 (77)	0.817	0.093
naming RT (normalized)	-0.15 (0.4)	-0.15 (0.3)	0.897	0.092
valence	2.49 (0.4)	5.10 (0.4)	< .001	> 100,000
arousal	5.30 (0.7)	3.61 (0.3)	< .001	> 100,000
dominance	3.60 (0.7)	5.33 (0.6)	< .001	> 100,000

Table 2.2: *Stimulus norms and ratings for negative and neutral words*. Concreteness rated on a 5-point scale (1 most abstract to 5 most concrete). Valence, arousal, and dominance rated on 9-point scales ranging from 1 most negative, least arousing, least dominate to 9 most positive, most arousing, most dominate. Values labelled "(normalized)" are normalized for each participant's mean and SD (see Balota et al., 2007). *P*-values are from *t*-tests (semantic neighbors: df = 577; morphemes: df = 597; all others: df = 598), except for % known, which is from a Mann-Whitney U test due to significant violations of the assumptions of the *t*-test. See manuscript text for sources of all norms and ratings and additional details.

In addition to being well matched on the properties discussed above, the conditions were well-differentiated on the emotional properties of interest. Negative words had a mean valence

[13] Bayes factor was not calculated for % known because the t-test calculation that serves as the basis of this version of Bayes factor was not appropriate for this property.

near the bottom of the scale (at the 3rd percentile for all words in the Warriner et al. database) and neutral words had a mean valence very close to the mid-point of the scale (47th percentile). In addition, negative words were among the most arousing words in the Warriner et al. norms (88th percentile), while neutral words were among the less arousing words in these norms (27th percentile).

Name stimuli

To obtain an additional measure of the traditional P300 and to assist in interpreting any unexpected effects, we also ran a closely matched oddball study without the emotion factor. Following from several classic studies in the P300 literature (e.g., Fabiani et al., 1986; Kutas et al., 1977) male and female names served as the stimulus categories. The name blocks were identical to the emotion blocks but substituted 300 male names and 300 female names for the negative and neutral words. To create these stimuli, we started with the 600 most common male and female names in the 1990 United States Census.[14] We obtained ratings of familiarity (from 0 – "I don't' know this name" to 3 "I'm very familiar with this name") and gender (1 – "Almost always male" to 5 – "Almost always female") from participants (n = 18-33 for each name) recruited from Amazon Mechanical Turk (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014). We also calculated orthographic neighborhood size (Coltheart's N) and bigram frequency from the $SUBTLEX_{US}$ corpus of words (Brysbaert & New, 2009). We obtained two measures of frequency: the percentage of people with each name in the 1990 US Census (see footnote 2) and corpus frequency from all American books published after 2000 in the Google Ngrams corpus (https://books.google.com/ngrams; Brysbaert, Keuleers, & New, 2011; Michel et al., 2011) We then used Match (van Casteren & Davis, 2007) to choose a set of 300 male and 300 female names.

[14] 1990 is the most recent census from which the U.S. Census Bureau makes this data available. See www.census.gov.

As can be seen in Table 2.3, gender ratings showed that the male and female names were well-differentiated and unambiguous in terms of their perceived gender. The male and female names were matched for length, orthographic neighborhood, bigram frequency, census frequency, familiarity, and gender typicality (distance from the midpoint of the gender ratings). They were similar, but not perfectly matched, for Google Ngram frequency. This may be due to differences in baseline frequency of male and female names in the books that the Google corpus draws from.

Property	Male Mean (SD)	Female Mean (SD)	*p*	BF_{10}
length (letters)	5.71 (1.32)	5.85 (1.43)	0.192	0.209
orthographic neighbors	3.33 (5.35)	3.03 (5.47)	0.502	0.113
mean bigram frequency	2594 (986)	2626 (1008)	0.695	0.098
US Census frequency (1990)	0.15% (0.18%)	0.15% (0.18%)	0.782	0.094
log Google Ngrams Frequency	5.79 (0.48)	5.47 (0.51)	< .001	>10,000
% known	98.9% (2.0%)	98.6 % (2.1%)	0.102	0.336
familiarity	1.42 (0.14)	1.43 (0.15)	0.501	0.114
gender typicality	1.69 (0.19)	1.72 (0.18)	0.134	0.273
gender	1.31 (0.19)	4.72 (0.18)	< .001	>10,000

Table 2.3: *Stimulus norms and ratings for male and female names*. Familiarity on scale from 0 ("I don't know this name") to 2 ("I'm very familiar with this name"). Gender on scale from 1 ("Almost always male") to 5 ("Almost always female") with 3 ("Equally male/female" as the midpoint). *P*-values are from *t*-tests with df = 598. See manuscript text for sources of all norms and ratings and additional details.

Procedure

Stimulus presentation

The emotion experiment consisted of three blocks differing in the proportion of neutral and negative words. One block consisted 40 negative words and 160 neutral (20%/80%); another 100 negative and 100 neutral (50%/50%); and the last 160 negative words and 40 neutral (80%/20%). Within each block, the order of the items was randomly determined for each subject individually. No word was repeated across blocks and which block a particular word appeared in was counterbalanced across subjects. The control experiment consisted of three blocks identical to the emotion blocks but with male and female names substituting for negative and neutral words.

The order of the experiments and the order of the proportion blocks within each experiment were counterbalanced across participants.

Stimuli were presented using PsychoPy v1.83.04 (Peirce, 2007, 2009). Each trial began with "+" (1000 ms, ISI: 400-600ms). The word was then presented for 1300 ms with an ISI varying randomly between 600 and 800 ms. Jittered ISIs were used to smooth effects of overlap and eliminate synchronized oscillations. Participants were asked to blink on the fixation cross and not during word presentation. Every 40 trials participants saw a screen indicating they could take a brief, self-paced break. Between blocks there was a longer break where the experimenter engaged with the participant.

Task

The participants' task was simply to press a key on a standard computer keyboard with one hand if the word was negative and a different key with the other hand if it was neutral. For the name blocks, the participants pressed keys to indicate whether the name was male or female. The hand-category mapping for each experiment was counterbalanced across participants. Participants were not told to emphasize speed, but were told to respond while the word was on the screen (i.e., within 1300 ms). Only responses occurring between 100 and 1840 ms after word onset were recorded (this was to prevent overlapping EEG triggers for stimuli and responses, which can cause the BioSemi system to pause recording).

ERP acquisition and processing

Data was collected using a BioSemi ActiveTwo EEG system and ActiView v7.05 EEG acquisition software (http://www.biosemi.com/). The EEG was recorded from 32 Ag/AgCl electrodes in an elastic cap placed according to the international 10-20 system. In addition, electrodes below the left eye and beside the right eye were recorded to monitor for blinks and eye movements, and electrodes on each mastoid were recorded to serve as the reference. The EEG

signal was amplified, digitally filtered online with a low pass 5th order sinc response filter with a half-amplitude cutoff at 102.4 Hz, and continuously sampled at 512Hz.

Memory test

After the ERP portion of the experiment and before leaving the lab, participants completed a recognition memory test that included all items in the 80% probability condition for each emotion and name condition (160 words each; 640 in total) along with 80 words in each emotion and name condition that had not been previously presented. Thus, there were 960 total words/names tested and these were presented in a random order. For each word, participants indicated whether they thought it was "old" (it had been presented in the ERP session) or "new" (it had not been presented in the ERP session) as well indicating their confidence in their judgment as "guessing", "somewhat sure", or "sure".

Data processing and analysis

EEG and ERP data processing was conducted in EEGLAB v13.5.4 (sccn.ucsd.edu/eeglab; Delorme & Makeig, 2004) and ERPLAB v5.0.0.0 (erpinfo.org/erplab; Lopez-Calderon & Luck, 2014) running in MATLAB 2015a.

The EEG was first referenced to the average of the two mastoid electrodes. For each segment of continuous EEG, we removed the DC offset by subtracting the average voltage of the entire segment, then applied a high-pass 2nd-order Butterworth infinite impulse response filter with a half-amplitude cut-off of 0.1 Hz (Kappenman & Luck, 2010; Tanner, Morgan-Short, & Luck, 2015). Prior to averaging, blinks, eye movements, and other artifacts were detected via artifact detection algorithms implemented in ERPLAB. The particular algorithms used and their parameters were tailored to each subject via visual inspection of the data, but were consistent across all conditions within each subject. Averages were then calculated from trials free of artifact for epochs 200 ms prior to stimulus onset to 1100 ms after stimulus onset for each condition of the experiment. The

200 ms period prior to stimulus onset served as the baseline for all data visualization and amplitude measurements.

ERP analysis

The factors in the averaged ERP analysis were the manipulated factors: Emotion (neutral, negative) and Probability (20%, 50%, 80%) for the emotional words study; Gender (male, female) and Probability (20%, 50%, 80%) for the names study. For analyses examining the effects of memory the factors were Emotion (neutral, negative) and Memory Status (remembered, forgotten) and Gender (male, female) and Memory Status (remembered, forgotten), respectively.

Data were analyzed at all time points from 300 to 1000 ms at all electrodes via the factorial ANOVA cluster mass test described in Chapter 3. The threshold for cluster inclusion was $p \leq .05$ and electrodes within approximately 7.5 cm of each other (assuming a head circumference of 56 cm) were considered neighbors. 100,000 permutations were performed for each test. Statistical analysis was conducted on data that was first low-pass filtered at 30 Hz (half amplitude cut-off, 2nd-order Butterworth IIR filter) and downsampled to 128 Hz (see Groppe, Urbach, & Kutas, 2011a; Luck, 2014, Ch. 13).

Results

Behavioral results

Names study

Behavioral results for the names study are shown in Figure 2.1.

Reaction times were slower to the 20% and 50% conditions than the 80% condition [Tukey's test: $t(23)$s > 6.4, ps < .001], but the 20% and 50% conditions did not differ [Tukey's test: $t(23) = 0.01$, $p > .999$]. Participants were also slower to identify female names than male names [$F(1, 23) = 4.32$, $p = .049$]. There was no interaction of Probability and Gender [$F(2, 46) = 0.12$, $p = .814$].

Categorization was less accurate for the 20% condition than the 50% or 80% conditions [Tukey's test: $t(23)$s > 2.8, ps < .02], which did not differ [Tukey's test: $t(23) = 1.26$, $p = .424$]. There was no effect of Gender on accuracy [$F(1, 23) = 0.17$, $p = .683$] and Gender did not interact with Probability [$F(2, 46) = 0.05$, $p = .874$]. A signal detection analysis showed that participants ability to discriminate between the categories did not differ between blocks [$F(1, 23) = 0.63$, $p = .528$], but the criterion did shift [$F(2, 46) = 36.96$, $p < .001$]: participants were biased toward whichever category was more common in a given block [one sample $t(23)$s > 4.3, ps < .001] with no bias for the 50/50 blocks [one sample $t(23) = 0.93$, $p = .362$].

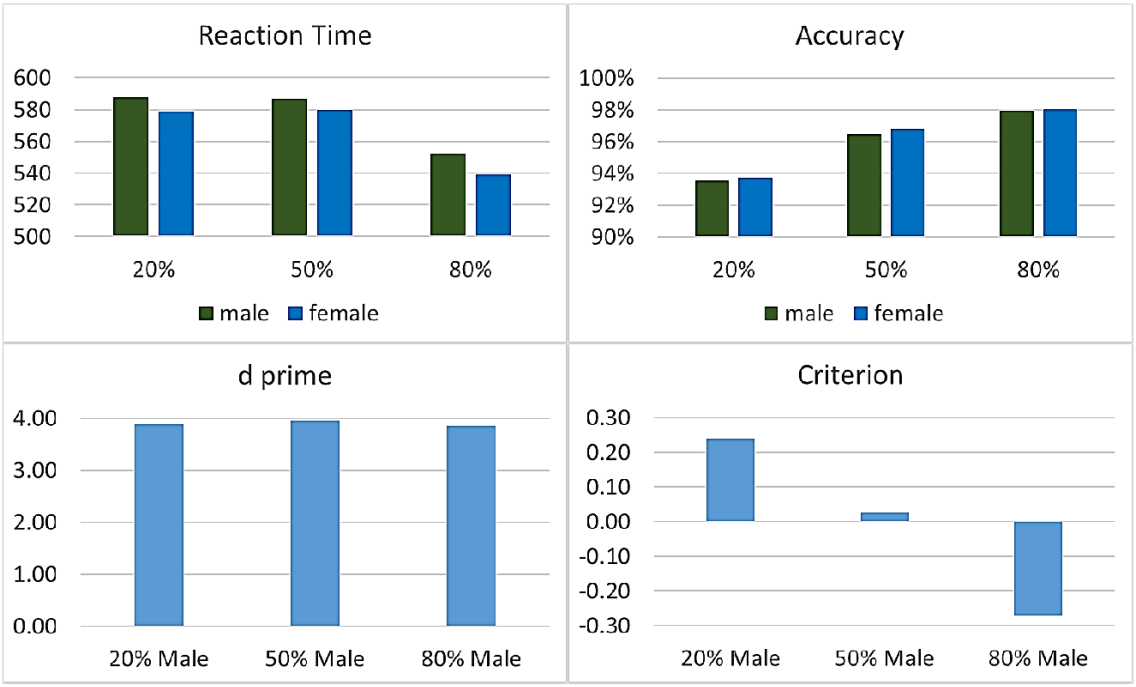

Figure 2.1: *Behavioral results for name-gender oddball study*. The top row shows reaction time (in ms) and accuracy for the gender categorization task. The bottom row shows signal detection measures for the same task. For the criterion measure, positive values indicate a tendency to respond "female" and negative values indicate a tendency to respond "male".

Emotion study

Behavioral results for the emotional words study are shown in Figure 2.2.

Reaction times were faster for negative words than neutral words [$F(1, 23) = 7.95$, $p = .010$]. Reaction times were slower for the 20% condition versus the 50% and 80% conditions

[Tukey's test: $t(23)$s > 2.9, ps < .02], which did not differ from each other [Tukey's test: $t(23)$ = 1.89, p = .154]. There was no interaction of Emotion and Probability for reaction times [$F(2, 46)$ = 1.09, p = 0.322].

Categorization was more accurate for higher probability words [Tukey's test: all $t(23)$s > 2.9, ps < .02]. There was no main effect of Emotion or Probability x Emotion interaction on accuracy [Fs < 1.37, ps > 0.25]. A signal detection analysis revealed that participants' ability to discriminate neutral and negative words did not differ between the three blocks [$F(2, 46)$ = 0.42, p = 0.643]. However, bias did differ between blocks [$F(2, 46)$ = 44.59, p < .001] with participants being more likely to respond with the more common category in the 80/20 blocks [one sample $t(23)$s > 4.2, ps < .001] and more likely to respond with neutral in the 50/50 block (although this bias did not reach significance in a one-sample t-test [one sample $t(23)$ = 1.59, p = 0.127]).

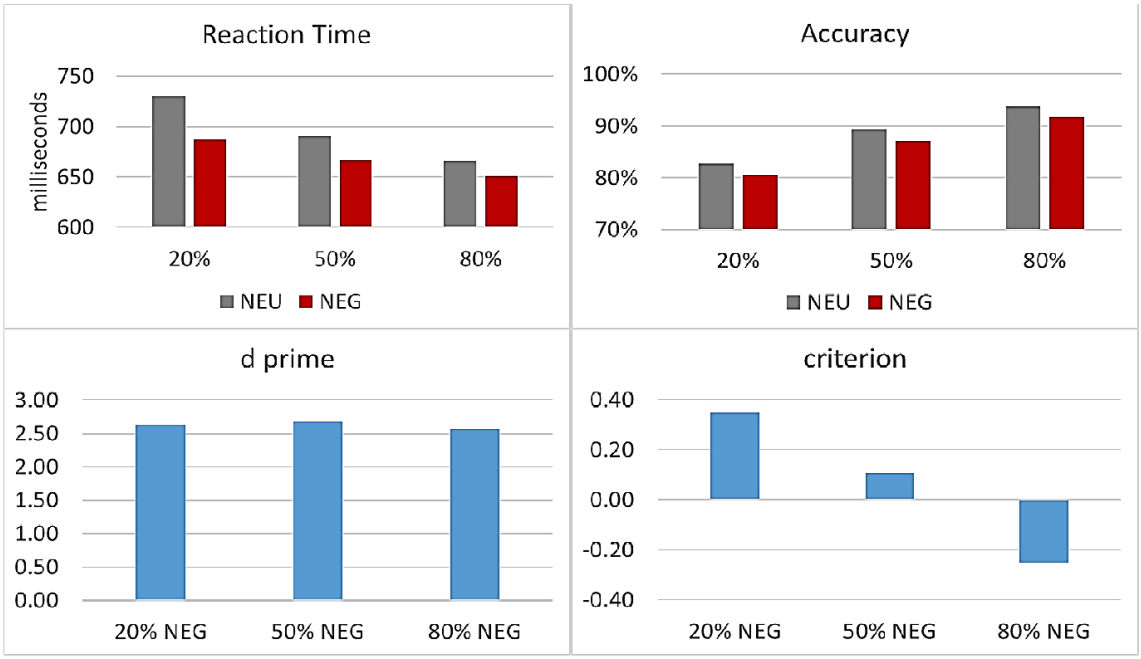

Figure 2.2: *Behavioral results for emotional words oddball study*. The top row shows reaction time (in ms) and accuracy for the valence categorization task. The bottom row shows signal detection measures for the same task. For the criterion measure, positive values indicate a tendency to respond "neutral" and negative values indicate a tendency to respond "negative".

Percent Correct: ERP stimuli

	ALL	guessing	somewhat sure	sure
NEG	87%	52%	74%	98%
NEU	76%	40%	59%	95%
Male	69%	47%	51%	91%
Female	65%	43%	53%	86%

ERP Stimuli by Confidence Rating

	ALL	guessing	somewhat sure	sure
NEG	160	15.3	44.4	100.4
NEU	160	21.7	52.8	85.5
Male	160	28.3	57.0	74.8
Female	160	29.6	57.5	72.9

Percent Correct: Foils

	ALL	guessing	somewhat sure	sure
NEG	52%	56%	57%	45%
NEU	66%	61%	71%	64%
Male	67%	55%	71%	64%
Female	63%	49%	66%	61%

Foils Confidence Rating

	ALL	guessing	somewhat sure	sure
NEG	80	16.7	38.4	24.9
NEU	80	19.1	40.0	20.9
Male	80	19.9	37.0	23.1
Female	80	17.5	38.1	24.4

Signal Detection: d'

	ALL	guessing	somewhat sure	sure
NEG	1.33	0.21	0.91	2.08
NEU	1.26	0.06	0.89	2.33
Male	1.02	0.32	0.61	2.05
Female	0.78	-0.07	0.53	1.67

Signal Detection: criterion (bias)

	ALL	guessing	somewhat sure	sure
NEG	-0.62	0.01	-0.32	-1.17
NEU	-0.17	0.18	0.17	-0.74
Male	-0.03	0.11	0.26	-0.54
Female	-0.02	0.10	0.21	-0.47

Table 2.4: *Memory test results.* Accuracy reflects the percentage of items correctly identified as "old" for the ERP stimuli and correctly identified as "new" for the foils. For the signal detection criterion measures, negative values represent a bias toward to responding "old" and positive numbers reflect a bias toward responding "new".

ERP results

Names study

Gender x Probability interaction

There was no apparent gender x probability interaction effect and no cluster reached significance in the statistical analyses (all clusters $p > .13$).

Gender main effect

Two small effects of Gender were apparent. First, a positivity largest at left posterior sites was larger to female names beginning before 400 ms, peaking at around 500 ms, and extending throughout the epoch. Second, a right-frontal positivity was larger to female names from approximately 800 ms to the end of the epoch. Both effects were significant in the cluster mass analysis (see Figure 2.4). Because gender per se was not of interest in the present work and these effects did not interact with Probability, they will not be discussed further.

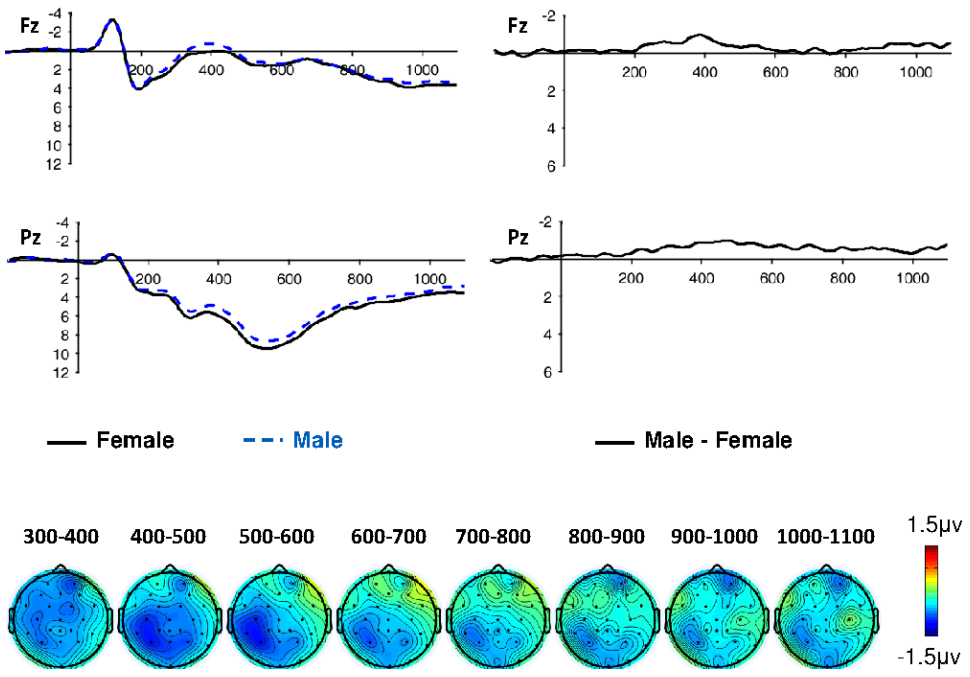

Figure 2.3: *Main effect of Gender*. Waveforms on the left show ERPs for each condition. Waveforms on the right show the male – female difference wave. All waveforms are low pass filtered at 15 Hz for viewing purposes. Scalp maps show the male – female voltage averaged across 100 ms time windows.

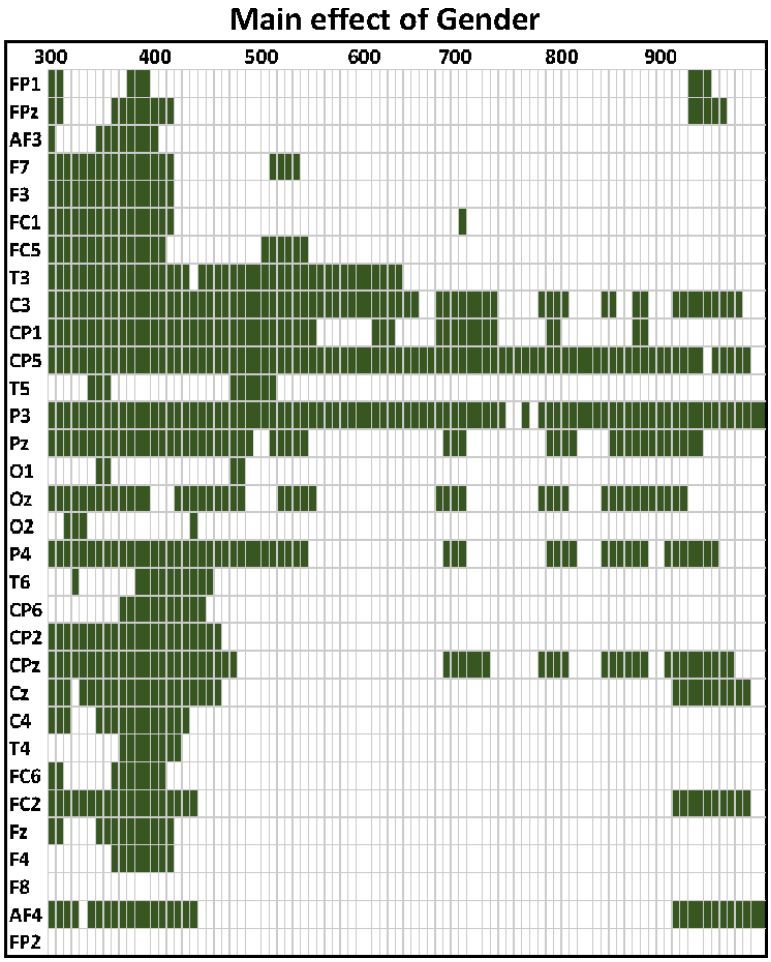

Figure 2.4: *Cluster mass statistical analysis for the main effect of Gender*. Time point-electrode combinations included in a significant cluster ($p < .05$) are highlighted in dark green.

Probability main effect

Main effects of probability are shown in Figure 2.5. As expected, a posterior positivity that can be identified as the P300 was larger to the 20% condition than the 50% condition and larger to the 50% condition than the 80% condition. This centroparietal effect (peaking at Pz) began around 500 ms, peaked just after 600 ms, and extended to around 900 ms. After and partially overlapping with this posterior effect (700 ms to the end of the epoch) was a frontally distributed effect (largest at Fz) showing the opposite pattern (i.e., the 80% condition was more positive than the 50% and the 50% was more positive than the 20%). Both effects were included in significant clusters in the cluster mass analysis (see Figure 2.6).

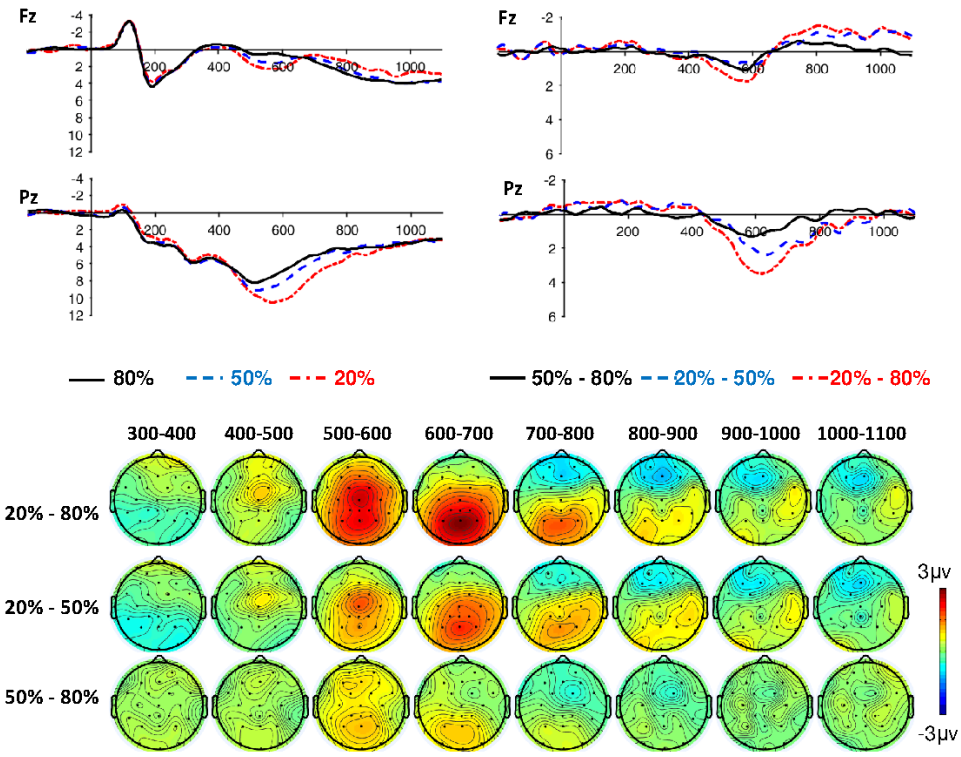

Figure 2.5: *Main effect of Probability for name-gender oddball study*. Waveforms on the left show ERPs for each condition. Waveforms on the right show difference waves for each pairwise comparison. All waveforms are low pass filtered at 15 Hz for viewing purposes. Scalp maps show the voltage difference for each comparison averaged across 100 ms time windows.

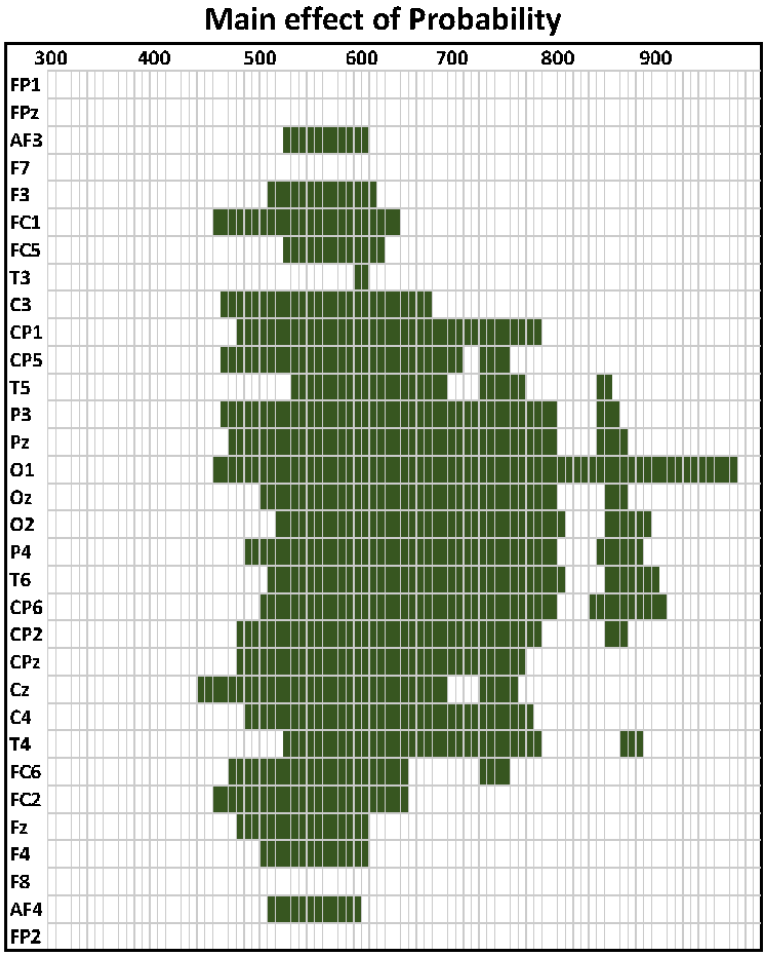

Figure 2.6: *Cluster mass statistical analysis for the main effect of Probability for the name-gender oddball study*. Time point-electrode combinations included in a significant cluster ($p < .05$) are highlighted in dark green.

Memory effects

To examine the relationship between online processing and later memory, we compared trials for which participants indicated that they were "sure" they had seen previously with those that they (incorrectly) indicated that had not seen previously and/or those on which they indicated they were guessing (either of which suggested no memory) within each condition.

There was no interaction of Memory Status and Gender at any time point or electrode (all clusters $p > .76$). Main effects of Memory Status are shown in Figure 2.7. Names that were remembered elicited a larger frontal positivity than names that were forgotten beginning around 500 ms and extending to the end of the epoch. Remembered names also elicited a larger posterior

positivity centered around Pz that extended from around 600 ms to the end of the epoch. Both of these effects were included in significant clusters in the cluster mass analysis (see Figure 2.8). The waveforms also suggested an increased negativity to remembered words between 200 and 600 ms at left temporal and parietal sites, but this effect was not significant in the cluster mass analysis (p = .273).

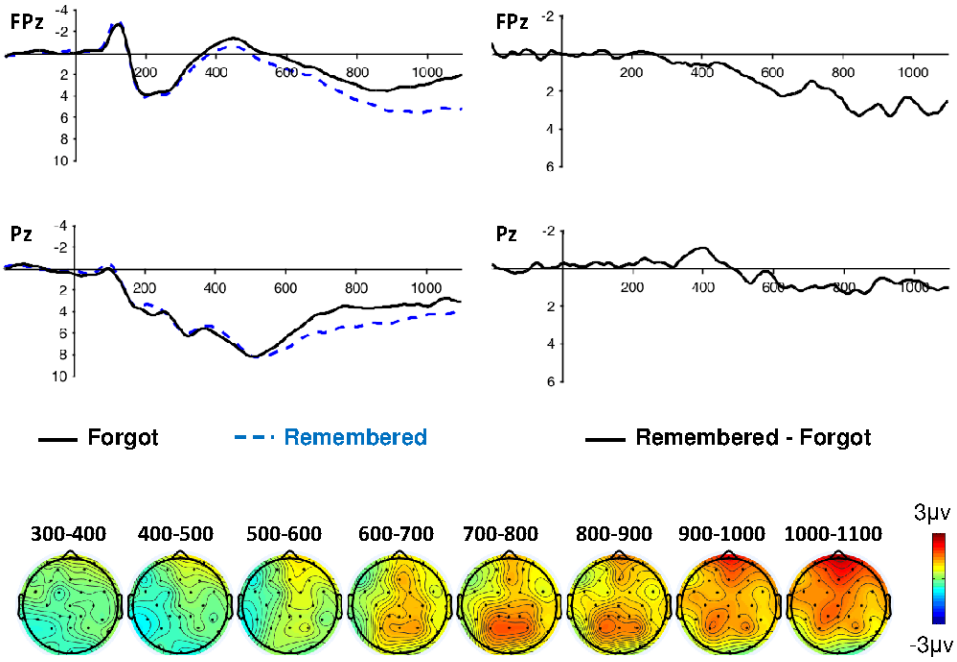

Figure 2.7: *Main effect of Memory Status for name-gender oddball study*. Waveforms on the left show ERPs for each condition. Waveforms on the right show remembered – forgotten difference waves. All waveforms are low pass filtered at 15 Hz for viewing purposes. Scalp maps show the remembered – forgotten voltage difference averaged across 100 ms time windows.

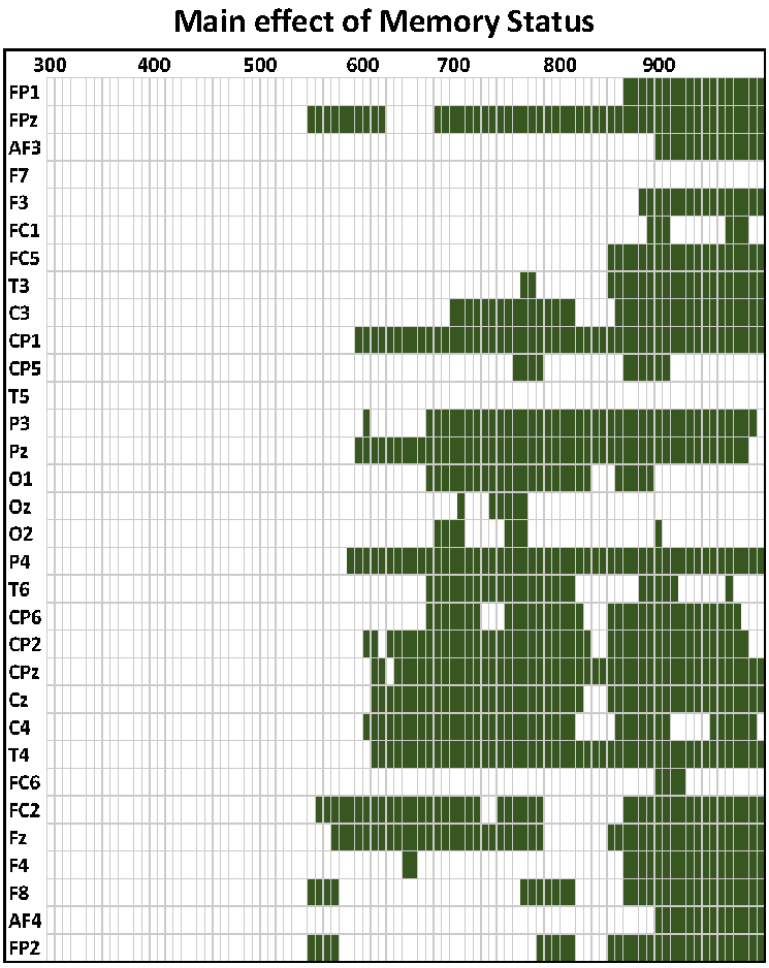

Figure 2.8: *Cluster mass statistical analysis for the main effect of Memory Status for the name-gender oddball study*. Time point-electrode combinations included in a significant cluster ($p < .05$) are highlighted in dark green.

Emotion study

Emotion x Probability interaction

There was no apparent emotion x probability interaction effect and no cluster reached significance in the statistical analyses. This lack of interaction seems unlikely to be attributable to insufficient power: $p > .63$ for all clusters and only a few isolated data points reached significance with no multiple comparisons correction at all. For example, the average *F*-value at CPz from 500 to 800 ms (a spatiotemporal ROI where the main effect of both probability and emotion were large—see below) was 1.90 which corresponds to an uncorrected parametric *p*-value of 0.8.

Emotion main effect

Main effects of emotion are shown in Figure 2.9. As expected, negative words elicited a larger posterior positivity that can be identified as the LPC. This effect began by 300 ms, peaked just before 600 ms, and continued until the end of the epoch. The effect was centroparietal (centered around CPz) at its peak, but more frontal at the beginning and end. This may be in part due overlap with an earlier effect, peaking at around 400 ms, at the most frontal electrodes. This effect coincided with a negative deflection in the waveform, but it's not clear whether it should be regarded as a reduced negativity or increased positivity for negative words. Both effects were included in significant clusters in the cluster mass analysis (see Figure 2.10).

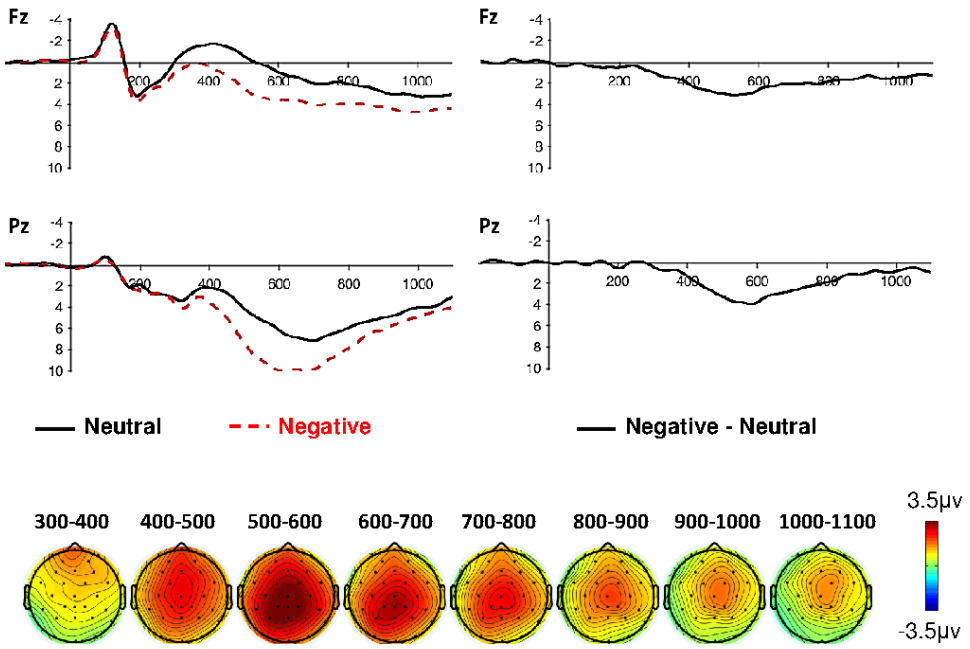

Figure 2.9: *Main effect of Emotion*. Waveforms on the left show ERPs for each condition. Waveforms on the right show the negative - neutral difference wave. All waveforms are low pass filtered at 15 Hz for viewing purposes. Scalp maps show the negative - neutral voltage averaged across 100 ms time windows.

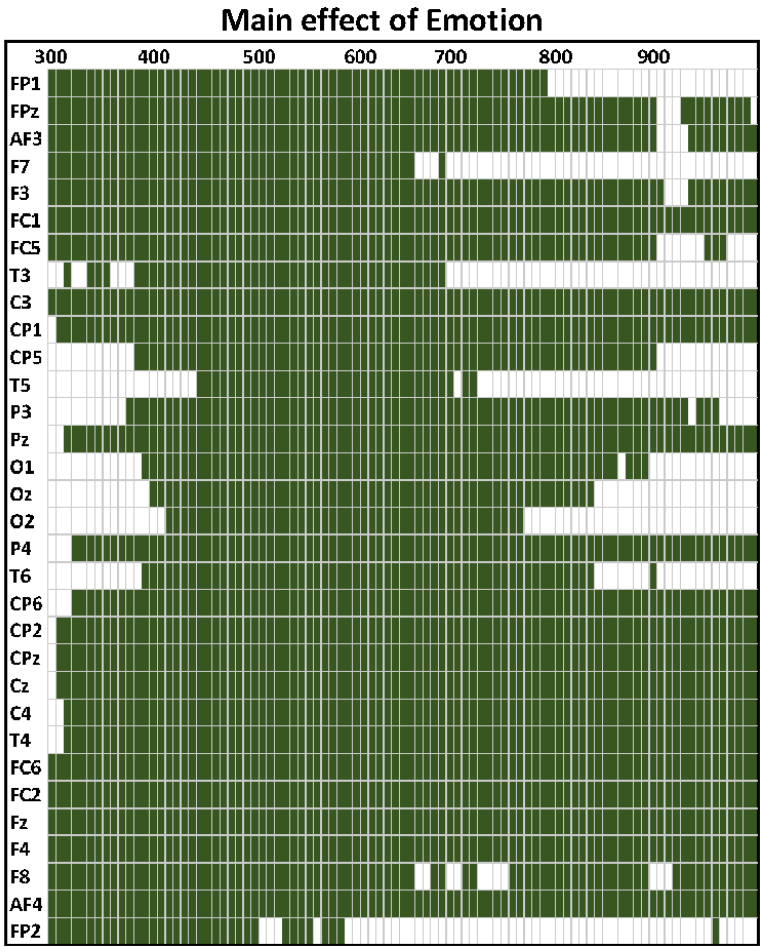

Figure 2.10: *Cluster mass statistical analysis for the main effect of Emotion.* Time point-electrode combinations included in a significant cluster ($p < .05$) are highlighted in dark green.

Probability main effect

Main effects of Probability are shown in Figure 2.11. As expected, there was an effect of the Probability manipulation on a centroparietal positivity. However, this effect was driven primarily by the 20% condition, which was larger than the 50% and 80% conditions while the 50% and 80% conditions showed a smaller difference. This pattern was confirmed by the cluster mass statistical analysis, where there were no clusters that reached significance for the 50% vs. 80% contrast. The positivity to words in the 20% condition began to diverge around 500 ms, peaked between 700 and 800 ms and extended to the end of the epoch. Similar to the emotion main effect, there

was also an earlier effect of probability at frontal sites from around 400 to 800 ms and peaking at around 600 ms.

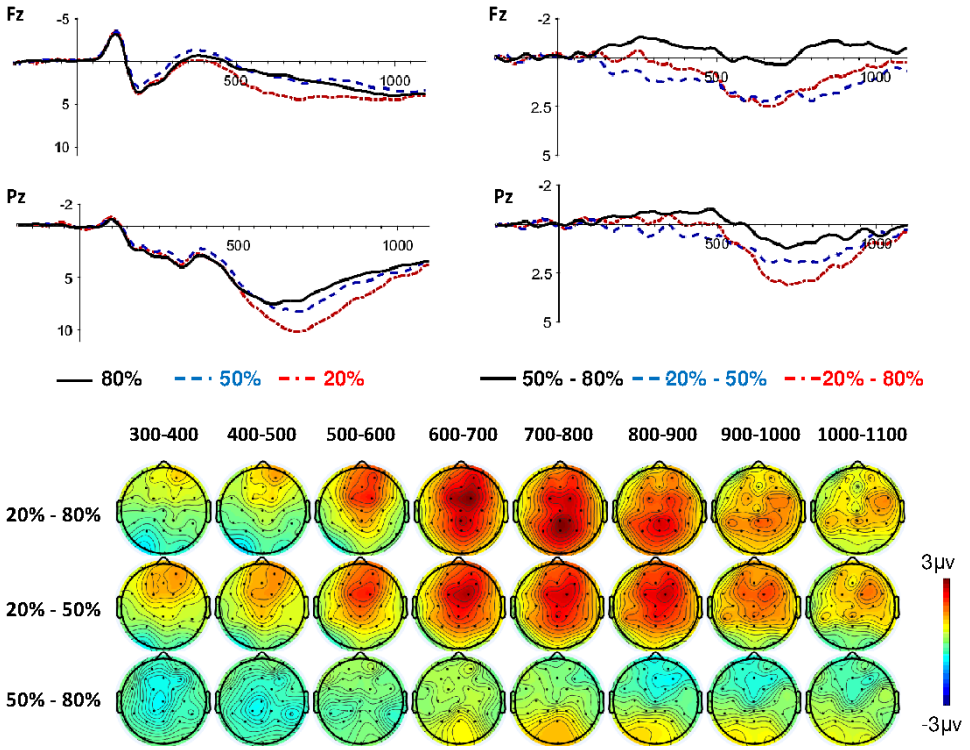

Figure 2.11: *Main effect of Probability for emotional words oddball study*. Waveforms on the left show ERPs for each condition. Waveforms on the right show difference waves for each pairwise comparison. All waveforms are low pass filtered at 15 Hz for viewing purposes. Scalp maps show the voltage difference for each comparison averaged across 100 ms time windows.

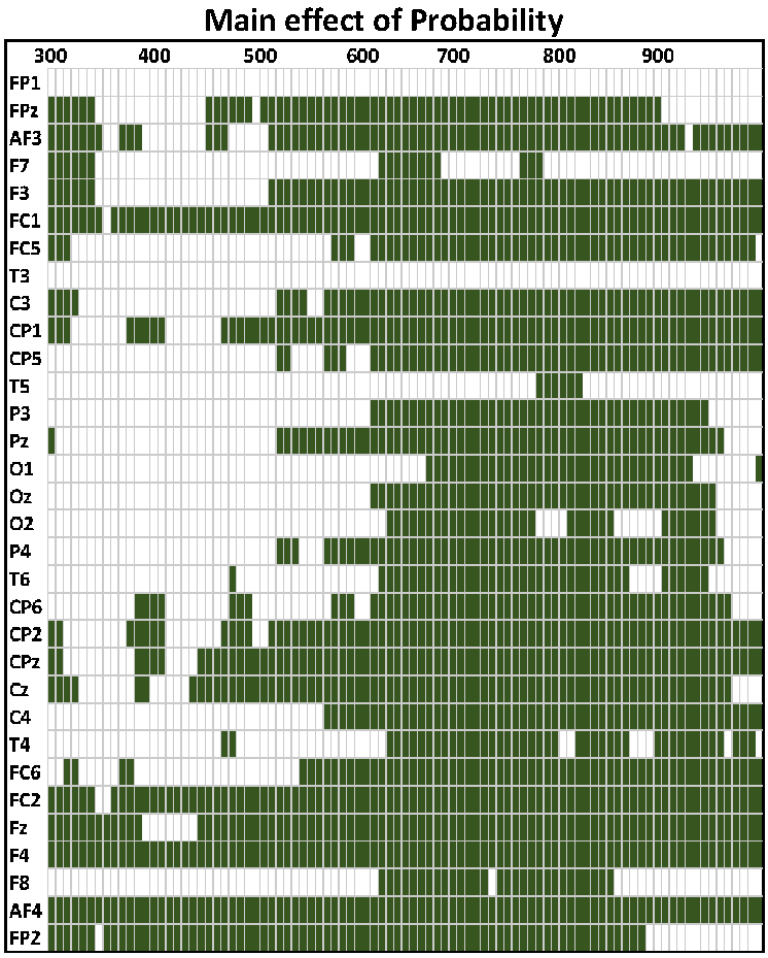

Figure 2.12: *Cluster mass statistical analysis for the main effect of Probability for the emotional words oddball study*. Time point-electrode combinations included in a significant cluster ($p < .05$) are highlighted in dark green.

Memory effects

To examine the relationship between online processing and later memory, we compared trials for which participants indicated that they were "sure" they had seen previously with those that they (incorrectly) indicated that had not seen previously and/or those on which they indicated they were guessing (either of which suggested no memory) within each condition.

There were no interaction effects between Emotion and Memory Status (all clusters $p > .66$). Main effects of Memory Status are shown in Figure 2.13. Words that were remembered elicited a larger right-frontal positivity (largest at AF4) beginning by 500 ms and extending to the end of the epoch. Starting around 600 ms, there was also a somewhat left-lateralized posterior

positivity. The cluster mass analysis showed significant clusters including both the frontal and posterior portions of the effect (see Figure 2.14). Similar to the Names study, there was also a left posterior negativity to remembered words peaking around 400 ms, but this effect was not significant in the statistical analysis (p = .157).[15]

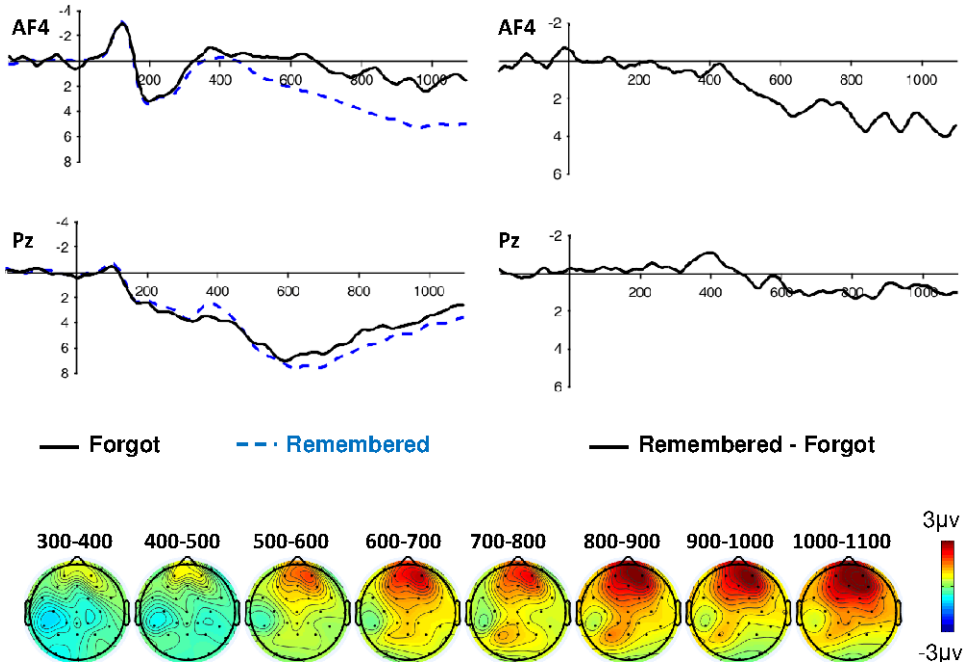

Figure 2.13: *Main effect of Memory Status for emotional words oddball study*. Waveforms on the left show ERPs for each condition. Waveforms on the right show remembered – forgotten difference waves. All waveforms are low pass filtered at 15 Hz for viewing purposes. Scalp maps show the remembered – forgotten voltage difference averaged across 100 ms time windows.

[15] The fact that this effect was observed independently in both studies suggests that its lack of significance could be due to insufficient power. However, because this effect was not predicted and is not directly relevant to the current research questions, it will not be discussed further.

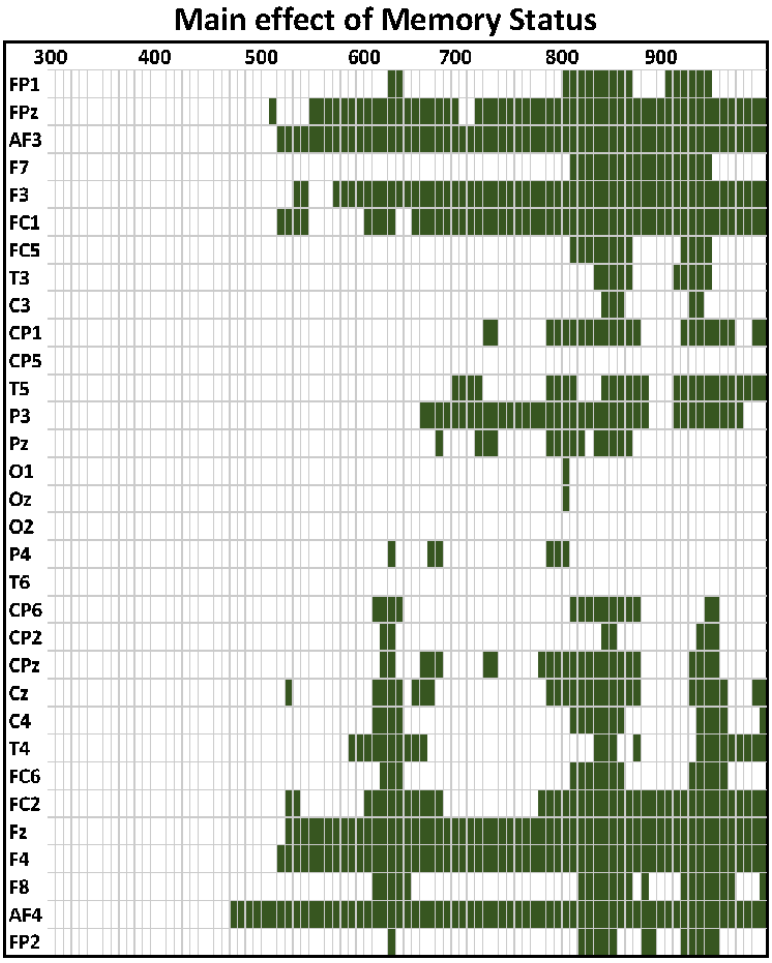

Figure 2.14: *Cluster mass statistical analysis for the main effect of Memory Status for the emotional words oddball study*. Time point-electrode combinations included in a significant cluster ($p < .05$) are highlighted in dark green.

Summary

Importantly, our results replicated the key findings of previous studies examining probability and memory effects using a name-gender oddball design (Fabiani et al., 1986; Kutas et al., 1977): a probability effect on the P300 component and a posterior positivity that was larger to remembered names than forgotten names. In addition, we observed a later frontal effect of probability with a pattern opposite of the P300 as well as a long-lasting frontal positivity to remembered names.

In the emotional words study, we showed the expected LPC effect to negative words versus neutral words. There was also an effect of probability on a posterior positivity as expected,

but this effect was later than that observed in the names study and restricted to the 20% condition (at least in terms of statistical significance). Comparing the effects of Emotion and Probability within the emotional words study, the effect of probability began and peaked somewhat later. However, both effects showed a similar centroparietal distribution, and both Probability and Emotion also elicited an earlier frontal positivity.

Both studies showed frontal and posterior positivities to remembered words versus forgotten words. For the names, the posterior portion was larger than the frontal positivity from around 600-900 ms, whereas for the emotional words study, the frontal positivity began earlier and was larger than the posterior effect until the end of the epoch.

Discussion

Although a number of potentially interesting effects emerged in the present study, for present purposes the discussion will focus on results relevant to the main questions motivating this work: 1) what is the relationship between the P300 and LPC? and 2) what is the functional role of the LPC?

What is the relationship between the P300 and LPC?

Both emotion and probability elicited similar effects, but there were also differences. In terms of timing, the probability effect began and peaked later than the emotion effect. However, this may be understood based on what we know of the P300. In categorical oddball studies the latency of the P300 varies with the time it takes to perform the relevant categorization (Kutas et al., 1977; Magliero, Bashore, Coles, & Donchin, 1984). In other words, the fact that a stimulus is from a rarer category cannot affect cognitive processes until the perceiver is able to determine its category membership. If emotional factors such as valence and arousal modulate the same process independently (i.e., not simply via category membership), they would not need to wait until categorization. So it is possible that early perception of the emotional features of the words began to modulate the process represented by the P300/LPC starting around 300 ms and

category membership later independently modulated the same process once participants had finished the more involved cognitive task of interpreting these emotional features in terms of binary categories.

Both the effect of emotion and the effect of probability were centered around centro-parietal electrodes and had very similar overall scalp distributions. Given the diffuse nature of ERP scalp distributions and their ambiguous relationship to underlying neural generators, this provides only weak evidence that the effects represent the same process or underlying neural sources. However, the argument is somewhat bolstered by the changes in scalp distribution over time: both effects first emerged at frontal sites before peaking at parietal sites at later time points. Notably, an earlier peak at frontal sites is a commonly observed pattern in P300 studies (Polich, 2007; Polich et al., 1997).[16]

There was no interaction of emotion and probability on the late positivity. Although an interaction would have provided the strongest evidence that these factors were modulating the same component, the lack of an interaction does not rule out the possibility that emotion and probability affected the same functional process independently. Indeed, among the factors known to modulate the P300, some have an additive effect. For example, R. Johnson Jr.'s (1986) model of the factors affecting P300 amplitude proposes that "stimulus value" has an additive relationship to stimulus probability. This is based on studies that manipulated value via the monetary reward associated with a stimulus, but as R. Johnson, Jr. (1986) notes "presumably there are many

[16] I take no strong position on the interpretation of this early frontal positivity, but one possibility is that it reflects the P3a component, which is related to but distinct from the P300 (or P3b). In some theoretical frameworks, the P3a and P3b are separable components resulting from the activation of the same norepinephrine (NE) neurotransmitter system (e.g., Nieuwenhuis et al., 2005; Polich, 2007). The P3a represents effects of NE in frontal regions controlling orienting and attention, and the P3b represents effects of NE at temporal and parietal regions underlying decision making and/or memory processes. Under this interpretation, the pattern observed for both probability and emotion suggests both are related to the activation of the locus coeruleus-norepinephrine system, in line with Nieuwenhuis et al. (2005).

instances in which the significance of events is determined by experience" (p. 373)—i.e., emotional stimuli.

Although the valence of the word did not interact with the probability effect, the overall emotional context may have interacted with the effects of probability: in contrast to the names study and many previous P300 studies, the positivity observed in the emotional words study was not graded with probability. Instead, only the 20% condition significantly differed from the other conditions. It is unclear whether this difference reflects anything meaningful or is simply a matter of power due to greater trial to trial variability in the emotion study. The general pattern looks very similar in the emotion and names studies (see Figures 2.5 and 2.11), but is more spread out in time in the emotion study. One possibility is that the emotional context "raises the stakes" such that words must be lower probability to elicit an increased late positivity. As discussed earlier in this chapter, this kind of context effect has been observed previously with taboo words (Fogel et al., 2012) and may also account for some effects of self-relevance on the LPC (Fields & Kuperberg, 2016).

Is the LPC related to memory?

Although the largest ERP difference between remembered and forgotten words was on a frontally distributed positivity, remembered words also elicited what appears to be a distinct parietal positivity. The "Dm effect"—the difference, at encoding, between remembered and forgotten stimuli—varies in timing, scalp distribution, and polarity across studies based on a number of different factors including stimulus type, task, and participant strategy (reviewed in A. D. Wagner et al., 1999; Wilding & Ranganath, 2012). Although the reasons for specific differences are not always known, the reason for the general variability is easily understood. Clearly processing within ~1s of stimulus onset is not the only determinant of whether a stimulus will be remembered at a later time. In addition, many different cognitive processes will be correlated with

later memory and which process most strongly predicts later memory will vary depending aspects of the stimulus, context, and perceiver's goals/strategy.

The context updating account that originally motivated examination of the relationship between P300 and memory argued that because the P300 reflected processing in relation to the task-relevant schema of the environment and related memory structures, items with a larger P300 should be better remembered. It did not argue that the P300 represented the only or even the main process of encoding and storage in long-term memory. Thus, under some circumstances, other memory processes may play a larger role in which items are remembered and the relationship of the P300 to memory may be weakened or disappear. For example, from the beginning of research on this question, studies showed that the P300 most strongly predicted later memory when participants utilized a rote, rather than elaborative, memorization strategy; when an elaborative strategy was used, memory was associated with an increased frontal positivity (Fabiani et al., 1990; Karis et al., 1984).

More generally, both frontal and posterior positivities are very common manifestations of the Dm effect in ERP studies, and a number of studies have observed both as was the case in the results presented here. It is not entirely clear why the frontal positivity effect was stronger for emotional words and the posterior positivity was stronger for names. Given the findings discussed above, one could imagine that the emotional words study led to more spontaneous elaboration than the names or perhaps the words were processed more deeply because the task was harder (Otten & Rugg, 2001), but these explanations are speculative.

In any case, we did observe a larger posterior positivity to remembered words as predicted. This provides some evidence that activity associated with the LPC is related to memory. Notably, this memory effect began later than the main effect of emotion on the LPC. Both the P300 (Nieuwenhuis et al., 2005; Polich, 2007) and LPC (Delplanque et al., 2006; Foti et al., 2009; Gable et al., 2015; Hajcak et al., 2012; MacNamara et al., 2009; Matsuda & Nittono, 2015) are complex

components that are generally thought to be generated by multiple sources and have dissociable sub-components (see discussion earlier in this chapter).[17] That is, of course, consistent with the changing scalp distribution over time observed in the present study for both effects of probability and emotion. If the positivity associated with later memory is part of the LPC, this suggests that only later portions of the LPC were associated with memory in the present work. This is explored in greater detail below (see also Footnote 16).

What is the function of the LPC?

The present study was intended to investigate a functional theory of the LPC. I proposed that the LPC may in fact be another name for the P300 or at least that these components may index significantly overlapping cognitive functions. More specifically, this theory suggests that the LPC reflects a process of updating our schema of the current environment in response to stimuli that are particularly informative or particularly important (motivationally relevant). To test this theory, we investigated the ERP effects of the factors most consistently associated with each component: stimulus probability (P300) and emotional valence/arousal (LPC). Although probability and emotion did not interact, they elicited similar effects. The biggest difference between the effects was the time course (particularly onset), but this difference makes sense given what we know about the timing of the P300. As predicted, a posterior positivity overlapping in space and time with the LPC was associated with later memory for a word, although this effect only overlapped with the later portion of the LPC. Taken together, the results can be seen as consistent with the P300-context updating theory of the LPC. However, this is not the same as strong evidence in favor of the theory, so it is important to consider alternative explanations.

[17] Note that this does not mean that these components are not, in some sense, a unitary phenomenon. For example, as discussed earlier in this chapter, one theory of the P300/LPC argues that it is generated in cortical areas stimulated by the locus coeruleus-norepinephrine system (Nieuwenhuis et al., 2005; Polich, 2007; see also Footnote 16). In response to motivationally relevant stimuli, NE is released to many cortical areas subserving different functions and the P300 is the result of the summated potential of these various regions. Under this understanding, the P300 is a coherent phenomenon, but also has subcomponents that may have somewhat different time courses and be more or less prominent in different situations.

Attention is probably the psychological construct most consistently associated with the LPC. As noted in earlier in this chapter, attention is a broad concept rather than a specific cognitive process, and most researchers have not been specific about what attention-related process in particularly the LPC might reflect. An exception is Gable et al. (2015), who argue that the LPC reflects "motivated [attentional] engagement with visual stimuli" (p. 53) and is generated by activity in occipital-parietal visual areas as a result of feedback from subcortical regions (e.g., the amygdala in response to emotional stimuli) and/or frontal regions (e.g., in the case of top-down effects of task relevance). This explanation is supported by results showing that the time-course of the LPC patterns with the expected time-course of attentional engagement given manipulations of stimulus duration and task-requirements (Gable et al., 2015), as well as by combined ERP-fMRI studies that have correlated the LPC with activity in visual areas (among many other neural regions; Y. L. Liu, Huang, McGinnis-Deweese, Keil, & Ding, 2012; Sabatinelli, Keil, Frank, & Lang, 2013; Sabatinelli, Lang, Keil, & Bradley, 2007). Taken together, this is probably the most detailed and well supported alternative to the P300-context updating theory of the LPC elaborated here.

However, the studies supporting this account were conducted with emotional pictures and the visual attention explanation seems less convincing for the LPC to words. While emotional pictures may draw sustained visual attention, there is little further information to extract from a word via visual examination after it has been recognized. Thus, visual engagement seems unlikely to explain an effect evoked by words that occurs wholly or mostly after the time windows generally associated with visual word recognition (Grainger & Holcomb, 2009). As noted above, there are likely multiple contributing neural sources and sub-components to the LPC. One possibility is that components associated with attentional engagement are most strongly associated with results like those of Gable et al. (2015) and the early portion of the LPC observed here, while the later portion of the effect seen here is more closely related to context updating and memory processes.

Open questions and future work

It will be important for future research to continue to test these ideas in a number of ways. Following from the questions raised above, work with greater spatial resolution, such as joint ERP-MEG and ERP-fMRI studies should examine the LPC elicited by words and auditory stimuli. Ultimately correlations between the LPC and activations in MRI are only correlations: they don't prove which regions generate the LPC. However, regions that are consistently correlated across different stimulus type and modalities are stronger candidates. Strategically testing situations in which a visual attention account seems less plausible will provide a stronger test of that theory of the LPC. It will also be important for more studies to examine ERPs to auditory emotional stimuli. Only a few studies have done so to date and these have generally been focused on examining effects of prosody and as such have not used designs that are comparable to visual studies (Kotz & Paulmann, 2007, 2011). No study has directly compared effects to the same words presented visually and auditorily.

Combining ERPs with methods with better spatial resolution may also help to test the relationship of the LPC to the P300 and context updating, particularly if, like the present work, manipulations expected to elicit a classic P300 effect are also included. MEG studies may be particularly interesting. MEG is sensitive to a smaller subset of neural regions than ERP (primarily cortical sulci near the surface of the scalp), so ERP and MEG effects will not always parallel each other. It is apparently well-known among the MEG community that the P300 does not always show up as expected in MEG studies, suggesting that at least some of its generators are not detected by MEG. If robust LPC effects were found in the absence of strong P300 effects in MEG studies, this would provide evidence that the components are distinct. On the other hand, if both effects are small, absent, or distributed differently compared to ERP, this would provide evidence in the opposite direction. Although a few studies have examined early effects of emotion using

MEG (Keuper et al., 2014; Peyk, Schupp, Elbert, & Junghofer, 2008), none that I'm aware of have examined (or at least reported) an MEG correlate of the LPC.

It will also be important for future studies to further examine the relationship between the LPC and memory. One reason that posterior memory effects may have been weak in the present study is that words were relatively homogenous within each emotional category. Since effects were tested in an orthogonal design, the effect of memory status was tested independently and there may have been relatively little trial to trial LPC amplitude variability once emotional category was controlled for. This would tend to produce a weak relationship between the LPC and later memory. While clear categorical differences were necessary for testing categorical probability effects in the present work, future work in which more diverse emotional stimuli are used may provide a better test of the LPC's relationship to memory.

Summary and conclusions

Questions about the function represented by ERP components are inherently difficult to answer (Luck, 2014, Ch. 3 Suppl.), and functional theories are still debated for many components that have been studied for decades such as the N400 (Kutas & Federmeier, 2011) and P300 (Polich, 2012). Of course, this problem is not unique to ERP components: many of the same problems are faced when using fMRI and other techniques to try to understand the function of particular brain regions as can be seen in the discussion of the function of the medial prefrontal cortex in Chapter 1. Nevertheless, these questions are important. Although techniques like fMRI and ERP can often provide key insights while remaining agnostic to the exact meaning of the regions/components activated (Donchin, 1981; Luck, 2014, Ch. 4), it should be obvious that understanding the cognitive processes that are being modulated in our studies will greatly increase our understanding of the brain and cognition. In the case of the LPC, very little work has been done to date to attempt to understand the precise function of this component. Consequently, the theoretical perspective and empirical results presented here should be seen only as a starting

point. The results of the present work provide at best circumstantial evidence for the P300-context updating theory of the LPC. However, it is hoped that future work testing predictions of this theory and others (e.g., Gable et al., 2015) will increase our understanding the LPC, emotion, and cognition more generally whether the specifics of the theory argued here prove to be correct or even widely off the mark.

Chapter 3: The problem of evidentiary value in the statistical analysis of ERP data and mass univariate statistics as a solution

Introduction

> *We are awash in a sea of uncertainty, caused by a flood tide of sampling and measurement errors, and there are no objective procedures that avoid human judgment and guarantee correct interpretations of results.* (Abelson, 1997, p. 13)

The predicament Abelson identifies is perhaps particularly acute in event-related potential (ERP) research. The event-related activity of interest is embedded in much larger noise in the raw EEG, leading to a smaller signal to noise ratio than in many other domains. In addition, the relatively large resource and time commitment to collect ERP data means most ERP studies have small sample sizes, further increasing sampling error.

Because we are so accustomed to the procedures we use to deal with these problems, it is easy to forget just how much of a challenge we face. Statistical techniques are what allow us to draw conclusions in the face of noisy data. With procedures like ANOVA and regression, we divide the variability into signal and noise. Hypothesis testing is meant to determine when a difference or relationship is likely to represent something real (in the population) versus a random result of sampling error. However, it is easy to forget that the statistical procedures we use are fragile and contingent. They rely on assumptions which are often false and can easily be used in ways that undermine the conclusions they allow us to draw—or render their results invalid altogether. It is easy to suspend human judgment and use statistics mindlessly (Gigerenzer, 2004)—"to apply statistical tests in a quasi-mechanical way, without giving adequate attention to

what questions these numerical procedures really answer" (Gigerenzer et al., 1989, p. 106)—and in so doing to draw conclusions that seem to be justified by the data but in fact do not follow at all.

In this chapter I discuss issues in the analysis of ERP data that undermine our ability to draw strong conclusions and contribute to spurious effects in the literature. I then propose permutation-based mass univariate statistics as an approach that fits the nature of ERP better than traditional analysis approaches and offers an improvement with regard to many of these issues. I present new software to implement these tests for factorial designs along with simulation studies examining the properties of these tests. As we will see, at least in some common analysis situations, these mass univariate permutation tests can provide greater power and greater flexibility compared to the traditional mean time window parametric approach while appropriately maintaining the Type I error rate. First though, I review the problems that these tests are intended to address.

P-values, positive predictive value, and the interpretation of null hypothesis tests

Null hypothesis significance testing remains the dominant statistical paradigm in psychology and neuroscience (for better or worse: Nickerson, 2000). Within this framework as it is usually practiced, everything comes down to the *p*-value. It is from the *p*-value that researchers conclude that an effect/difference/relationship is "real"—i.e., whether it reflects a non-zero value in the population rather than the result of sampling error. When the *p*-value is smaller than some pre-determined criterion, we describe an effect as "significant" and it is then treated as real or reliable. Unfortunately, what the *p*-value actually means is widely misunderstood (Greenland et al., 2016; Haller & Krauss; Hubbard & Bayarri, 2003; Nickerson, 2000).

It is generally recognized that it is difficult to draw any conclusions from a non-significant effect for several reasons. In fact, it is difficult to even publish a non-significant effect (a phenomenon that leads to its own problems: Rosenthal, 1979; Simonsohn, Nelson, & Simmons, 2014). We draw conclusions and make claims only when we find significant results, so we are

interested in the reliability/replicability specifically of significant results. That is, a key question regarding the confidence we have in published results and our own results is this: how likely is it that an effect is real (i.e., the null hypothesis is actually false) given that it achieves statistical significance?

The *p*-value itself is often interpreted as giving us the answer to this question, but this is incorrect. The *p*-value indicates the probability of obtaining the results we obtained (or results reflecting a larger difference/effect) *if the null hypothesis were true*. It does not give us the probability that the null hypothesis is true given the results we obtained (Berger & Sellke, 1987; Colquhoun, 2014; Hubbard & Lindsay, 2008; Nickerson, 2000; Sellke, Bayarri, & Berger, 2001). In formula form:

$$p = P(\geq data \mid H_0) \neq P(H_0 \mid data)$$

The probability that the null hypothesis is actually false given a significant result (e.g., $p < .05$) is referred to as the positive predictive value (PPV; Ioannidis, 2005) and can be calculated as follows:

$$PPV = P(H_1 \mid sig.) = \frac{P(H_1) \times Power}{P(H_1) \times Power + P(H_0) \times Type\ I\ error\ rate}$$[18]

Thus, in addition to the *p*-value, the confidence we can have in a statistically significant finding is related both to the prior probability that the null is false and to power. The relationship to power in particular is often not appreciated, as power is commonly understood to be related only to Type II errors, and not Type I errors. (i.e., it is assumed low power is irrelevant once a significant effect is found; for further discussion of the relationship between power and reliability of results, see Button et al., 2013; Colquhoun, 2014; Szucs & Ioannidis, 2017).

[18] Following common practice, the term "Type I error rate" here and throughout this chapter does not refer to the overall proportion of studies that will produce a Type I error, but rather to the probability of a Type I error given that the null is true. Similarly, power refers to the probability of obtaining a significant result when the alternative hypothesis is true.

Let us consider a hypothetical experiment. Assume that we have moderate power of 0.50 and we appropriately maintain the Type I error rate at α = 0.05. Since we want an unbiased test of which hypothesis is true, we will give them equal priori probability: $P(H_0) = 0.5$ and $P(H_1) = 0.5$. In this case, the probability that the null is true given a significant result is 0.09 (the probability that the null is true is, of course, 1 – PPV). Note that this is significantly higher than the *p*-value, (which is 0.05 or less) or the α-level (which is .05). In fact, given an unbiased prior distribution on the hypotheses (i.e., the alternative and the null are assumed to be equally likely[19]), the probability that the null is true will be greater than the *p*-value and α-level unless power is .95 or greater, a value very rarely realized. In other words, the *p*-value is nearly always an underestimate of the probability that the null is true (see also Berger & Sellke, 1987; Colquhoun, 2014; Sellke et al., 2001).

Now consider a common situation in ERP studies: a small effect is found in study with a relatively small sample size (so power is low) in a time window and/or spatial region where no effect was expected a priori (so the priori probability of the alternative should be considered low). For purposes of calculation, we'll describe this situation as having a Type I error rate = 0.05, power = 0.20, and the prior probability that the alternative is true = 0.10. In this case the probability that the null is true is nearly 60% even though $p < .05$. In other words, despite a significant effect, the null hypothesis is more likely to be true than false.

Another underappreciated problem is that while α may be set at 0.05 in most neuroscience experiments, the Type I error rate is rarely truly maintained at this level.[20] Psychologists and

[19] In this case (i.e., $P(H_0) = 0.5$ and $P(H_1) = 0.5$), the calculation of PPV can be reduced to:

$$PPV = \frac{Power}{Power + Type\ I\ error\ rate}$$

[20] Here and throughout this chapter, I use alpha to refer to the value set by the experimenter (i.e., the value to which the calculated *p*-value is compared to determine significance—nearly always 0.05), intended as the long-run Type I error rate. In contrast, I use Type I error rate to refer to the actual probability of finding a significant result given that the null is true. When researcher degrees of freedom and multiple comparisons are appropriately dealt with (and all the assumptions of the test are met), these values match, but we are concerned here with situations where they do not.

neuroscientists are becoming increasingly aware of how flexibility in research design and analysis can increase the Type I error rate—so called "researcher degrees of freedom" (Simmons, Nelson, & Simonsohn, 2011). As will be discussed in greater detail below, in the realm of ERP research, one example is the tendency to conduct statistical tests in multiple time windows or to choose time windows based on observation of the data. This creates a multiple comparisons problem, allowing for multiple chances to find a significant effect, and thereby increasing the probability of finding at least one significant effect even if all null hypotheses are true. The case described above, where a small effect was found in a time window where no effect was expected, likely reflects a situation where multiple time windows were tested (including time windows that did not correspond to a particular a priori hypothesis). Let us imagine that by testing several time windows, the overall Type I error rate is increased to 0.10 instead of the nominal 0.05. Again assuming power is 0.20 and the prior probability of the alternative is 0.10, there is now an 82% chance that the null is actually true even though we have rejected the null with $p < .05$. Even if we assume the alternative hypothesis has a priori probability of 0.5, the probability that the null is true given a significant result in this situation (low power, increased Type I error rate) is greater than 30%—much higher than the 5% many researchers would assume.

To summarize: when power is low and/or the Type I error rate is inflated, a significant effect may offer little to no evidence that the null hypothesis is actually false, particularly when dealing with unpredicted effects. Unfortunately, these conditions are common. Bakker, van Dijk, and Wicherts (2012) suggest that the average power in psychology is around 0.35 and Button et al. (2013) suggest that the average power for experiments in neuroscience is even worse at 0.2 (see also Szucs & Ioannidis, 2017). As will be discussed below, inflation of the Type I error rate is similarly widespread. This suggests that false results are regularly published and cited. In fact, it is exactly the logic laid out above that led Ioannidis (2005) to famously conclude that *most published research findings are false* (see also Colquhoun, 2014). Unfortunately, there is some

empirical evidence to support this claim (Open Science Collaboration, 2015) as well as evidence that the problem may be worse in cognitive neuroscience than other areas of psychology (Szucs & Ioannidis, 2017). These issues are directly related to the much discussed "replication crisis" in psychology and neuroscience (Pashler & Harris, 2012; Pashler & Wagenmakers, 2012) and associated problems facing the field (Srivastava, 2016).

Maximizing PPV

Clearly PPV values as low as those calculated above significantly undermine the research process. The point of null hypothesis tests is to allow us to distinguish really existing effects from those generated by random sampling error. If the probability that the null hypothesis is true when we find a significant effect is 30% or 60% or 80%, then our statistical tests are essentially useless. As long as null hypothesis testing is going to be used (a matter of legitimate debate that will not be discussed here; see Gigerenzer, 2004; Harlow, Mulaik, & Steiger, 1997; Nickerson, 2000), we need to take very seriously research procedures and analysis approaches that maximize PPV.

PPV is affected by three key values: power, the Type I error rate, and the priori probability of the alternative and null. The prior probability of the alternative hypothesis is increased by clear and strongly justified theories and hypotheses (Fiedler, 2017), but beyond that it is not something we control: determining which hypothesis is true is, after all, the point of experiments and null hypothesis testing. In any case, for the purposes of statistical calculation we must be unbiased and cannot simply assume that our theories are true a priori (in which case we might as well not collect data at all). Thus, if we want meaningful hypothesis tests, our goal is to maximize power and minimize the Type I error rate.

Power itself is determined by three values: the effect size, the sample size, and the Type I error rate. The true effect size in the population (i.e., the ratio of the size of the effect to the population variability) is a property of the phenomenon under study and is out of the experimenter's control. However, observed effect size in a given experiment is also influenced by

measurement error, which is under the experimenter's control. Thus, power can be increased by increasing the number of trials contributing to ERPs, collecting data that is as clean and artifact free as possible, properly detecting and removing trials with artifact, designing studies with stimuli and procedures that reduce variability, and so on. These are obviously important issues about which there is much to be said, but they are issues of experimental design and data processing, not statistics, and that is not the focus of this chapter (see Luck, 2014 for general advice). Regarding sample size, EEG studies are expensive in terms of time and money, so sample sizes tend to be small. There is a strong argument to be made that we would be better off if we ran less overall studies and focused on conducting studies with larger sample sizes that are particularly well-designed and motivated (Button et al., 2013; Szucs & Ioannidis, 2017). But that argument is also not the focus of this chapter, so it will suffice to say that all else being equal, it's better to use larger sample sizes.

The last factor affecting power, the Type I error rate, brings us to a crucial tension. Our goal is to minimize the Type I error rate and maximize power, but, all else being equal, reducing the Type I error rate also reduces power. In theory, the Type I error rate is set by the researchers via the alpha level, which is nearly always set at 5% in ERP studies. In fact, practices that inflate the Type I error rate above alpha are widespread, and this issue will be the focus of the following section. Because inflated Type I error rates lower PPV, it is important that we recognize and correct these practices. However, the relationship between power and Type I error makes this particularly challenging: if we properly maintain the Type I error rate by reducing power, we may have done nothing to increase confidence in our results.

Multiple comparisons problems in ERP analysis

Research practices that inflate the Type I error rate and increase the likelihood of false positive results appear to be widespread across psychology and neuroscience (John, Loewenstein, & Prelec, 2012; Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Luck &

Gaspelin, 2017; Masicampo & Lalande, 2012; Simmons et al., 2011; Vul, Harris, Winkielman, & Pashler, 2009). While a number of such practices are likely common in ERP research, here I focus specifically on analysis approaches that inflate the Type I error rate due to both explicit and implicit multiple comparisons across space and time.

Consider an ERP experiment where data is recorded from 32 electrodes at a sampling rate of 500 Hz. Epochs extending to 1000 ms after stimulus onset are extracted from this data. Even after individual trials are averaged to form ERPs, there are 16,000 data points for each subject in each condition. This presents a massive multiple comparisons problem. That is, in such an experiment, it is almost guaranteed that in some time window at some point on the scalp there will be an effect that reaches significance in a conventional analysis even if the null hypothesis is true at all time points and electrodes. Unfortunately, the most common ways of analyzing ERP data do not sufficiently address this multiple comparisons problem. This problem has been reviewed in detail by Luck and Gaspelin (2017), so I will only briefly describe it here (see also, Kilner, 2013; Kriegeskorte et al., 2009; Luck, 2014, Ch. 10).

Analyzing results across time

The most common way to analyze ERP data is to reduce the large amount of recorded data before statistical analysis. In the time domain, this is conventionally achieved by identifying a time window of interest and either calculating the maximum or minimum value (peak amplitude) or average value (mean amplitude) across that time window (Luck, 2014, Ch. 10). The key issue is how to choose the time window(s) to use. It is common to identify time windows based on examination of the obtained data, but this introduces a significant bias. As noted above, it is nearly guaranteed that there will be a difference that surpasses a conventional significance threshold *somewhere* in the spatiotemporal matrix of ERP data, even if all null hypotheses are true. Choosing a time window based on where differences are observed in the data is essentially equivalent to conducting an analysis at all possible time windows and reporting the one that

produces the largest effect. Framed in this way, the multiple comparisons problem and inflation of the Type I error rate should be obvious. Because all these many possible analyses are not actually conducted, this is often called the problem of *implicit* multiple comparisons, and many researchers appear unaware of the extent to which this can inflate the false positive rate and undermine confidence in significant results.

It is easy to understand why this approach is so common: if a time window is chosen a priori, an experimenter will have reduced power unless the effect coincides exactly with the chosen time window and will have no power to detect effects wholly outside the time window. An approach that only allows researchers to find effects they already expect is not particularly conducive to the practice of basic science. And, as laid out above, low power is as much a hinder to confidence in our conclusions as Type I error rate. As a result, there is very much a *damned if you do, damned if you don't* aspect to choosing time windows in ERP analysis. A variety of methods have been proposed to deal with this dilemma within the traditional mean time window parametric approach (Brooks, Zoumpoulaki, & Bowman, 2017; Luck & Gaspelin, 2017), but none is wholly satisfactory or addresses all situations.

Whether time windows are chosen a priori or based on the observed data, it is also common to examine multiple time windows. Again, the motivation is understandable: we want to be able to detect effects wherever they may be and it seems contrary to the scientific endeavor to ignore the majority of the data we collect. When each time window examined is motivated by distinct a priori hypotheses, this should not be considered a multiple comparisons issue. But when time windows are examined with no associated hypotheses (e.g., for completeness) or multiple time windows are examined because the researchers are not sure where to expect effects, this obviously increases the chance of obtaining significant effects where no real effect exists.

Analyzing results across space (electrodes)

Reducing data in the spatial domain can be done similarly to the time domain: by averaging across a subset of electrodes. As in the time domain, if the electrodes in this subset are chosen based on where effects occur in the obtained data, this introduces a significant bias. If not, all the problems associated with strict a priori time windows apply. Another common approach is to include spatial factors in a factorial ANOVA. This approach has the advantage of obviating the need to choose an a priori region and it does reduce the multiple comparisons problem. However, it does not eliminate the problem and it introduces additional problems.

To understand how to interpret an interaction with region/electrode, it is important to understand the biophysical basis of EEG/ERP (for a more detailed explanation see Buzsáki, Anastassiou, & Koch, 2012; Luck, 2014). ERP components are primarily a reflection of post-synaptic potentials in cortical pyramidal cells. These post-synaptic potentials create electrical dipoles: that is, they generate positive voltage on one side of the pyramidal layer and a corresponding negative voltage on the other side. The electrode montage on the scalp only covers approximate half of a full sphere. So when an electrical dipole is oriented vertically (i.e., parallel to the spine), only one side of the dipole will be recorded and all or most of the electrodes on the scalp will record voltage in the same direction. If the dipole is oriented horizontally (perpendicular to the spine), both sides will be recorded, and a component will appear as a positive deflection at some sites and a negative deflection at others. As a result, the orientation of the source dipoles of an effect will determine whether it is registered in a factorial ANOVA as a main effect, an interaction with region/electrode, or both. It is also important to note that a dipole on a gyrus and a dipole in an adjacent sulcus will have different orientations by 90º. Thus physically close sources in functionally similar regions of the brain can register very differently on the scalp and in a factorial ANOVA with region/electrodes factors.

The important point is that in such a factorial ANOVA, a main effect of factor A and an A x electrode/region must generally be interpreted in the same way: either simply indicates an effect of factor A with telling us much else (for a more detailed description of why, see Luck, 2014, pp. 316-317). All effects are generated by dipoles that will affect different electrodes differently and thus will differ across the scalp, so every effect will generate an interaction with electrode/region. Unless a dipole is oriented such that the positive and negative ends cancel each other perfectly across the recorded electrodes, it will also generate a main effect. Whether one or both of these effects reaches statistical significance is simply a combination of power and dipole orientation.

This means two things. First, every effect has two chances to reach significance. If multiple spatial factors are included (e.g., an anterior-posterior factor and a hemisphere factor) then there are even more chances. This obviously inflates the Type I error rate.[21] Second, the variance due to a given effect is divided between two or more sources in the ANOVA model, which undermines power. Thus dealing with data across electrodes by including spatial factors in an ANOVA can actually inflate Type I error rate *and* reduce power, a particularly bad combination.

There is an additional problem with electrode/region factors in ANOVA models. One of the assumptions of the ANOVA is that all data points are independent observations. Because all electrodes record voltage from a (differently weighted) mixture of the same underlying neural sources, different electrodes or regions at the same time point are fundamentally not independent observations. This means that the degrees of freedom for any interaction with electrode/region will be inflated, leading to a narrower F distribution than is justified, which inflates the Type I error rate.

[21] It is a common misperception that factorial ANOVA controls the Type I error rate or corrects for multiple comparisons across the multiple effects tested within the model (i.e., all main effects and interactions). In fact, a factorial ANOVA consists of multiple completely independent (i.e., orthogonal) tests which each have Type I error rate at the chosen alpha. Generally, it is assumed that each effect tests a distinct hypothesis and so this is not considered a multiple comparisons issue, but for the reasons laid out above, this is not the case with interactions involving spatial factors.

Permutation-based mass univariate statistics

One way to address the implicit multiple comparisons problem is to make it explicit: conduct a separate statistical test at multiple or all time points and electrodes and apply some multiple comparisons correction to control the family-wise Type I error rate. Because this method deals with a large number of dependent variables by conducting many univariate analyses, it is called the "mass univariate approach".[22] The mass univariate approach is at the core of standard analysis approaches in functional MRI research (Woolrich, Beckmann, Nichols, & Smith, 2009) and other domains such as genetics, but has been much less commonly used in ERP research.

The challenge of applying the mass univariate approach is that conventional multiple comparison corrections such as the Bonferroni correction reduce power to unacceptably low levels when applied to the large number of time points and electrodes in a standard ERP study. Several alternative methods of correction more appropriate for mass univariate analysis have been developed. These generally fall into one of two types: false discovery rate-based corrections and permutation-based corrections. Here I focus on permutation-based approaches (see Groppe et al., 2011a for a review of false discovery rate approaches as applied to ERP data). I will first review the basic logic of permutation tests and then discuss how they can be used to correct for multiple comparisons.

Introduction to permutation-based statistical tests

A permutation test is a form of non-parametric null hypothesis test. Rather than calculating a *p*-value from a theoretical distribution justified by the assumptions of the test (e.g., that the data are drawn from normally distributed populations), the null distribution is empirically constructed by resampling the data. Consider a simple within-subjects priming experiment with two mean reaction time measures, one for the primed condition and one for the unprimed condition, from

[22] This is in contrast to multivariate approaches, which are generally not feasible with ERP data because the number of dependent variables (i.e., time point and/or electrodes) exceed the sample size (Wheldon, Anderson, & Johnson, 2007).

each of 24 subjects. To conduct the permutation test, the primed and unprimed data is randomly switched or not switched for each subject independently. In other words, we create a situation in which the null is true by definition (because now the primed and unprimed conditions have data draw from the same population) but the structure and variability of the data is otherwise the same. A paired sample *t*-statistic is calculated for this randomly permuted data and this is repeated for a large number of permutations.[23] The proportion of *t*-values in this null distribution that are as extreme or more extreme as the observed *t*-value from the unpermuted data is the *p*-value—i.e., the probability of obtaining the observed *t* if the null were true (for a more in depth description of permutation tests than provided here, see Good, 2005; Groppe et al., 2011a; Manly, 1997; Maris & Oostenveld, 2007).

When the assumptions of the *t*-test are met, the empirical distribution of *t*-values will be a good approximation of the *t* distribution,[24] and the permutation test will therefore give the same answer as the parametric *t*-test with some generally small error assuming reasonable sample sizes. When the normality assumption of the *t*-test is not met, the permutation test will give a more accurate *p*-value than the parametric test. Thus, under conditions where both tests are justified, the parametric and permutation tests are approximately equally powerful and both maintain the Type I error rate appropriately, but the permutation test is justified in a wider set of circumstances.

Permutation statistics for factorial designs

It is also possible to conduct permutation based versions of ANOVA using the *F*-statistic. For one-way ANOVA, this works exactly the same as the *t*-test (i.e., the data for each subject is randomly reshuffled among all the conditions of the independent variable). For factorial designs, the situation is somewhat more complicated. The issue of permutation tests for factorial designs

[23] In theory, all possible permutation should be examined, but even with moderate sample sizes this is often impractical (e.g., with 24 subjects in a two-condition experiment, there are over 16 million permutations). Fortunately, a large number of random permutations will give a sufficiently precise estimate of the *p*-value.
[24] With an infinitely large sample size and infinite permutations, the empirical distribution would perfectly match the parametric *t*-distribution.

and multiple regression has been discussed in depth by other authors (Anderson, 2001; Anderson & Ter Braak, 2003; Winkler, Ridgway, Webster, Smith, & Nichols, 2014) and so will only be briefly described here.

Let us consider a simple two-way design with factors A and B. There are three effects in the ANOVA model: A, B, and AxB. In parametric ANOVA, the *F*-value calculated for each effect is compared to a different *F*-distribution depending on its degrees of freedom. In the same way, in the permutation ANOVA separate null distributions must be constructed for each effect. The question that guides these permutations is which data is exchangeable under the null hypothesis. For example, we can't test the effect of A by freely permuting data across all cells of the factorial design because the data is only exchangeable across all cells if the null is true for all three effects in the design; we wish to independently test whether the null hypothesis is true for A. Appropriate exchangeability under the null can be achieved in one of two ways (discussed in greater detail in Anderson & Ter Braak, 2003). The first is via restricted permutations. For example, to test the effect of factor A, we would only permute data across conditions of A while keeping each data point in the same condition of factor B.[25] This provides an exact test: that is, the Type I error rate is maintained at exactly the specified level. The second method is to subtract the effects that are not being tested—in this case the effect of B and AxB interaction—to obtain residuals such that the null effect is true for B and AxB. These residuals can then be permuted freely across all cells of the design under the assumption of the null hypothesis for A. However, the effects that are subtracted to form the residuals are only estimates of the true population effects. As a result, this permutation of residuals provides only an approximate test: the Type I error rate will be asymptotic to α as the sample size increases (because the estimate of the effects being subtracted is asymptotic to the population parameter as the sample size increases).

[25] Mathematically, this is equivalent to averaging across all the conditions of B and then simply conducting a one-way ANOVA, and this is an alternative way of computing the same result.

From the perspective of maintaining the Type I error rate, the restricted permutation method is clearly preferable, but this method is not possible for all effects. For the AxB interaction, if permutation is restricted within both A and B, the only permutation possible is the original data. In this case a different form of the restricted permutation method is possible in some cases. If either factor A or B has only two levels, it is possible to subtract across the levels of that factor and then conduct a one-way ANOVA (i.e., a test of whether the difference in factor A changes across levels of factor B is a test of the interaction effect). This will provide an exact test of the interaction. However, for designs in which more than one factor has more than two levels, only the permutation of residuals method is available. The actual Type I error for the permutation of residuals method with ERP data will be a focus of simulation studies later in this chapter.

Permutation-based correction of multiple comparisons

For present purposes, the value of permutation tests is that they can be used to correct for multiple comparisons while maintaining acceptable power by taking into account the structure and correlation of ERP data across time and space. There are two main methods for doing this currently in use. I will discuss these in terms of ANOVA designs and the *F*-statistic, but they can also be conducted as *t*-tests or other statistics. These techniques are described in greater detail in Groppe et al. (2011a).

The first method uses the F_{max} statistic (first described by Blair & Karniski, 1993). In this approach, for each permutation of the data an *F*-statistic is conducted at each time point and electrode. The largest *F*-value observed across all time points and electrodes is recorded as F_{max}. This is repeated for a large number of permutations to empirically construct a distribution of F_{max} under the null hypothesis. The observed *F*-value at each time point and electrode is then compared to the F_{max} value at the 1 – α percentile of the distribution and data points with an *F*-observed exceeding this critical value are considered significant. If the voltage recorded at all electrode and time point combinations were completely independent of each other, this approach

would give approximately the same answer as the Dunn–Šidák correction used with parametric ANOVA. However, data at adjacent temporal and spatial position are obviously strongly positively correlated. This means that the Dunn–Šidák correction (which assumes independence) or Bonferroni correction (which assumes negative correlation) will be conservative (the Type I error rate will be significantly below the nominal α), whereas the F_{max} permutation procedure will have greater power while accurately maintaining the Type I error rate at α.

The advantage of the F_{max} procedure (over the cluster mass procedure discussed next) is that it provides strong familywise error correction: each electrode-time point combination can be regarded on its own with the same confidence as any significant effect. The procedure therefore allows for statistically justified claims to be made about the exact time points and locations that are found to be significant. The downside to the F_{max} procedure is that it may have relatively low power compared to traditional analysis approaches. This also somewhat undermines its advantages, since the extent of significant results will underestimate the true extent of effects unless the sample size or effect size are very large. This will be discussed more in relation to the simulation results later in this chapter.

The second approach is the cluster mass permutation test (first described in the context of MRI research by Bullmore et al., 1999; first described for EEG data by Maris & Oostenveld, 2007). This technique is based on the knowledge that true effects are not likely to appear at individual time points and/or electrodes, but instead will appear at several adjacent time points and electrodes. An a priori threshold for inclusion in a cluster is specified. Often this is the *F*-value that would indicate significance if no multiple comparisons procedure were applied. For each permutation, any adjacent electrodes and/or time points that exceed this threshold are considered to be in the same cluster and all the *F*-values in a given cluster are summed together to generate a cluster mass statistic. The largest such cluster mass from each permutation is retained and these form the null distribution. All clusters mass statistics in the unpermuted data are then

compared to the cluster mass value at the 1 – α percentile of the null distribution and any that exceed this value are considered significant.

Unlike the F_{max} procedure, the cluster mass procedure provides only weak family-wise error correction. If the null hypothesis is true at all electrodes and time points examined, there is only an α chance that any cluster will reach significance; however, significance applies only to the cluster as a whole—it cannot be used to conclude with any specified level of confidence or likelihood of Type I error that an effect is actually present at any given time point and/or electrode included in the cluster. In many cases this downside is not much a problem. In most ERP studies, hypotheses and conclusions are not dependent on precisely when an effect starts and stops, for example. In fact, the traditional mean window/ROI approach tells us even less about the temporal and spatial extent of effects. The upside to the cluster based procedure is that it can be significantly more powerful than the F_{max} procedure, particularly for effects that are widespread in space and time.

Software and Simulation Studies

There are currently two widely available implementations of permutation tests for ERP data: FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), which is a complete EEG data processing and analysis package, and the Mass Univariate Toolbox (Groppe et al., 2011a; Groppe, Urbach, & Kutas, 2011b), which is generally used in conjunction with EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014).

These toolboxes are fairly easy to use and offer many options and features. However, both are restricted in the research designs that can be accommodated: the Mass Univariate Toolbox only implements *t*-tests; FieldTrip implements one-way ANOVA, but not factorial ANOVA. It is possible to use both FieldTrip and the Mass Univariate Toolbox to test main effects in factorial designs by first averaging across all other factors and then conducting a *t*-test or one-way ANOVA. Similarly, it is possible to use the method of subtraction described above to reduce some

interaction effects to a one-way design that can be tested with these toolboxes. But there is no widely available software that uses the permutation of residuals method and easily accommodates all factorial designs.

Here I introduce the Factorial Mass Univariate Toolbox (FMUT): a new extension of the Mass Univariate Toolbox to analyze within-subject factorial designs via either the F_{max} or cluster mass procedures.[26] The functions provided in this extension work similarly to existing Mass Univariate Toolbox functions and replicate much of their functionality, but extend this to factorial designs. For main effects, the FMUT functions provide an exact test via the restricted permutation method. For interaction effects they can utilize either the restricted permutation method (where possible) or the permutation of residuals method.

In the following section, I use this software to analyze the properties of these tests via simulation studies. These studies address two key issues relevant to the widespread adoption of these analysis techniques:

1. Although the permutation of residuals method is only guaranteed to maintain Type I error rate as the sample size grows large, previous simulation studies have shown that the Type I error rate is not inflated even with moderate sample sizes (Anderson & Ter Braak, 2003; Still & White, 1981). However, the actual Type I error rate will depend on the nature of the data. Previous studies have shown that permutation of residuals appropriately controls the Type I error rate and corrects for multiple comparisons with fMRI data (Winkler et al., 2014). In the first set of simulations, I test whether this also holds with ERP data.
2. The increased flexibility while appropriately controlling the Type I error rate that permutation tests offer is not of much value if they don't also provide sufficient power. In the second set of simulations, I test the relative power of the F_{max} and cluster mass

[26] This software is under a free and open source license and will soon be made publicly available. For now, a beta version is available by contacting the author.

methods compared to traditional mean window parametric ANOVA for some common ERP components.

General methods

Simulation studies were conducted according to the same general approach described by Groppe et al. (2011b). Realistic EEG noise epochs were obtained from EEG data collected in our lab from 49 subjects completing an AX-CPT task as part of several studies. Briefly, participants saw a series of letters and had to press a button whenever they saw X preceded by an A (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956). EEG data was recorded from 32 Ag/AgCl electrodes using a BioSemi ActiveTwo system (biosemi.com), low-pass filtered online at 102.4 Hz, and continuously sampled at 512 Hz. In EEGLAB (sccn.ucsd.edu/eeglab; Delorme & Makeig, 2004) and ERPLAB (erpinfo.org/erplab; Lopez-Calderon & Luck, 2014), the continuous EEG was referenced to the average of the mastoids and high pass filtered at 0.05 Hz, then epochs were extracted from 200 ms before to 1100 ms after each letter and baseline corrected by subtracting the mean voltage from -200 to 0 ms of each epoch.

The AX-CPT task consisted of 400 trials with two events in each trial for a total of 800 epochs. Trials with artifact (blinks, eye movements, bad channels, etc.) were detected via algorithms implemented in ERPLAB and discarded. The remaining trials were low-pass filtered at 30 Hz, down-sampled to 128 Hz, and used to extract epochs of background EEG noise. For each epoch, the averaged waveform (i.e., the ERP) of the condition for that epoch was subtracted from the raw EEG. For example, for an X following an A, the average waveform for all X's following A's was subtracted. This removes (an estimate of) the event-related activity and leaves the EEG background noise (which sums and averages to zero across all trials within and across conditions).

Studies were simulated by drawing a random subset of subjects, then drawing a random subset of their noise trials and randomly assigning these to conditions. This procedure creates a situation in which the null hypothesis is known to be true (all conditions are drawn from the same

population) and the distribution and structure of variability of the EEG across time and space is realistic. The statistical test of interest was then performed for each stimulated study, and the Type I error rate was recorded as the percentage of studies showing a significant effect at any time point or electrode.

To test power, realistic effects were drawn from ERP data collected in our lab (studies described below) by the following procedure. 1) The conditions of interest were taken from grand mean ERPs and low pass filtered at 7.5 Hz to remove high frequency noise. 2) All time points before and after the time window containing the effect of interest were set to the average of all the conditions so that the null hypothesis was known to be true at these time points. Studies were then simulated via the procedure described above with the additional step that the ERP effect was added to the averaged noise for each study. Individual differences were simulated by multiplying the waveform for the added effect by a value randomly drawn from a normal distribution with a mean of 1 and standard deviation of .1 for each subject for each condition. In other words, the size of the effect varied somewhat from subject to subject, as would be the case in a real experiment. Power was calculated as the percentage of studies showing a significant effect at any time point or electrode.

Simulations of the Type I error rate of approximate tests

In the first four simulations, I tested the Type I error rate of the approximate method of calculating an interaction effect. For all these simulations, I tested the two-way interaction effect in a 3 x 3 design, which is the simplest design for which the approximate method would be required.[27] This effect was calculated with simulated studies where the null was known to be true

[27] Generally speaking, the Type I error rate will be closer to the nominal α as the number of levels in each factor increases. This is because as the number of levels increases, the amount of data increases faster than the number of means that need to be calculated to form the residuals (e.g., in a 3 x 2 design, 5 means must be calculated based on 6 cells of data, whereas in a 3 x 3 design, 6 means must be calculated based on 9 cells of data). The 3 x 3 design therefore gives an upper limit for the expected Type I error rate for two-way interactions in cases where the approximate test is required.

as described above. Early and late time windows at all electrodes were examined for both the F_{max} and cluster mass procedures. Studies consisted of 24 subjects, with 40 trials in each of the 9 cells of the design. 2,500 studies were simulated. Both the cluster mass and F_{max} procedure were conducted with 2,500 permutations per test. For the cluster mass procedure, the threshold for cluster inclusion was $p \leq .05$ uncorrected and channels within approximately 7.5 cm of each other (assuming a head circumference of 56 cm) were considered spatial neighbors.[28]

The results are shown in Table 3.1. The permutation of residuals method led to only a minimally inflated Type I error rate for both methods in both early and late time windows.

Mult Comp Corr	Time Window	Type I Error Rate
cluster mass	0 - 300	0.054
cluster mass	500 - 900	0.053
F_{max}	0 - 300	0.052
F_{max}	500 - 900	0.058

Table 3.1: *Simulation studies of Type I error rate*. Simulations of a null interaction effect for a 3x3 design. See text for details of simulation methods and parameters.

Simulations of power for realistic ERP effects

The relative power of the F_{max}, cluster mass approach, and the mean window/ROI approaches was examined for two common ERP components: the N400 and the P300. For each effect several time windows and electrode selections were examined. For the parametric approach, amplitude was averaged across all time points and electrodes within the specified spatiotemporal ROI. For the cluster mass test, clusters were found within the same time window and electrode set. All studies had 24 subjects and 20 trials per condition. 2,500 studies were simulated for each effect examined with 2,500 permutations for each F_{max} or cluster mass test. All cluster mass analyses were conducted with a cluster inclusion threshold of $p \leq .05$ uncorrected and all channels within approximately 7.5 cm of each other (assuming a head circumference of

[28] There are a number of parameters, such as filtering, sampling rate, cluster inclusion threshold, spatial neighbor threshold, etc. that can affect the power and sensitivity of the F_{max} and cluster mass tests. These considerations are discussed in Groppe et al. (2011a) and Ch. 13 (online) of Luck (2014).

56 cm) were considered spatial neighbors. Power was calculated as the percentage of studies where any significant effects were observed. To compare the spatial extent of effects detected by the F_{max} and cluster mass approaches, for each study showing any significant effect I calculated the mean number of time points showing a significant effect (i.e., included in a significant cluster or exceeding the F_{max} critical value) at any of the centroparietal electrodes where the effects were largest (see definition of centroparietal ROI below). This number was then multiplied by the number of ms between sampling points (7.8) to put it in units of ms instead of sampling points.

N400

A canonical N400 effect was extracted from a study in which words within sentences were either highly expected (high cloze probability) or unexpected (but plausible) in an unconstraining context (Wlotko & Kuperberg, 2017). This effect was centroparietally distributed and peaked around 430 ms. For the purposes of simulating data, all time points before 200 and after 700 ms were set to null effects. Because this effect was quite large, power would approach or equal 1 for standard analysis approaches. The size of the effect relative to the error trials was therefore reduced by a third.

Results of simulations are reported in Table 3.2. In line with many previous studies, I first analyzed this effect between 300 and 500 ms at a spatial ROI of 5 centroparietal electrodes (Cz, CP1, CPz, CP2, Pz). This represents an a priori prediction that matches where the effect was actually observed. Even in this best case scenario for the parametric approach, the cluster mass approach had slightly greater power and the F_{max} approach had approximately the same power. The advantage of the mass univariate approaches became particularly evident when a priori assumptions were relaxed. When the effect was examined in a 200 – 600 ms time window at all electrodes, the cluster mass approach displayed approximately the same power as in the more restrictive time window and continued to have greater power than the parametric approach did in the more restrictive time window. The F_{max} approach had reduced power, but outperformed the

mean amplitude approach. These advantages may be in part due to the fact that the N400 effect actually extended beyond the 300 – 500 ms time window. This is likely a common situation (i.e., that effects extend beyond the a priori time window in which they are examined). However, to examine the impact of including additional null time points, I reran the 200-600 ms full electrode analysis with all time points outside 300 – 500 ms set to null. In other words, in this simulation, half of the time points examined showed no effect. This had no effect on the mass univariate approaches, but significantly reduced power further for the mean amplitude approach. Finally, even in a completely bottom-up approach examining the entire epoch at all electrodes, the mass univariate approaches showed only moderately reduced power, while power was reduced even further for the mean amplitude approach.

Effect	Time Window	Electrodes	cluster mass power	F_{max} power	mean amplitude parametric power	cluster temporal extent	F_{max} temporal extent
N400	300 - 500	centroparietal ROI	0.82	0.77	0.76	144	73
N400	200 - 600	full 32	0.81	0.68	0.45	186	41
N400 (restricted)	200 - 600	full 32	0.79	0.68	0.29	159	39
N400	0 - 1000	full 32	0.69	0.57	0.15	216	34
P300	500 - 750	centroparietal ROI	0.39	0.34	0.34	144	45
P300	400 - 900	full 32	0.25	0.22	0.09	210	29
P300	0 - 1000	full 32	0.22	0.17	0.08	247	24

Table 3.2: *Simulations studies of power*. Power refers to the number of simulated studies in which any significant effect was observed. Temporal extent is the average number of time points (in ms) included in significant effects. See text for details of simulation methods and parameters.

The power of the F_{max} test was nearly as good as the cluster mass test in the more restrictive time windows. However, because calculated power reflects whether *any* time point and electrode combination showed a significant effect, this is somewhat misleading. Examination of the temporal extent of significant effects shows that both tests underestimate the true duration of the effect, but the cluster mass test had significantly greater sensitivity to the true range of the effect.

P300

A P300 effect was drawn from the 20%, 50%, and 80% conditions of the name-gender oddball task described in Chapter 2. This effect peaked around 620 ms and was centroparietally distributed. For the purposes of the simulation, all time points before 430 ms and after 980 ms at all electrodes were set to the average of the two conditions (i.e., the null hypothesis was true). Because this effect was quite large, the size of the effect relative to the error trials was reduced by half.

Results of simulations are reported in Table 3.2. I first analyzed this effect between 500 and 750 ms at a spatial ROI of 5 centroparietal electrodes (Cz, CP1, CPz, CP2, Pz). This represents an a priori prediction that matches the observed effect. Here the power of the cluster mass test was slightly greater than the F_{max} test or mean amplitude test, which were equivalent. When the time window was doubled in size to 400 to 900 ms and all electrodes were examined, the power of the mean amplitude test was decreased significantly more than the mass univariate tests. In a fully bottom-up test of the entire epoch and all electrodes, the mass univariate tests continued to show power reduced moderately from the more restrictive time windows. As observed for the N400, the cluster mass test did a significantly better job of approaching the true duration of the effect than the F_{max} test.

Discussion

The traditional ERP analysis approach utilizing parametric, factorial ANOVA to analyze mean (or peak) amplitude across defined time windows suffers from a number of problems. The parametric ANOVA relies on several assumptions that are either violated or not known to be met by most ERP datasets. And choosing spatial and temporal regions for analysis either involves overly restrictive a priori guesses as to where effects will be, or suffers from a multiple comparisons problem. Both of these issues can inflate the Type I error rate, however the methods available for dealing with them within the traditional mean window parametric approach

significantly reduce flexibility and/or reduce power. For reasons laid out in detail in the Introduction, to maximize the ability of null hypothesis tests to discriminate between real and spurious findings, and therefore to produce analyses we can have confidence in, it is important to *both* minimize Type I error and maximize power. Thus in many cases, a significant effect in the mean amplitude parametric approach may not actually offer much evidence of a true effect. The permutation-based mass univariate approach does not require the parametric assumptions of the traditional ANOVA and offers natural ways of making explicit and correcting for multiple comparisons by taking account of the structure of the ERP data and ERP effects. Here I explored the properties of such tests and their suitability for general use in the analysis of ERP data.

First, for some interaction effects, it is not possible to construct a permutation test that guarantees the Type I error rate will conform to the nominal α. In these cases, an approximate test based on permutation of residuals is required. Here I have shown that, consistent with a number of previous simulation studies (Anderson & Ter Braak, 2003; Still & White, 1981; Winkler et al., 2014), the Type I error rate for these tests is only very modestly inflated for realistic ERP data. This was true for both the F_{max} and cluster mass approaches for both early and late time windows in the ERP. These results show that both approaches can be used for complex factorial designs with ERP data.

The simulations examining the power of the permutation-based tests were particularly interesting. When examining ROIs conforming to the observed temporal and spatial location of the effects, the cluster mass test had greater power than the mean amplitude approach and the F_{max} approach, which displayed similar power. When wider time windows were used and all electrodes were examined, the power of the mass univariate approaches was only moderately reduced whereas the mean amplitude approach had significantly less power. The cluster mass approach in particular maintained reasonable power even in fully bottom-up analyses with no a priori temporal or spatial assumptions. Thus, the mass univariate approaches used here have

both greater power than the traditional mean amplitude approach and less sensitivity to the time windows and electrodes chosen for analysis.

Mass univariate approaches are generally portrayed as sacrificing power to maintain the Type I error rate in the face of uncertainty. However, this is based on the assumption that mass univariate approaches will be used with little a priori temporal or spatial assumptions. Here we have shown that when assumptions are matched, permutation-based mass univariate approaches actually have greater power than standard analysis approaches. The reasons for this become clear upon examination. For the F_{max} approach, locations and times where an effect is strongest are tested independently, instead of being averaged with locations/times where the effect is weaker as with the averaged time window approach. This fact is balanced by the multiple comparisons correction, such that the probability that any effect will be found significant is comparable in the two approaches when strong a priori predictions are available, but better for the F_{max} approach when the time window includes many time points/electrodes with null or weak effects.

The cluster mass approach has greater power than the mean amplitude approach because it is able to take advantage of more information about an effect. The probability estimate (i.e., the *p*-value) for the cluster mass test takes advantage of the fact that an effect spread across many time points and or electrodes is less likely to be due to random EEG noise than a real effect. For this reason, it can be more sensitive than the mean amplitude approach that first averages across these data points and does not take account of spatial and temporal extent in its probability calculation.

As should be obvious from this description, this test will be particularly powerful for effects with a significant temporal and/or spatial extent, such as the P300 and N400 effects examined here. This makes it very useful for many common ERP effects; indeed, in many research areas, most of the widely studied effects are of this type (e.g., the N400, P300, LPP, LRP, P600, etc.).

However, the cluster mass test may have low power for short-lived, spatially focal components (e.g., the P1, N170). Groppe et al. (2011b) report significantly better power for a t_{max} test than a cluster mass test for a focal simulated N170 component. However, they did not compare either mass univariate approach to a parametric mean amplitude approach and their analyses used a very broad ROI of 100 to 900 ms at all electrodes. Obviously the mean amplitude approach requires relatively strong a priori predictions to have sufficient power to detect focal effects. It is unclear how the power of the F_{max} or cluster mass tests would compare given the same a priori assumptions, and it will be important for future simulation studies to examine this question.

Although the F_{max} test and the cluster mass test displayed similar power in many simulations, it should be noted that this is only in terms of whether an effect is detected at all. The F_{max} test is still significantly underpowered in terms of showing the temporal (and presumably also spatial) extent of an effect. Although the F_{max} test is often described as allowing for strong conclusions about specific time points and electrodes, this only holds if only Type I errors are of concern. The cluster mass test does not allow for statements to be made about individual time points and electrodes with any given confidence or risk of Type I error, but the simulation results reported here show that in practice it will give a significantly more accurate view of the actual extent of an effect. In fact, at least with the standard $p \leq .05$ uncorrected threshold for inclusion and normal ERP samples sizes, it will probably generally underestimate the true extent of effects, but to a significantly smaller degree than the F_{max} test.

Summary and conclusions

Mass univariate statistics are in many ways a more natural fit for ERP research than traditional approaches. Where the standard mean amplitude approach averages over the complex spatiotemporal nature of ERP to make it fit a standard parametric ANOVA, mass univariate approaches examine data across time and space and take account of the structure of the data,

allowing these tests to be less reliant on a priori spatial and temporal assumptions. This allows researchers to flexibly identify effects while appropriately maintaining the Type I error rate.

There have previously been two primary barriers to the wide adoption of mass univariate tests in the ERP literature. First, implementations of these tests that handle factorial designs and provide tests of interaction effects have not been available. Second, there has been a perception that mass univariate statistics involve sacrificing power for flexibility. The software and simulation data presented here address both of these concerns.

Taken together, the results and considerations presented here suggest that, at least for relatively broadly distributed components, the cluster mass test in particular has many advantages and little disadvantages as compared to traditional mean amplitude analyses or the F_{max} test:

1. In comparison to traditional parametric approaches, the cluster mass test makes less assumptions about the population and thus is applicable in a wider array of circumstances.
2. When spatial and temporal assumptions are matched, the cluster mass test has greater power than the F_{max} or mean amplitude tests. Thus, contrary to common portrayals, the utility of the mass univariate approach is not limited to situations in which researchers are uncertain where to expect effects.
3. The cluster mass test is less sensitive to the time window chosen than the F_{max} test and particularly than the mean amplitude test. It therefore allows greater flexibility and uncertainty about the temporal and spatial location of effects while maintaining acceptable power. This makes it well suited to exploratory analyses, but, importantly, also to analyses where some a priori information is available but not enough to completely constrain analyses to a very specific ROI. This additional flexibility also reduces the dependence of results on experimenter decisions, which is a positive feature even if those decisions are unbiased.

4. Taken together, the previous two points mean that the cluster mass test can offer a substantial improvement of the positive predictive value (PPV) of significant findings over traditional analysis approaches and goes a long way to addressing the tensions and dilemmas examined in the Introduction.
5. Whereas the mean amplitude approach gives only broad boundaries to where an effect occurs and the F_{max} approach significantly underestimates the extent of effects, the cluster mass approach allows for a reasonable bottom-up identification of effects. In most cases this will underestimate the true extent of effects, but this is likely true with any analysis technique given average sample sizes in ERP research. With a highly powered study, the cluster mass technique is likely to give a very good approximation of the true extent of effects.

With the freely available factorial extension to the Mass Univariate Toolbox introduced here and the preceding considerations, it appears there is now little reason why the cluster mass test should not be widely adopted as the default analysis approach for many common ERP analysis situations. No statistical technique can substitute for good theories, strong experimental design, and informed and ethical conduct of research, and no statistical technique is appropriate for all situations and all research questions. However, it is hoped that the techniques and software introduced here can play an important role in addressing the significant issues of evidence and reliability raised in the Introduction to this chapter.

Summary and General Discussion

The three chapters presented here deal with a number of theoretical, empirical, and methodological issues in the investigation of the processing of self-relevant and emotional information. In Chapter 1 we used functional MRI to investigate activity in a region of the medial prefrontal cortex that much previous work has associated with processing self-related information. We showed that the self-other effect in this region was modulated by the emotional valence of social vignettes: specifically, we only observed a self-other effect for positive scenarios. Interestingly, this same region of the mPFC has previously been associated with belief updating specifically in respond to positive self-relevant information (Sharot et al., 2011), suggesting that activity in this region may play a role in maintaining positive illusions about the self.

Chapter 2 presented a theoretical and empirical review of the late positive component (LPC) of the ERP commonly evoked by emotional stimuli. An open question in the literature is what function (i.e., cognitive process or mechanism) is reflected by this component. In addition to being an inherently interesting question, understanding the function of the LPC will likely help us to better make sense of the complex and sometimes contradictory set of findings reviewed in Chapter. More generally, understanding the neurobiology and function of the LPC will make this literature more interesting and informative to psychologists and neuroscientists not inherently interested in the LPC as an ERP component, but interested in the neural systems that generate it or the cognitive mechanism it reflects.

Little empirical or theoretical work has been done to understand the function of the LPC, but the LPC has often been seen as similar or related to the P300 component of the ERP. I outlined a theoretical perspective in which the LPC and P300 are the same component or at least represent very similar cognitive mechanisms—namely, a process by which our model of the current environment is kept up to date. An initial study examining this hypothesis showed results

broadly consistent with it. However, much more research will be required to further test this and other theories of the LPC (e.g., Gable et al., 2015). It is hoped that the discussion and results presented in Chapter 2 will encourage more researchers to take up the question of a functional theory of the LPC.

Although the first two chapter investigate different questions using different methodologies, there are some similarities. In both chapters I suggested that the dependent variable of interest (mPFC activation and the LPC) may be related to some form of belief updating. This of course make sense: both self-relevance and emotion are indications of information which is important to consider and remember, and both lead to better memory. In addition, both the mPFC and LPC show interesting interactions between self-relevance and emotion. I don't mean to suggest that the mPFC is a likely generator for the LPC. However, both may be better understood by modelling work examining how social and affective factors affect learning and belief updating and it may be the future research can explore interconnections between the two in a broader neural network sensitive to these factors.

Above I noted that work examining the function of the LPC will be important to making sense of the complicated pattern of findings that have been reported for this component. However, it is important to note that drawing conclusions from a review of the literature like that presented in Chapter 2 is complicated by a number of methodological and statistical issues (Meehl, 1990; Morey & Lakens, 2017). One significant challenge is that both lower power (Szucs & Ioannidis, 2017) and research practices that inflate the Type I error rate (Luck & Gaspelin, 2017; Simmons et al., 2011) are quite common in the ERP literature. Taken together, these factors lead to a high rate of spurious findings (Button et al., 2013; Colquhoun, 2014; Ioannidis, 2005). It is obviously difficult to interpret a pattern of results when it is likely that a high proportion of them are not reliable.

Chapter 3 presented work to address one aspect of this problem: how to correct for the significant issue of multiple comparisons across time and space in ERP data while maintaining sufficient power to detect effects and have confidence in effects we do detect. Results from simulation studies with realistic EEG/ERP data showed that permutation-based mass univariate statistics—in particularly the cluster mass based test—provide the best balance of Type I error rate and power, at least for effects distributed relatively widely across space and time (i.e., most later ERP effects). In addition, this chapter presents new software to conduct these tests for factorial designs and interaction effects that cannot best tested with currently available software. Taken together, this software and simulation results address the two main barriers to widespread adoption of these techniques.

Appendix

Functional MRI activation tables

Table A1: Other-Neutral versus Fixation (whole brain)

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Supplementary motor area	--	0.0006	5.82	0, 10, 58	p(FWE) < 0.0001, k = 565
Inferior frontal gyrus (pars triangularis)	L	0.0301	4.89	-50, 24, 4	p(FWE) < 0.0001, k = 4426
Middle temporal cortex (anterior)	L	0.0011	5.67	-52, 4, -20	
Middle temporal cortex (pole)	L	0.0745	4.65	-40, 18, -32	
Precentral gyrus	L	0.0039	5.39	-38, 0, 50	p(FWE) = 0.0001, k = 860
Occipital cortex (calcarine)	R	0.0515	4.75	18, -66, 10	p(FWE) < 0.0001, k = 1890

Table A2: Other-Positive versus Fixation (whole brain)

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Inferior frontal gyrus (pars triangularis)	L	0.0540	4.80	-54, 22, 6	*p*(FWE) < 0.0001, k = 4370
Middle temporal cortex (anterior)	L	0.0001	6.21	-54, -4, -16	
Occipital cortex (lingual)	L	0.0161	5.11	-14, -30, -4	*p*(FWE) = 0.0437, k = 228
Supplementary motor area	L	0.0165	5.11	-4, 12, 56	*p*(FWE) = 0.0009, k = 516
Occipital cortex (lateral)	L	0.0324	4.94	-16, -94, -8	*p*(FWE) < 0.0001, k = 4675
Precentral gyrus	L	0.0771	4.71	-38, -2, 54	*p*(FWE) < 0.0001, k = 969
Middle temporal cortex (pole)	R	0.0813	4.69	56, 6, -16	*p*(FWE) < 0.0001, k = 773

Table A3: Other-Negative versus Fixation (whole brain)

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Inferior frontal gyrus (pars opercularis)	L	0.0113	5.09	-56, 18, 18	*p*(FWE) < 0.0001, k = 5760
Inferior frontal gyrus (pars triangularis)	L	0.0090	5.14	-58, 20, 24	
Inferior frontal gyrus (pars orbitalis)	L	0.0038	5.34	-46, 24, -6	
Superior temporal cortex (pole)	L	0.0161	5.00	-46, 8, -22	
Middle temporal cortex (posterior)	L	0.0639	4.64	-56, -48, 14	
Middle temporal cortex (anterior)	L	0.0003	5.89	-54, -26, -4	
Middle temporal cortex (anterior)	R	0.0449	4.73	62, 0, -16	*p*(FWE) < 0.0001, k = 1906
Middle temporal cortex (pole)	R	0.0186	4.96	44, 18, -30	
Supplementary motor area	L	0.0782	4.58	-2, 10, 62	*p*(FWE) = 0.0003, k = 921
Superior frontal cortex (medial)	L	0.0328	4.82	-8, 60, 42	
Precentral gyrus	L	0.0358	4.79	-38, -2, 56	*p*(FWE) = 0.0008, k = 781
Occipital cortex (calcarine)	R	0.0568	4.67	20, -90, -2	*p*(FWE) < 0.0001, k = 2113
Occipital cortex (lateral)	L	0.0676	4.62	-34, -92, -4	*p*(FWE) < 0.0001, k = 2481

Table A4: Self-Neutral versus Fixation (whole brain)

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Inferior frontal gyrus (pars opercularis)	L	0.0030	5.49	-52, 18, 20	*p*(FWE) < 0.0001, k = 4967
Inferior frontal gyrus (pars triangularis)	L	0.0098	5.22	-54, 22, 2	
Inferior frontal gyrus (pars orbitalis)	L	0.0207	5.04	-48, 24, -8	
Middle temporal cortex (posterior)	L	0.0459	4.83	-56, -44, 4	
Middle temporal cortex (anterior)	L	0.0003	5.94	-54, -4, -16	
Precentral gyrus	L	0.0129	5.15	-38, 0, 50	*p*(FWE) < 0.0001, k = 1104
Occipital cortex (lingual)	L	0.0553	4.78	-20, -88, -14	*p*(FWE) < 0.0001, k = 2460
Occipital cortex (lateral)	L	0.0217	5.03	-24, -104, -2	
Occipital cortex (calcarine)	R	0.0852	4.67	20, -90, -2	*p*(FWE) < 0.0001, k = 1876
Occipital cortex (lateral)	R	0.0251	4.99	40, -86, -8	
Superior frontal cortex (lateral)	L	0.0547	4.79	-14, 32, 64	*p*(FWE) = 0.0028, k = 444

Table A5: Self-Positive versus Fixation (whole brain)

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Inferior frontal gyrus (pars triangularis)	L	0.0228	4.98	-52, 22, 0	*p*(FWE) < 0.0001, k = 5426
Inferior frontal gyrus (pars orbitalis)	L	0.0192	5.02	-50, 24, -4	
Superior temporal cortex (pole)	L	0.0728	4.67	-44, 10, -22	
Middle temporal cortex (posterior)	L	0.0031	5.45	-54, -30, -2	
Middle temporal cortex (anterior)	L	0.0022	5.53	-52, -4, -18	
Precentral gyrus	L	0.0026	5.49	-44, -16, 62	*p*(FWE) < 0.0001, k = 1442
Supplementary motor area	R	0.0158	5.07	2, 10, 60	*p*(FWE) < 0.0001, k = 1091
Superior frontal cortex (medial)	L	0.0331	4.88	-6, 64, 28	
Occipital cortex (calcarine)	R	0.0323	4.89	18, -90, 2	*p*(FWE) < 0.0001, k = 5228
Occipital cortex (lingual)	L	0.0694	4.69	-20, -90, -14	
Occipital cortex (lateral)	R	0.0608	4.72	40, -86, -8	
Superior temporal cortex (pole)	R	0.0696	4.69	42, 12, -24	*p*(FWE) < 0.0001, k = 1028

Table A6: Self-Negative versus Fixation (whole brain)

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Inferior frontal gyrus (pars opercularis)	L	0.0938	4.64	-50, 18, 20	*p*(FWE) < 0.0001, k = 5855
Inferior frontal gyrus (pars orbitalis)	L	0.0025	5.54	-46, 24, -4	
Superior temporal cortex (pole)	L	0.0846	4.67	-46, 24, -14	
Middle temporal cortex (anterior)	L	0.0000	6.78	-52, -8, -18	
Middle temporal cortex (pole)	L	0.0022	5.57	-44, 12, -30	
Inferior temporal cortex (anterior)	L	0.0796	4.68	-44, -8, -30	
Precentral gyrus	L	0.0002	5.97	-40, 0, 58	*p*(FWE) < 0.0001, k = 929
Superior temporal cortex (pole)	R	0.0052	5.37	46, 12, -22	*p*(FWE) < 0.0001, k = 1419
Middle temporal cortex (anterior)	R	0.0007	5.79	54, 4, -18	
Middle temporal cortex (pole)	R	0.0756	4.70	48, 18, -34	
Supplementary motor area	L	0.0223	5.02	-4, 10, 56	*p*(FWE) < 0.0001, k = 1419
Superior frontal cortex (medial)	L	0.0202	5.04	-10, 56, 46	
Occipital cortex (lingual)	L	0.0349	4.91	-20, -90, -12	*p*(FWE) < 0.0001, k = 2698
Occipital cortex (lateral)	L	0.0226	5.02	-16, -96, -8	
Occipital cortex (calcarine)	R	0.0817	4.68	20, -88, 0	*p*(FWE) < 0.0001, k = 2117
Occipital cortex (lateral)	R	0.0312	4.93	42, -86, -8	

Table A7: Each Condition versus Fixation in the mPFC ROI

Region	R/L	Peak voxel p-value	z-score	MNI (x, y, z)	Cluster level p-value (FWE)
Self Relevant Positive – Fixation					
Superior frontal cortex (lateral)	L	0.0266	4.34	-10, 60, 38	*p*(FWE) < 0.0001, k = 727
Supplementary motor area	L	0.0487	4.16	-8, 18, 68	
Superior frontal cortex (medial)	L	0.0036	4.88	-6, 64, 28	
Self Relevant Neutral – Fixation					
Superior frontal cortex (lateral)	L	0.0059	4.79	-14, 32, 64	*p*(FWE) = 0.0003, k = 444
Superior frontal cortex (medial)	L	0.0187	4.48	-10, 64, 26	
Self Relevant Negative – Fixation					
Superior frontal cortex (lateral)	L	0.0021	5.04	-10, 56, 46	*p*(FWE) < 0.0001, k = 970
Supplementary motor area	L	0.0506	4.18	-4, 18, 70	
Superior frontal cortex (medial)	L	0.0063	4.77	-8, 60, 26	
Other Relevant Positive – Fixation					
Supplementary motor area	L	0.0122	4.61	-8, 18, 66	*p*(FWE) = 0.0165, k = 157
Superior frontal cortex (lateral)	L	0.0202	4.47	-12, 58, 42	*p*(FWE) = 0.0359, k = 115
Other Relevant Neutral – Fixation					
Superior frontal cortex (medial)	L	0.0523	4.11	-10, 64, 30	*p*(FWE) = 0.0433, k = 123
Other Relevant Negative – Fixation					
Superior frontal cortex (lateral)	L	0.0857	3.89	-12, 36, 62	*p*(FWE) = 0.0006, k = 562
Supplementary motor area	L	0.0111	4.51	-8, 24, 70	
Superior frontal cortex (medial)	L	0.0035	4.82	-8, 60, 42	

References

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science, 8*(1), 12-15.

Abraham, A. (2013). The world according to me: Personal relevance and the medial prefrontal cortex. *Frontiers in Human Neuroscience, 7*(341).

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. Dunning & J. Krueger (Eds.), *The Self and Social Judgement* (pp. 85-106). New York: Psychology Press.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277.

Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences, 58*(3), 626-639.

Anderson, M. J., & Ter Braak, C. J. F. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation, 73*(2), 85-113.

Araujo, H. F., Kaplan, J., & Damasio, A. (2013). Cortical midline structures and autobiographical-self processes: An activation-likelihood estimation meta-analysis. *Frontiers in Human Neuroscience, 7*(548).

Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgement* (pp. 334-347). New York: Cambridge University Press.

Bailey, K., & Chapman, P. (2012). When can we choose to forget? An ERP study into item-method directed forgetting of emotional words. *Brain and Cognition, 78*(2), 133-147.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*(6), 543-554.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459.

Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science, 12*(3), 197-204.

Bartholow, B. D., Pearson, M. A., Gratton, G., & Fabiani, M. (2003). Effects of alcohol on person perception: A social cognitive neuroscience approach. *Journal of Personality and Social Psychology, 85*(4), 627-638.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.

Bayer, M., Sommer, W., & Schacht, A. (2010). Reading emotional words within sentences: The impact of arousal and valence on event-related potentials. *International Journal of Psychophysiology, 78*, 299-307.

Bayer, M., Sommer, W., & Schacht, A. (2012). P1 and beyond: Functional separation of multiple emotion effects in word recognition. *Psychophysiology, 49*(7), 959-969.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York: Guilford Press.

Beer, J. S. (2014). Exaggerated positivity in self-evaluation: A social neuroscience approach to reconciling the role of self-esteem protection and cognitive bias. *Social and Personality Psychology Compass, 8*(10), 583-594.

Beer, J. S., & Flagan, T. (2015). More than the medial prefrontal cortex (MPFC): New advances in understanding the neural foundations of self-insight. In G. H. E. Gendolla, M. Tops & S. L. Koole (Eds.), *Handbook of Biobehavioral Approaches to Self-Regulation* (pp. 209-220). New York: Springer.

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of *P* values and evidence. *Journal of the American Statistical Association, 82*(397), 112-122.

Bernat, E., Bunce, S., & Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *International Journal of Psychophysiology, 42*(1), 11-34.

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology, 30*(5), 518-524.

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In J. T. Cacciopo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 581-607). New York: Cambridge University Press.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2014). Emotion word recognition: Discrete information effects first, continuous later? *Brain Research, 1564*, 62-71.

Briggs, K. E., & Martin, F. H. (2009). Affective picture processing and motivational relevance: Arousal and valence effects on ERPs in an oddball task. *International Journal of Psychophysiology, 72*(3), 299-306.

Brooks, J. L., Zoumpoulaki, A., & Bowman, H. (2017). Data-driven region-of-interest selection without inflating Type I error rate. *Psychophysiology, 54*(1), 100-113.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives: Pronouns modulate perspective taking during narrative comprehension. *Psychological Science, 20*(1), 27-32.

Brunyé, T. T., Ditman, T., Mahoney, C. R., & Taylor, H. A. (2011). Better you than I: Perspectives and emotion simulation during narrative comprehension. *Journal of Cognitive Psychology, 23*(5), 659-666.

Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of Google Books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology, 2*.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*(3), 531-544.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11*(2), 49-57.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging, 18*(1), 32-42.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376.

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience, 13*(6), 407-420.

Cacioppo, J. T., Crites, S. L., Berntson, G. G., & Coles, M. G. H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science, 4*(2), 108-112.

Cacioppo, J. T., Crites, S. L., Gardner, W. L., & Berntson, G. G. (1994). Bioelectrical echoes from evaluative categorizations: I. A late positive brain potential that varies as a function of trait negativity and extremity. *Journal of Personality and Social Psychology, 67*(1), 115-125.

Carretie, L., Hinojosa, J. A., Albert, J., Lopez-Martin, S., De La Gandara, B. S., Igoa, J. M., et al. (2008). Modulation of ongoing cognitive processes by emotionally intense words. *Psychophysiology, 45*(2), 188-196.

Chavez, R. S., & Heatherton, T. F. (2015). Multimodal frontostriatal connectivity underlies individual differences in self-esteem. *Social Cognitive and Affective Neuroscience, 10*(3), 364-370.

Chavez, R. S., & Heatherton, T. F. (2017). Structural integrity of frontostriatal connections predicts longitudinal changes in self-esteem. *Social Neuroscience, 12*(3), 280-286.

Chavez, R. S., Heatherton, T. F., & Wagner, D. D. (in press). Neural population decoding reveals the intrinsic positivity of the self. *Cerebral Cortex*.

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language, 122*(3), 211-226.

Citron, F. M. M., Weekes, B. S., & Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: Time course and ERP correlates. *Neuroscience Letters, 533*, 90-95.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204.

Clegg, L., Swain, A., Fields, E. C., Delaney-Busch, N., Holt, D. J., & Kuperberg, G. R. (2010). *The construction of valence: ERPs to neutral words in discourse modulated by emotional content of previous word or sentence*. Poster presented at the 17th Annual Meeting of the Cognitive Neuroscience Society, Montreal, QC.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of *p*-values. *Royal Society Open Science, 1*(140216).

Compton, R. J. (2003). The interface between emotion and attention: A review of evidence from psychology and neuroscience. *Behavioral and Cognitive Neuroscience Reviews, 2*(2), 115-129.

Coulson, S., King, J. W., & Kutas, M. (1998). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13*(6), 653-672.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences, 10*(7), 294-300.

Crites, S. L., Cacioppo, J. T., Gardner, W. L., & Bernston, G. G. (1995). Biolectrical echoes from evaluative categorization: II. A late positive brain potentail that varies as function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology, 68*(6), 997-1013.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology, 52*(2), 95-111.

D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: The valuation hypothesis. *Frontiers in Human Neuroscience, 7*(372).

Dayan, P., & Yu, A. J. (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network: Computation in Neural Systems, 17*(4), 335-350.

De Swart, J. H., Kok, A., & Das-Smaal, E. A. (1981). P300 and uncertainty reduction in a concept-identification task. *Psychophysiology, 18*(6), 619-629.

Delaney-Busch, N., & Kuperberg, G. R. (2013). Friendly drug-dealers and terrifying puppies: Affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive Affective & Behavioral Neuroscience, 13*(3), 473-490.

Delaney-Busch, N., Wilkie, G., & Kuperberg, G. R. (2016). Vivid: How valence and arousal influence word processing under different task demands. *Cognitive Affective & Behavioral Neuroscience, 16*(3), 415-432.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Delplanque, S., Lavoie, M. E., Hot, P., Silvert, L., & Sequeira, H. (2004). Modulation of cognitive processing by emotional valence studied through event-related potentials in humans. *Neuroscience Letters, 356*(1), 1-4.

Delplanque, S., Silvert, L., Hot, P., Rigoulot, S., & Sequeira, H. (2006). Arousal and valence effects on event-related P3a and P3b during emotional categorization. *International Journal of Psychophysiology, 60*(3), 315-322.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A Meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience, 24*(8), 1742-1752.

Dillon, D. G., Cooper, J. J., Grent-'t-Jong, T., Woldorff, M. G., & LaBar, K. S. (2006). Dissociation of event-related potentials indexing arousal and semantic cohesion during emotional word encoding. *Brain and Cognition, 62*(1), 43-57.

Ding, J., Wang, L., & Yang, Y. (2014). The dynamic influence of emotional words on sentence processing. *Cognitive, Affective, & Behavioral Neuroscience*.

Donchin, E. (1981). Surprise! . . . Surprise? *Psychophysiology, 18*(5), 493-513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(3), 357-374.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: Variation of event-related potentials with subjective probability. *Psychophysiology, 14*(5), 456-467.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69-106.

Durda, K., & Buchanan, L. (2006). WordMine2. Retrieved March 11, 2016, from http://web2.uwindsor.ca/wordmine/

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 224-240.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology, 23*(3), 298-308.

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a Von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology, 75*(2), 22-35.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation, 51*, 1-44.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*.

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science, 12*(1), 46-61.

Fields, E. C. (2015). *Event-related potential studies of the effects of mood, self-relevance, and task on the processing of emotional words in social vignettes* (Master's Thesis). Tufts University, Medford, MA.

Fields, E. C., & Kuperberg, G. R. (2012). It's all about you: An ERP study of emotion and self-relevance in discourse. *NeuroImage, 62*(1), 562-574.

Fields, E. C., & Kuperberg, G. R. (2015). Loving yourself more than your neighbor: ERPs reveal online effects of a self-positivity bias. *Social Cognitive and Affective Neuroscience, 10*(9), 1202-1209.

Fields, E. C., & Kuperberg, G. R. (2016). Dynamic effects of self-relevance and task on neural processing of emotional words in context. *Frontiers in Psychology, 6*(2003).

Fischler, I., & Bradley, M. (2006). Event-related potential studies of language and emotion: Words, phrases, and task effects. *Progress in Brain Research, 156*, 185-203.

Flagan, T., & Beer, J. S. (2013). Three ways in which midline regions contribute to self-evaluation. *Frontiers in Human Neuroscience, 7*(450).

Fogel, A. R., Midgley, K. J., Delaney-Busch, N., & Holcomb, P. J. (2012). F#©K! Processing emotion and taboo in a native vs. a second language: An ERP study. *Psychophysiology, 49*(Suppl. 1), S41.

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., et al. (2003). In search of the emotional self: An fMRI study using positive and negative emotional words. *American Journal of Psychiatry, 160*(11), 1938-1945.

Fossati, P., Hevenor, S. J., Lepage, M., Graham, S. J., Grady, C., Keightley, M. L., et al. (2004). Distributed self in episodic memory: Neural correlates of successful retrieval of self-encoded positive and negative personality traits. *NeuroImage, 22*(4), 1596-1604.

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology, 46*(3), 521-530.

Frijda, N. H. (1986). *The Emotions*. New York: Cambridge University Press.

Frühholz, S., Jellinghaus, A., & Herrmann, M. (2011). Time course of implicit processing and explicit processing of emotional faces and emotional words. *Biological Psychology, 87*(2), 265-274.

Gable, P. A., Adams, D. L., & Proudfit, G. H. (2015). Transient tasks and enduring emotions: The impacts of affective content, task relevance, and picture duration on the sustained late positive potential. *Cognitive, Affective, & Behavioral Neuroscience, 15*(1), 45-54.

Garrett, N., Sharot, T., Faulkner, P., Korn, C. W., Roiser, J. P., & Dolan, R. J. (2014). Losing the rose tinted glasses: Neural substrates of unbiased belief updating in depression. *Frontiers in Human Neuroscience, 8*(639).

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*, 587-606.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The Empire of Chance: How Probability Changed Science and Everyday Life*. New York: Cambridge University Press.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*(1), 76-97.

Goldin, P. R., Jazaieri, H., Ziv, M., Kraemer, H., Heimberg, R. G., & Gross, J. J. (2013). Changes in positive self-views mediate the effect of cognitive-behavioral therapy for social anxiety disorder. *Clinical Psychological Science, 1*(3), 301-310.

González-Villar, A. J., Triñanes, Y., Zurrón, M., & Carrillo-de-la-Peña, M. T. (2014). Brain processing of task-relevant and task-irrelevant emotional words: An ERP study. *Cognitive, Affective, & Behavioral, Neuroscience, 14*(3), 939-950.

Good, P. I. (2005). *Permutation, parametric and bootstrap tests of hypotheses* (3rd ed.). New York, NY: Springer.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass, 3*(1), 128-156.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (2016). Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*(12), 1711-1725.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology, 48*(12), 1726-1737.

Gu, R. L., Ge, Y., Jiang, Y., & Luo, Y. J. (2010). Anxiety and outcome evaluation: The good, the bad and the ambiguous. *Biological Psychology, 85*(2), 200-206.

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology, 35*(2), 129-155.

Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (2012). ERPs and the study of emotion. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 441-474). New York: Oxford University Press.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research - Online, 7*(1), 1-20.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5*(9), 394-400.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What If There Were No Significance Tests?* Hillsdale, NJ: Erlbaum.

Herbert, C., Herbert, B. M., Ethofer, T., & Pauli, P. (2011). His or mine? The time course of self-other discrimination in emotion processing. *Social Neuroscience, 6*(3), 277-288.

Herbert, C., Herbert, B. M., & Pauli, P. (2011). Emotional self-reference: Brain structures involved in the processing of words describing one's own emotions. *Neuropsychologia, 49*(10), 2947-2956.

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology, 45*(3), 487-498.

Herbert, C., Kissler, J., Junghofer, M., Peyk, P., & Rockstroh, B. (2006). Processing of emotional adjectives: Evidence from startle EMG and ERPs. *Psychophysiology, 43*(2), 197-206.

Herbert, C., Pauli, P., & Herbert, B. M. (2011). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Social Cognitive and Affective Neuroscience, 6*(5), 653-661.

Hinojosa, J. A., Carretie, L., Mendez-Bertolo, C., Miguez, A., & Pozo, M. A. (2009). Arousal contributions to affective priming: Electrophysiological correlates. *Emotion, 9*(2), 164-171.

Hinojosa, J. A., Carretie, L., Valcarcel, M. A., Mendez-Bertolo, C., & Pozo, M. A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective, & Behavioral Neuroscience, 9*(2), 173-189.

Hinojosa, J. A., Mendez-Bertolo, C., & Pozo, M. A. (2010). Looking at emotional words is not the same as reading emotional words: Behavioral and neural correlates. *Psychophysiology, 47*(4), 748-757.

Hirsh, J. B., & Inzlicht, M. (2008). The devil you know: Neuroticism predicts neural response to uncertainty. *Psychological Science, 19*(10), 962-967.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L.-H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitve, Affective, & Behavioral Neuroscience, 9*(4), 389-397.

Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience, 21*(11), 2245-2262.

Hubbard, R., & Bayarri, M. J. (2003). Confusion over measures of evidence (*p*'s) versus errors (α's) in classical statistical testing. *American Statistician, 57*(3), 171-178.

Hubbard, R., & Lindsay, R. M. (2008). Why *P* values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology, 18*(1), 69-88.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), 696-701.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*(4), 887-900.

JASP Team. (2017). JASP (Version 0.8.1.0) [Computer software].

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524-532.

Johnson, D. D. P., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature, 477*(7364), 317-320.

Johnson, R., Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*(4), 367-384.

Johnson, R., Jr., & Donchin, E. (1985). Second thoughts: Multiple P300s elicited by a single stimulus. *Psychophysiology, 22*(2), 182-194.

Kaltwasser, L., Ries, S., Sommer, W., Knight, R. T., & Willems, R. M. (2013). Independence of valence and reward in emotional word processing: Electrophysiological evidence. *Frontiers in Psychology, 4*.

Kamp, S. M., Brumback, T., & Donchin, E. (2013). The component structure of ERP subsequent memory effects in the Von Restorff paradigm and the word frequency effect in recall. *Psychophysiology, 50*(11), 1079-1093.

Kamp, S. M., Potts, G. F., & Donchin, E. (2015). On the roles of distinctiveness and semantic expectancies in episodic encoding of emotional words. *Psychophysiology, 52*(12), 1599-1609.

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research, 1148*, 138-148.

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology, 47*(5), 888-904.

Kappenman, E. S., & Luck, S. J. (2012). ERP components: The ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford University Press.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the Von Restorff effect. *Cognitive Psychology, 16*(2), 177-216.

Keuper, K., Zwanzger, P., Nordt, M., Eden, A., Laeger, I., Zwitserlood, P., et al. (2014). How 'love' and 'hate' differ from 'sleep': Using combined electro/magnetoencephalographic data to reveal the sources of early cortical responses to emotional words. *Human Brain Mapping, 35*(3), 875-888.

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology, 124*(10), 2062-2063.

Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: Insights from ERP studies. *Progress in Brain Research, 156*, 147-183.

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biological Psychology, 80*(1), 75-83.

Knost, B., Flor, H., Braun, C., & Birbaumer, N. (1997). Cerebral processing of words and the development of chronic pain. *Psychophysiology, 34*(4), 474-481.

Kopp, B. (2008). The P300 component of the event-related brain potential and Bayes' theorem. In M. Sun (Ed.), *Cognitive Sciences at the Leading Edge*. New York: Nova Science Publishers.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151*, 107-118.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass, 5*(3), 108-125.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12*(5), 535-540.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32-59.

Kuperberg, G. R., Kreher, D. A., Swain, A., Goff, D. C., & Holt, D. J. (2011). Selective emotional processing deficits to social vignettes in schizophrenia: An ERP study. *Schizophrenia Bulletin, 37*(1), 148-163.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Laney, C., Campbell, H. V., Heuer, F., & Reisberg, D. (2004). Memory for thematically arousing events. *Memory & Cognition, 32*(7), 1149-1159.

Lau, E. F., Gramfort, A., Hamalainen, M. S., & Kuperberg, G. R. (2013). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *Journal of Neuroscience, 33*(43), 17174-17181.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920-933.

Lee, K. H., & Siegle, G. J. (2012). Common and distinct brain networks underlying explicit emotional evaluation: a meta-analytic study. *Social Cognitive and Affective Neuroscience, 7*(5), 521-534.

Legrand, D., & Ruby, P. (2009). What Is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review, 116*(1), 252-282.

Li, W., & Han, S. (2010). Perspective taking modulates event-related potentials to perceived pain. *Neuroscience Letters, 469*(3), 328-332.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121-143.

Liu, B. L., Jin, Z. X., Wang, Z. N., & Hu, Y. (2010). The interaction between pictures and words: Evidence from positivity offset and negativity bias. *Experimental Brain Research, 201*(2), 141-153.

Liu, Y. L., Huang, H. Q., McGinnis-Deweese, M., Keil, A., & Ding, M. Z. (2012). Neural substrate of the late positive potential in emotional processing. *Journal of Neuroscience, 32*(42), 14563-14572.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event related potentials. *Frontiers in Human Neuroscience, 8*(213).

Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 329-327). New York: Oxford University Press.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (2nd ed.). Cambridge, MA: The MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157.

Luck, S. J., & Kappenman, E. S. (2012). ERP components and selective attention. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 295-328). New York: Oxford University Press.

MacNamara, A., Foti, D., & Hajcak, G. (2009). Tell me about it: Neural activity elicited by emotional pictures and preceding descriptions. *Emotion, 9*(4), 531-543.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*(2), 171-186.

Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.). London, UK: Chapman & Hall.

Mardaga, S., & Hansenne, M. (2009). Personality modulation of P300 wave recorded within an emotional oddball protocol. *Neurophysiologie Clinique/Clinical Neurophysiology, 39*(1), 41-48.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of *p* values just below .05. *Quarterly Journal of Experimental Psychology, 65*(11), 2271-2279.

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*(2), 114-133.

Matsuda, I., & Nittono, H. (2015). Motivational significance and cognitive effort elicit different late positive potentials. *Clinical Neurophysiology, 126*(2), 304-313.

Mazaika, P. K., Hoeft, F., Glover, G. H., & Reiss, A. L. (2009). *Methods and software for fMRI analysis of clinical subjects*. Poster presented at the 15th Annual Meeting of the Organization for Human Brain Mapping, San Francisco, CA.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*(1), 195-244.

Metzger, L. J., Orr, S. P., Lasko, N. B., McNally, R. J., & Pitman, R. K. (1997). Seeking the source of emotional stroop interference effects in PTSD: A study of P3s to traumatic words. *Integrative Physiological and Behavioral Science, 32*(1), 43-51.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711-747.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176-182.

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1309-1316.

Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., & Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *Journal of Cognitive Neuroscience, 18*(9), 1586-1594.

Morey, R. D., & Lakens, D. (2017). Why most of psychology is statistically unfalsifiable. [Electronic Version]. *Manuscript in preparation*. Retrieved 15 March 2017, from https://github.com/richarddmorey/psychology_resolution/blob/master/paper/response.pdf

Naumann, E., Bartussek, D., Diedrich, O., & Laufer, M. E. (1992). Assessing cognitive and affective information-processing functions of the brain by means of the late positive complex of the event-related potential. *Journal of Psychophysiology, 6*(4), 285-298.

Naumann, E., Maier, S., Diedrich, O., Becker, G., & Bartussek, D. (1997). Structural, semantic, and emotion-focussed processing of neutral and negative nouns: Event-related potential correlates. *Journal of Psychophysiology, 11*(2), 158-172.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*(2), 241-301.

Nieuwenhuis, S. (2011). Learning, the P3, and the locus coeruleus-norepinephrine system. In R. B. Mars, J. Sallet, M. F. S. Rushworth & N. Yeung (Eds.), *Neural Basis of Motivational and Cognitive Control* (pp. 209-222). Cambridge, MA: MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510-532.

Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences, 8*(3), 102-107.

Northoff, G., Heinzel, A., Greck, M., Bennpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain: A meta-analysis of imaging studies on the self. *NeuroImage, 31*(1), 440-457.

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., et al. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience, 16*(10), 1746-1772.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology, 77*(3), 247-265.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).

Osgood, C. E., Suci, G. J., & Tannenbaum, P. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.

Otten, L. J., & Rugg, M. D. (2001). Electrophysiological correlates of memory encoding are task-dependent. *Cognitive Brain Research, 12*(1), 11-18.

Palazova, M., Mantwill, K., Sommer, W., & Schacht, A. (2011). Are effects of emotion in single words non-lexical? Evidence from event-related brain potentials. *Neuropsychologia, 49*(9), 2766-2775.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184-188.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*(6), 531-536.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528-530.

Pauli, P., Amrhein, C., Muhlberger, A., Dengler, W., & Wiedemann, G. (2005). Electrocortical evidence for an early abnormal processing of panic-related words in panic disorder patients. *International Journal of Psychophysiology, 57*(1), 33-41.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*(10).

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*(3), 302-321.

Peyk, P., Schupp, H. T., Elbert, T., & Junghofer, M. (2008). Emotion processing in the visual brain: A MEG analysis. *Brain Topography, 20*(4), 205-215.

Phan, K. L., Taylor, S. F., Welsh, R. C., Ho, S. H., Britton, J. C., & Liberzon, I. (2004). Neural correlates of individual ratings of emotional salience: A trial-related fMRI study. *NeuroImage, 21*(2), 768-780.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage, 16*(2), 331-348.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148.

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 159-188). New York: Oxford University Press.

Polich, J., Alexander, J. E., Bauer, L. O., Kuperman, S., Morzorati, S., O'Connor, S. J., et al. (1997). P300 topography of amplitude/latency correlations. *Brain Topography, 9*(4), 275-282.

Ponz, A., Montant, M., Liegeois-Chauvel, C., Silva, C., Braun, M., Jacobs, A. M., et al. (2014). Emotion processing in words: A test of the neural re-use hypothesis using surface and intracranial EEG. *Social Cognitive and Affective Neuroscience, 9*(5), 619-627.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology, 3*.

Qin, P. M., Duncan, N., & Northoff, G. (2013). Why and how is the self-related to the brain midline regions? *Frontiers in Human Neuroscience, 7*(909).

Recio, G., Conrad, M., Hansen, L. B., & Jacobs, A. M. (2014). On pleasure and thrill: The interplay between arousal and valence during visual word recognition. *Brain and Language, 134*, 34-43.

Rellecke, J., Palazova, M., Sommer, W., & Schacht, A. (2011). On the automaticity of emotion processing in words and faces: Event-related brain potentials evidence from a superficial task. *Brain and Cognition, 77*(1), 23-32.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20*(5), 343-350.

Rouder, J. N., Speckman, P. L., Sun, D. C., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296-320.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

Sabatinelli, D., Keil, A., Frank, D. W., & Lang, P. J. (2013). Emotional perception: Correspondence of early and late event-related potentials with cortical and subcortical functional MRI. *Biological Psychology, 92*(3), 513-519.

Sabatinelli, D., Lang, P. J., Keil, A., & Bradley, M. M. (2007). Emotional perception: Correlation of functional MRI and event-related potentials. *Cerebral Cortex, 17*(5), 1085-1091.

Sass, S. M., Heller, W., Stewart, J. L., Silton, R. L., Edgar, J. C., Fisher, J. E., et al. (2010). Time course of attentional bias in anxiety: Emotion and gender specificity. *Psychophysiology, 47*(2), 247-259.

Saxe, R. (2009). Theory of Mind. In W. P. Banks (Ed.), *Encyclopedia of Consciousness* (pp. 401-409). Oxford: Academic Press.

Schacht, A., & Sommer, W. (2009a). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition, 69*(3), 538-550.

Schacht, A., & Sommer, W. (2009b). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 28-43.

Schapkin, S. A., Gusev, A. N., & Kuhl, J. (2000). Categorization of unilaterally presented emotional words: An ERP analysis. *Acta Neurobiologiae Experimentalis, 60*(1), 17-28.

Scherer, K. R. (2000). Psychological models of emotion. In J. Borod (Ed.), *The Neuropsychology of Emotion* (pp. 137-162). New York: Oxford University Press.

Schindler, S., Wegrzyn, M., Steppacher, I., & Kissler, J. (2014). Its all in your head: How anticipating evaluation affects the processing of emotional trait adjectives. *Frontiers in Psychology, 5*(1292), 1-10.

Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology, 80*(1), 95-104.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of *p* values for testing precise null hypotheses. *The American Statistician, 55*(1), 62-71.

Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in Cognitive Sciences, 20*(1), 25-33.

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience, 14*(11), 1475-1479.

Shestyuk, A. Y., & Deldin, P. J. (2010). Automatic and strategic representation of the self in major depression: Trait and state abnormalities. *American Journal of Psychiatry, 167*(5), 536-544.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-Curve and effect size: Correcting for publication bia using only significant results. *Perspectives on Psychological Science, 9*(6), 666-681.

Sokolov, E. N., Spinks, J. A., Näätänen, R., & Lyytinen, H. (2002). *The Orienting Response in Information Processing*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Research, 48*(12), 1391-1408.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21*(3), 489-510.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). Effect of stimulus sequence on waveform of cortical event-related potential. *Science, 193*(4258), 1142-1146.

Srivastava, S. (2016). Everything is fucked: The syllabus. *The Hardest Science*. Retrieved December 13, 2016, from https://hardsci.wordpress.com/2016/08/11/everything-is-fucked-the-syllabus/

Still, A. W., & White, A. P. (1981). The approximate randomization rest as an alternative to the *F* test in analysis of variance. *British Journal of Mathematical and Statistical Psychology, 34*(2), 243-252.

Swann, W. B., Jr. (2011). Self-verification theory. In P. A. M. V. Lange, A. W. Kruglanski & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (Vol. 2, pp. 23-42). Thousands Oak, CA: SAGE Publications.

Swann, W. B., Jr., & Bosson, J. (2010). Self and identity. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., pp. 589-628). New York: McGraw Hill.

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS biology, 15*(3), e2000797.

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology, 52*(8), 997-1009.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization minimization hypothesis. *Psychological Bulletin, 110*(1), 67-85.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210.

Tempel, K., Kuchinke, L., Urton, K., Schlochtermeier, L. H., Kappelhoff, H., & Jacobs, A. M. (2013). Effects of positive pictograms and words: An emotional word superiority effect? *Journal of Neurolinguistics, 26*(6), 637-648.

Thomas, S. J., Johnstone, S. J., & Gonsalvez, C. J. (2007). Event-related potentials during an emotional Stroop task. *International Journal of Psychophysiology, 63*(3), 221-231.

Tritt, S. M., Peterson, J. B., & Inzlicht, M. (2012). *The uncertainty bias: Brain evidence that ambiguity is more captivating than threat or reward*. Poster presented at the Annual Meeting of the Social and Affective Neuroscience Society, New York, NY.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences, 11*(4), 153-157.

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science, 20*(9), 1092-1099.

van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods, 39*(4), 973-978.

Van Duynslaeger, M., Sterken, C., Van Overwalle, F., & Verstraeten, E. (2008). EEG components of spontaneous trait inferences. *Social Neuroscience, 3*(2), 164-177.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30*(3), 829-858.

Van Overwalle, F., Van den Eede, S., Baetens, K., & Vandekerckhove, M. (2009). Trait inferences in goal-directed behavior: ERP timing and localization under spontaneous and intentional processing. *Social Cognitive and Affective Neuroscience, 4*(2), 177-190.

Vanderploeg, R. D., Brown, W. S., & Marsh, J. T. (1987). Judgments of emotion in words and faces: ERP correlates. *International Journal of Psychophysiology, 5*(3), 193-205.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*(3), 274-290.

Wagner, A. D., Koutstaal, W., & Schacter, D. L. (1999). When encoding yields remembering: Insights from event-related neuroimaging. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 354*(1387), 1307-1324.

Wagner, D. D., Haxby, J. V., & Heatherton, T. F. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(4), 451-470.

Wagner, D. D., Kelley, W. M., Haxby, J. V., & Heatherton, T. F. (2016). The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *Journal of Neuroscience, 36*(26), 6917-6925.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191-1207.

Weinberg, A., & Hajcak, G. (2010). Beyond good and evil: The time-course of neural activity elicited by specific picture content. *Emotion, 10*(6), 767-782.

Weinberg, A., Hilgard, J., Bartholow, B. D., & Hajcak, G. (2012). Emotional targets: Evaluative categorization as a function of context and content. *International Journal of Psychophysiology, 84*(2), 149-154.

Weinstein, A. M. (1995). Visual ERPs evidence for enhanced processing of threatening information in anxious university students. *Biological Psychiatry, 37*(12), 847-858.

Welborn, B. L., & Lieberman, M. D. (2015). Person-specific theory of mind in medial pFC. *Journal of Cognitive Neuroscience, 27*(1), 1-12.

Wheldon, M. C., Anderson, M. J., & Johnson, B. W. (2007). Identifying treatment effects in multi-channel measurements in electroencephalographic studies: Multivariate permutation tests and multiple comparisons. *Australian & New Zealand Journal of Statistics, 49*(4), 397-413.

Wilding, E. L., & Ranganath, C. (2012). Electrophysiological correlates of episodic memory processes. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 373-396). New York: Oxford University Press.

Williamson, S., Harpur, T. J., & Hare, R. D. (1991). Abnormal processing of affective words by psychopaths. *Psychophysiology, 28*(3), 260-273.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage, 92*, 381-397.

Wlotko, E. W., & Kuperberg, G. R. (2017). *The brain dissociates between different levels of prediction violation during language comprehension...but not always*. Poster presented at the 30th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

Woolrich, M. W., Beckmann, C. F., Nichols, T. E., & Smith, S. M. (2009). Statistical analysis of fMRI data. In M. Filippi (Ed.), *fMRI Techniques and Protocols* (pp. 179-236). New York, NY: Humana Press.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681-692.

Zaki, J., & Ochsner, K. N. (2011). You, me, and my brain: Self and other representation in social cognitive neuroscience. In A. Todorov, S. T. Fiske & D. Prentice (Eds.), *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*. New York: Oxford University Press.

Zhang, M., & Guo, T. (2014). Event-related brain potentials differentiate three types of emotional words categorized from linguistic perspective. *Journal of Neurolinguistics, 31*, 17-27.