

An Open Science Electrophysiology Study of  
Anxiety and Error Monitoring

---

A Thesis  
Presented to  
The Division of Philosophy, Religion, Psychology, and Linguistics  
Reed College

---

In Partial Fulfillment  
of the Requirements for the Degree  
Bachelor of Arts

---

Melissa D. Lewis

May 2013



Approved for the Division  
(Psychology)

---

Michael Pitts



# Acknowledgements

Arianna, Natasha, Annelyse, Rose, and Juliet, the women with whom I've shared a home, victories, setbacks, and a lot of Thai food: from you I learned how to love fearlessly and how to persist.

Michael, my advisor: I first learned to be practically and optimistically critical of science from you, and I am grateful for how you have enriched my academic and intellectual experience at this school.

Reed College: you are an improbable assembly of administrators, faculty, staff, and students. I am humbled and galvanized by my time with you, and I will always work to give to others as you have given to me.

Everything I can say I've accomplished is a product of many people's magnanimity, long emails, Trader Joe's gift cards, copyediting, mackerel, wisdom, flowers, crushing hugs, jokes, stories, surprise visits, and cat pictures. You inspire me to be present, stubborn, and gracious. I hope my life reflects your benefaction.



# Table of Contents

<b>Introduction</b>	<b>1</b>
0.1 The Neural Basis of Error Monitoring and Defensive Motivation	1
0.1.1 Error Monitoring	1
0.1.2 Defensive Motivation	1
0.1.3 The ERN and Startle	2
0.1.4 State versus Trait Anxiety	2
0.2 Open Science	4
0.2.1 Historical Precedence	4
0.2.2 The Center for Open Science	5
<b>Chapter 1: The Replication</b>	<b>7</b>
1.1 Methods	7
1.1.1 Participants	7
1.1.2 The Task	7
1.1.3 Errors and EEG	8
1.1.4 Startle and EMG	8
1.2 Results	9
1.2.1 Startle	9
1.2.2 The Error-related Negativity	9
1.2.3 Error-potentiated Startle and the Predictive Value of the ERN	10
1.3 Discussion	10
<b>Chapter 2: The Extension</b>	<b>15</b>
2.1 Methods	15
2.2 Results	15
2.2.1 Anxiety	15
2.2.2 Anxiety and Electrophysiological Measures	16
2.3 Discussion	18
2.4 General Discussion	19
2.4.1 On the Practice of Replication	20
<b>Chapter 3: The Future of Open Science</b>	<b>21</b>
<b>Appendix A: Comparison Tables</b>	<b>25</b>

<b>Appendix B: Free and Open Source Software Used</b>	<b>27</b>
<b>Appendix C: Report to the Reproducibility Project</b>	<b>29</b>
C.1 Introduction	29
C.2 Methods	30
C.2.1 Power Analysis	30
C.2.2 Planned Sample	30
C.2.3 Materials	30
C.2.4 Procedure	31
C.2.5 Analysis Plan	31
C.2.6 Differences from Original Study	32
<b>References</b>	<b>33</b>



# List of Tables

1	Reasons for Retraction of 742 Papers, 2000 - 10. . . . .	4
2.1	Overall Startle Magnitude Effects . . . . .	16
2.2	Startle Magnitude ANOVAs by Trial Type . . . . .	17
A.1	Original and Replication Results . . . . .	25
B.1	Free (F) and Open Source (OS) Software Used. . . . .	27



# List of Figures

1	Mean event-related potentials time-locked to correct and erroneous responses in Hajcak 2012 . . . . .	3
1.1	Comparison of mean startle magnitude by trial type from original and replication study . . . . .	9
1.2	Comparison of mean event-related potentials time-locked to correct and erroneous responses between the original and replication studies . . . . .	12
1.3	Comparison of correlation of ERN amplitude and error-potentiated startle magnitude between original and replication studies . . . . .	13
1.4	Comparison of correlation of ERN amplitude and error-potentiated startle magnitude before and after median split . . . . .	13
1.5	Comparison of correlation of ERN amplitude and error-potentiated startle magnitude after median split, with and without outlier . . . . .	13
2.1	Mean startle magnitude ( $\mu V$ ) by trial type for low and high anxiety conditions. . . . .	17
2.2	Mean error startle magnitude ( $\mu V$ ) between low and high anxiety conditions, split by order . . . . .	18
2.3	Mean correct predictable startle magnitude ( $\mu V$ ) between low and high anxiety conditions, split by order . . . . .	19



# Abstract

Replicability is essential to scientific progress, but published research findings are less likely to be replicated every year. This experiment was both a replication of a previous research finding and an extension of its associated questions. The original experiment investigated whether an event-related potential called the error-related negativity (ERN) indexed motivational or emotional aspects of error monitoring. They did so by investigating whether it correlated to a physiological index of defensive motivation, the startle response to a loud auditory stimulus as measured by electromyography (EMG) at the orbicularis oculi muscle. This experiment replicated increased startle magnitude following errors compared to correct responses, suggesting a relationship between defensive motivation and error detection. This experiment also replicated the finding of an ERN with the same scalp distribution of the original study and in the same time frame (peaking roughly 50ms after response). However, it did not find a correlation between the magnitude of the startle response and the amplitude of the ERN, as the original study found. As an extension, this experiment manipulated state anxiety with a feedback system based on points. It also used the Spielberger State-Trait Anxiety Inventory to examine whether there was a relationship between self-reported state and/or trait anxiety and the ERN and/or startle response. The results showed that trial type modulated startle response. However, state anxiety could not be teased out for its effect on startle or the ERN, and there was no correlation between ERN and startle. No relationship between trait anxiety and the ERN or the startle response was found.



# Dedication

To Charleston Elementary, convergence of blossoming trees, tetherball, and teachers who said that making the world a better place is a perfectly viable career option.





# Introduction

## 0.1 The Neural Basis of Error Monitoring and Defensive Motivation

### 0.1.1 Error Monitoring

The mechanisms by which people monitor and correct for errors remain rich areas of investigation in neuroscience and psychology in general. Related to executive control, detection of errors during response monitoring – whatever its underlying mechanisms – consistently proves aversive ([Leue et al., 2012](#)). One prominent model of error monitoring, the integrative account, posits that the aversiveness arises from anticipation of greater cognitive demand ([Botvinick, 2007](#)). Another model, revised reinforcement sensitivity theory (rRST), describes the aversive nature of errors as originating in anticipation of negative consequences ([Corr, 2004](#)). Evidence from experiments manipulating task complexity points to a significant role of expectation of cognitive load ([Leue et al., 2012](#)). However, experiments manipulating incentives and performance import lend credence to rRST ([Weinberg et al., 2011](#); [Bush et al., 2000](#)). These two models seem not be mutually exclusive and error aversion may be supported by simultaneously viable mechanisms ([Bush et al., 2000](#)). One of the key features shared by both of these theories is the central role of defensive motivation in error monitoring and detection.

### 0.1.2 Defensive Motivation

In the context of error monitoring, defensive motivation is thought to arise as a response to a commission error as a means of improving diligence and mitigating future errors. Evidence for this comes from a variety of experimental manipulations, including reward incentives and punishments ([Leue et al., 2012](#); [Weinberg et al., 2011](#)). Additionally, insofar as defensive motivation is yoked to affective states like anxiety, evidence for a relationship between affective states and error comes from experiments measuring and manipulating state anxiety.

ety and also measuring the more stable trait anxiety (Hajcak, 2012). Measures of defensive motivation include skin conductance and startle response (Hajcak & Foti, 2008).

### 0.1.3 The ERN and Startle

The error-related negativity (ERN) is a particular type of event-related potential (ERP). ERPs reflect brain activity corresponding to specific sensory, cognitive, or motor events and are small signals embedded in larger ongoing electrical activity measured at the scalp via electroencephalography (EEG). During a recording of EEG, electrodes at the scalp are thought to be recording the voltage resulting from the summation of dipoles. Dipoles are spatially proximal pairs of positive and negative electrical charges originating in postsynaptic potentials (Luck, 2005). Because of the convoluted nature of the cortex it is unclear what relationship these charges have to the function or location of the neural region underlying them. It is clear, though, that this voltage fluctuation occurs as a consequence of the activity of postsynaptic potentials, because when excitatory neurotransmitters are released at apical dendrites of pyramidal cells in the cortex, i.e. presynaptic terminals, it causes positive ions to flow into the postsynaptic neuron. This leaves the charge outside of the cell relatively negative. Then current flows out of the cell body and basal dendrites, creating a net positivity.

The ERN is a negative-going ERP component peaking roughly 50-100 milliseconds after a commission error (Figure 1) and is usually found over the frontal-central scalp (Falkenstein et al., 1991). While it is often difficult to determine the source of an ERP component, source localization algorithms whose purpose is identification of scalp-recorded ERP neural generators indicate that the source of the ERN is in the midline prefrontal cortex, where the anterior cingulate cortex (ACC) is located. Additionally, measurement of local field potentials in the ACC of monkeys find increased firing when the animals make an error, at a similar time frame as the ERN (Bush et al., 2000).

Experiments have found a positive relationship between ERN amplitude and a variety of other phenomena, including diagnosis of obsessive-compulsive disorder, diagnosis of generalized anxiety disorder, and higher ratings of anxiety irrespective of diagnoses (Weinberg et al., 2011). Such findings prompted some researchers to suggest that increased ERN itself might be a predictive marker of anxiety and disorders thereof (Hajcak, 2012).

### 0.1.4 State versus Trait Anxiety

A consensus as to the definition of emotions in general is still absent in the literature, but in the context of the current study it makes sense to adopt a view seeing emotion in motivational states and to define their parameters as 1) direction away from or toward a stimulus,

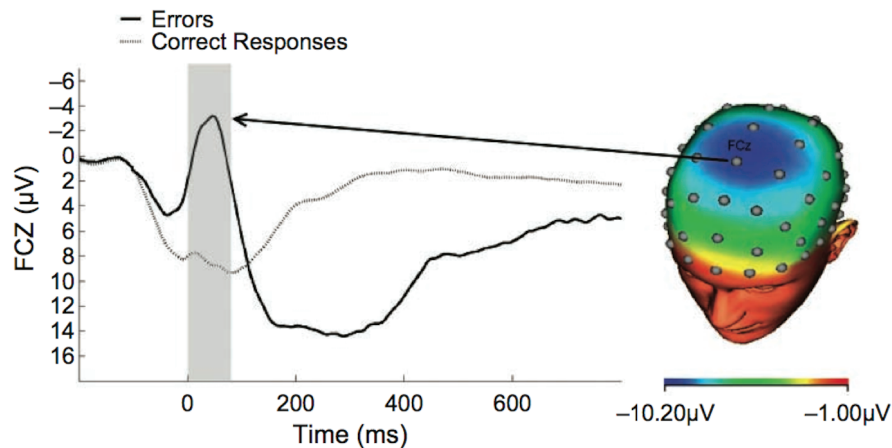


Figure 1: Mean event-related potentials time-locked to correct and erroneous responses at electrode FCz (left). This graph plots voltage over time, where the response occurs at 0 milliseconds (ms). The error-related negativity (ERN) is observed as a sharp negative deflection that peaks around 50 ms (illustrated by the gray bar) after the commission of a mistake. The voltage difference between errors and correct responses in the time window of the ERN can be plotted over the scalp (right); the ERN is maximal at the frontal-central midline of the scalp. Caption also from Figure 1 of Hajcak, 2012

and 2) magnitude of motivation, commonly referred to as valence and arousal respectively (Luck & Kappenman, 2012).

Most generally, state anxiety is differentiated from trait anxiety in its duration and relationship to stimuli: state anxiety can be a reaction to a threatening stimulus, but trait anxiety can be sustained in the absence of a particular cause (Endler & Kocovski, 2001). One measure used to differentiate between these is a 40-item questionnaire called the Spielberger State-Trait Anxiety Inventory (STAI). Though it is among the longest- and most widely-used studies differentiating between these types of anxiety, it remains unclear whether the questions identifying trait anxiety are also measuring depression (Bieling et al., 1998). However, their comorbidity and shared mechanisms – neurophysiological and even genetic – make these generally difficult to isolate (Andrews & Thomson, 2009).

Anxiety is uniquely suitable for ERP studies because other states are likelier to decay over as many trials as are necessary to obtain a meaningful signal. ERPs can serve as a useful complement in discerning causal relationships between stimuli and response (Luck & Kappenman, 2012), because they can be used to measure physiological arousal that would accompany an anxious, but not a depressed, affect that cannot be dissociated in a questionnaire (Grös et al., 2007).

Evidence exists for a significant role of state anxiety in increasing vigilance in cognitive tasks. Activity increases in the same region of the brain (the right ventrolateral prefrontal

cortex) during both exposure to threatening stimuli and during difficult vigilance tasks, e.g. when search stimuli are relatively difficult to distinguish from background stimuli, or when search criteria must be held in working memory for a long period of time (Aron et al., 2004). Similar to such evidence for potential adaptiveness of state anxiety, evidence exists for an adaptive role of the longer-term trait anxiety as well. The analytical rumination (AR) hypothesis posits an adaptive role for the frequently comorbid trait anxiety and depression in attentional control and sustained analysis of complex issues (Andrews & Thomson, 2009).

## 0.2 Open Science

### 0.2.1 Historical Precedence

Most published research findings, whether in medical or psychological science, are not replicable (Ioannidis, 2005). Many causes appear to contribute to nonreplication, including demands for novel positive findings (Taubes, 1995; Fanelli, 2011). Additionally, false positives abound because of the margin allowed by the  $<.05$   $p$ -value criterion for publication (Ioannidis, 2005). Most retractions are due to error, rather than outright fraud. See Table 1.

Table 1: Reasons for Retraction of 742 Papers, 2000 - 10. Source of Retractions: PubMed. Adapted from Steen 2010.

Reason for retraction	Retracted paper, n(%)	IF, mean (SD)
<b>Fraud</b>		
Fabrication	111 (15.0)	9.70 (11.49)
Falsification	98 (13.2)	8.01 (8.33)
<b>Error</b>		
Scientific mistake	234 (31.5)	10.60 (10.66)
Duplicate publication	117 (15.8)	3.18 (2.63)
Plagiarism	107 (14.4)	2.31 (1.90)
Ethical violations	76 (10.2)	5.74 (8.51)
Unstated reasons	61 (8.2)	5.08 (8.82)
Journal error	27 (3.6)	2.66 (2.21)
IF = impact factor		

Additionally, the structure of professional science today emphasizes publications, and those publications hinge on novel positive findings. These publications are the arena of labs' competition for funding. Meta-analyses find that statistically significant results are three times more likely to be published than papers affirming null results (Jennings & Horn, 2011).

Because it is circular to explain this disparity being caused by experimenters' desires to submit significant findings, it is also important to consider the roles that reviewers and editors have in preferring those significant findings. Not only are concurrent articles differentially prioritized according to whether they report positive results, but positive results are privileged if they are novel. This privileging occurs irrespective of the frequency of null findings (i.e. failure to replicate), a phenomenon called latent bias. Finally, retractions of improperly reported positive findings are higher among more influential journals (Fang & Casadevall, 2011) as indexed by their impact factor, i.e. how frequently their recent articles are being cited in other journals.

Transparency would address methodological problems across kind and degree, from innocuous problems of standardizing experimental conditions and protocols to outright forgery. Thanks to the internet, researchers interested in implementing these practices are able to coordinate. Some endeavors are ostensibly more social, e.g. ResearchGate, of note because it is a place where researchers can post their research for free, at least for articles published in journals that allow this. Other potential solutions are concerned more with methodology, as in Figshare, a site where every component of the scientific process, including methods and datasets, can be shared publicly for free and assigned their own digital object identifiers (DOIs) to be used for citation.

Reaching consensus that science needs to be more open is the first and simplest obstacle to proliferating the practice. Adoption in the community and standardization within and among practices proves to be much more complex. While analyses of extant publications are important in both understanding the kind and extent of such problems, an important complement to these endeavors is systematic replication. The Center for Open Science was founded to integrate these solutions.

### 0.2.2 The Center for Open Science

The Center for Open Science has three primary activities: the Open Science Collaboration (OSC), the Open Science Framework (OSF), and the Reproducibility Project (RP). The Open Science Collaboration is the broader community interested in open practices, spanning scientific disciplines and open-source developers. The OSF is focused on developing infrastructure for collaboration, including version control, a centralized place for sharing and finding experimental materials, documenting work, and standardizing workflow. The RP is focused on both the rate of reproducibility and how this rate can be improved.

## The Reproducibility Project

At the time of this writing, over 100 scientists have joined a single endeavor to replicate experiments published in three prominent psychological science journals: *Psychological Science*, the *Journal of Personality and Social Psychology*, and the *Journal of Experimental Psychology: Learning, Memory, and Cognition*. The purpose is not just to test reproducibility of published research findings, but to develop a methodology to improve it ([Open Science Collaboration, 2012](#)).

Most studies come from the fields of social and cognitive psychology because they do not require specialized equipment and are less expensive, and most replications require undergraduate subjects ([Carpenter, 2012](#)). Though such a homogeneous research population presents its own problems, a significant advantage exists in limiting costs of such studies because it serves dual purposes, including training ([Frank & Saxe, 2012](#)).

Participation in the Reproducibility Project entails collaboration with other members to develop strategies for adhering as closely to the original study as possible given inevitable differences between the respective labs and their access to materials. Because the course of the replication will be transparent, participating researchers will send a report outlining their plan for replication according to a template. See Appendix C: [Report to the Reproducibility Project](#) for a comparison between methods in this and the original study, as the methods in the following chapter describe only the methods for this replication.

# Chapter 1

## The Replication

### 1.1 Methods

#### 1.1.1 Participants

Participants were 30 undergraduate students aged 18-35 (18 female, 12 male). Human Subjects Research Committee approval was obtained before the study began.

#### 1.1.2 The Task

Participants performed a flanker task in which they viewed five horizontally aligned arrows (< < > > >) on a computer screen. On each trial, participants were to specify the direction of the central arrow with the click of a left or right mouse button. Whether the central arrow was aligned with flankers was pseudorandomly chosen and of equal frequency. Trials on which central and flanker arrows were aligned will hereafter be called compatible, while trials on which the central and flanker arrows were not aligned will hereafter be called incompatible.

Each participant first performed one practice block of 30 trials followed by eight additional blocks of 30 trials. Stimuli (i.e. the arrows) were presented for 200 ms with an intertrial interval (ITI) that varied pseudorandomly from 500 to 1000 ms. At the end of each of these blocks, participants saw on the screen one of three messages according to performance: if 75% accurate or below, they saw “Please try to be more accurate”; if above 90% they saw “Please try to respond faster”; if between these levels of accuracy, they saw “You’re doing a great job.” The purpose of this feedback was to encourage participants to balance speed and accuracy, and also to induce enough errors to examine for the ERN.

### 1.1.3 Errors and EEG

While performing this task each participant was monitored for his or her electroencephalographic (EEG) and electromyographic (EMG) activity using the EasyCap electrode cap and BrainAmps Standard system (Brain Products, Gilling, Germany). Recordings were taken from 96 scalp electrodes at equidistant positions as well as from two reference electrodes placed on the left and right mastoids. Electrooculogram (EOG) from blinks and other eye movements were recorded from four facial electrodes placed approximately 1 cm: above the right eye, below the right eye, left of the left eye, and right of the right eye. All bioelectric signals were digitized on a laboratory microcomputer using Recorder software (Brain Products). Sampling was at 1000 Hz.

Offline analysis was performed using Brain Vision Analyzer software (Brain Products, Gilching, Germany). Data were referenced to the numeric mean of the mastoids and band-pass filtered with cutoffs of 0.1 and 30 Hz. EEG was segmented for each trial, beginning 200 ms before response and continuing for 800 ms. It was corrected for blinks and eye movements using EMCP (Gratton et al., 1983). The ERN was defined as the average activity in a 0- to 100-ms window following response onset on error trials, but it generally peaks approximately 50 ms following commission of an error. The ERN was statistically evaluated using R, with Greenhouse-Geisser correction applied to p values associated with multiple degrees of freedom, repeated measures comparisons.

### 1.1.4 Startle and EMG

Startle probes consisted of a 50-ms 99-dB white noise with instantaneous onset 300ms after response. They were presented on 10% of all trials in the practice block. In the eight blocks following practice, they were presented on 4% of correct trials preceded by correct trials (called “correct unpredictable” because the presentation of the startle probe is relatively unlikely), 50% of error trials, and 50% of correct trials following errors (also called “correct predictable” as the probability of 50%, though random, is a much higher rate than that of the other correct trial type).

Startle response was measured with two electrodes placed approximately 12 mm apart under each participant’s left eye on the orbicularis muscle. Startle response EMG data was band-pass filtered (28—512 Hz; 24 dB/octave roll-off), rectified, then low-pass filtered at 30 Hz (24 dB/octave) and baseline-corrected. Response magnitude and latencies were quantified according to a peak in the 20- to 120-ms window after the startle probe was presented. It was statistically evaluated using R, with Greenhouse-Geisser correction applied to p values associated with multiple degrees of freedom, repeated measures comparisons.



## 1.2 Results

### 1.2.1 Startle

On average, participants made 36.13 ( $SD = 19.70$ ) errors. As in the original study, the magnitude of the startle response differed as a function of trial type,  $F(2,56)=9.36$ ,  $p < .001$  (see Figure 1.1). Follow-up  $t$  tests found that startle magnitude was greater following errors than following either subtype of correct trial, whether predictable,  $t(28) = 3.79$ ,  $p < .001$ , or unpredictable  $t(28) = 2.90$ ,  $p < .01$ . Additionally, there was no significant difference between correct unpredictable and correct predictable trials,  $t(28)=-1.7025$ ,  $p > .09$ . There was no correlation between number of errors and error-potentiated startle magnitude,  $r = -.10$ ,  $p > .5$ . As in the original study again the latency of startle response was shorter for predictable (for error:  $M = 62\text{ms}$ ,  $SD = 5.8\text{ms}$  and for correct predictable:  $M = 62.7\text{ms}$ ,  $SD=4.7\text{ms}$ ) than for unpredictable trials ( $M = 64.4$ ,  $SD = 6.6\text{ms}$ ), but the difference was not significant  $F(2,56) = 2.81$ ,  $p > .05$ .

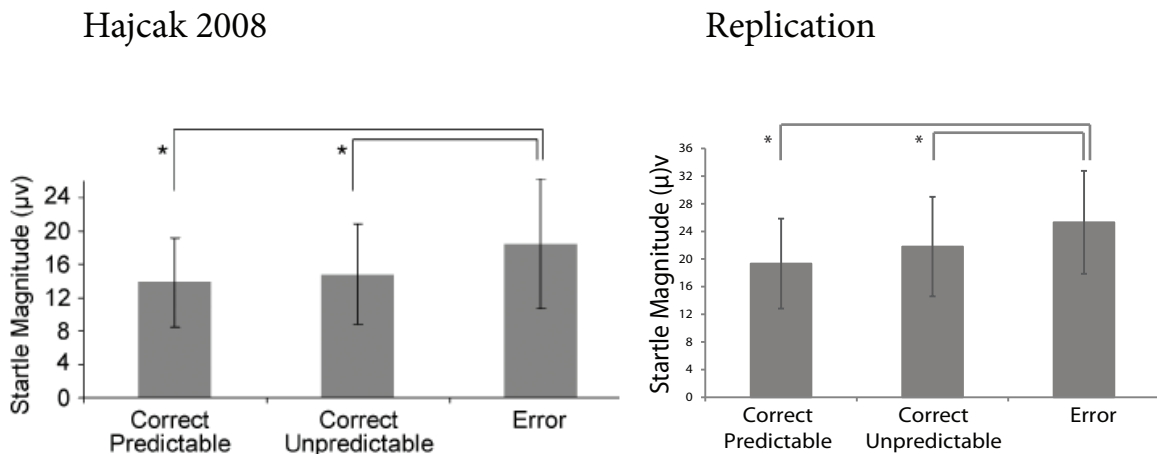


Figure 1.1: Comparison of mean startle magnitude by trial type from original and replication study. Error bars represent the 95% confident interval. Asterisks indicate significant differences between trial type,  $p < .01$ . Caption adapted from Figure 1 of (Hajcak & Foti, 2008)

### 1.2.2 The Error-related Negativity

The ERN is often measured as the difference between ERPs for correct and error trials (here measured in the 0-100ms window following response). The presence of the ERN is assessed by comparing this difference wave to 0. Because the presence of the ERN was investigated in the original article at a single electrode site, FCz, it was investigated here at the closest

analogous site in the 96-channel setup, electrode 2. At electrode 2, a one-sample  $t$  test found that ERN amplitude (0-100ms post-response) differed significantly from 0,  $t(29) = -6.85$ ,  $p < .001$ . Figure ?? depicts the scalp topography of the ERN on the left and the grand average ERP waveform for correct and error trials on the right. The peak at 400ms in the replication study ERPs is absent in the original study. While the nature of this extra peak is as yet undetermined, its timing is such that it could be a post-auricular muscle (or PAM) response, which is well-documented electrical activity evoked in the muscle behind the ear (on the mastoid) following a brief auditory stimulus (O'Beirne & Patuzzi, 1999). This signal would appear in most scalp channels, even one on the top of the head, because they are all referenced to the mastoid electrodes. Further supporting this possibility is the difference in amplitude of this peak between correct and error trials. The probability of presentation of the auditory stimulus was a combination of 4% for correct unpredictable and 50% of correct predictable trials, which would result in fewer presentations than for just error trials because that probability was always 50%. Because the original study also referenced to the mastoids after recording, it is unclear why the PAM response does not also appear in their data.

### 1.2.3 Error-potentiated Startle and the Predictive Value of the ERN

Analogous to finding the ERN by subtracting correct from error trials, the original authors defined the error-potentiated startle (EPS) as the startle magnitude after errors minus the average startle magnitude of correct predictable and correct unpredictable trials. The original study found that the ERN predicted the potentiation of startle response, but the current study failed to replicate this finding,  $r = -.28$ ,  $p > .12$ .

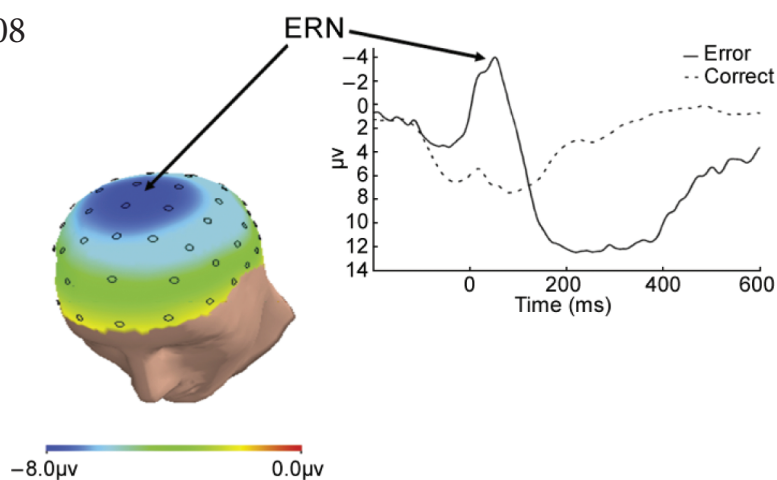
## 1.3 Discussion

Results for the ERN and for startle aligned closely with those of the original study, but the correlation between the EPS and ERN did not hold. See Figure 1.3 for comparison of scatterplots between the original and replication studies. However, the original author has also faced difficulty replicating this finding (Riesel et al., 2013). A significant correlation between EPS and the ERN emerged only after the authors performed exploratory post-hoc analyses and formed two subgroups of participants based on a median split of participants into groups with high and low amplitude relative to  $-8.25\mu V$ . A correlation between ERN amplitude and startle magnitude was found among those participants whose ERN amplitude was relatively high, i.e. above the median,  $r = -.60$ ,  $p < .05$ . Dividing participants in the current study along the median ( $5.85\mu V$ ) also generated a significant correlation,  $r = -.64$ ,  $p < .001$ . See

Figure 1.4.

While this median split led to significant correlations in both studies, the scatterplots from each study reveal a potential outlier. When the array of values for the ERN and for the EPS are investigated independently for the value with the largest difference from the sample mean, each is found to belong to the same participant ( $z(\text{ERN}) = -2.98$ ,  $z(\text{EPS}) = 2.95$ ). When this participant is removed, the correlation fails to be significant even when participants are split by the median,  $r = .09$ ,  $p > .75$ . See Figure 1.5 for a comparison of correlations before and after removal of this potential outlier.

Hajcak 2008



Replication

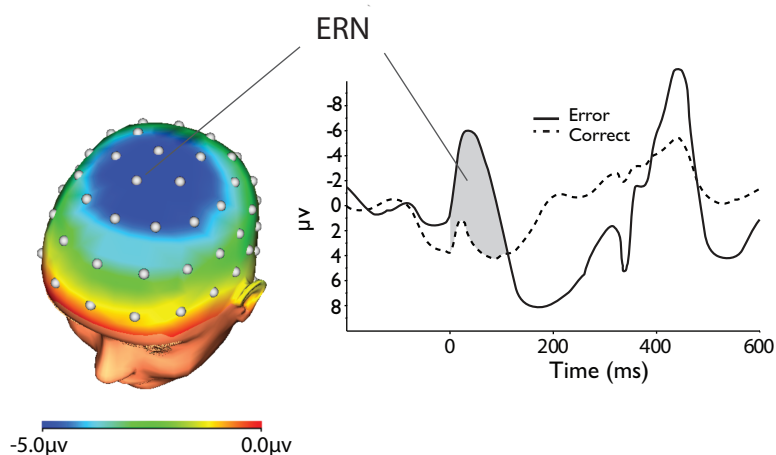
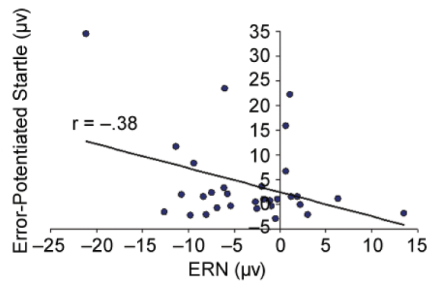


Figure 1.2: Comparison of mean event-related potentials time-locked to correct and erroneous responses at analogous electrode sites FCz and 2 between the original and replication studies, respectively (left). The graph (right) plots voltage over time, where response occurs at 0 milliseconds (ms). The error-related negativity (ERN) is a sharp negative deflection that peaks around 50 ms after the commission of a mistake, and its amplitude is calculated as the voltage difference between errors and correct responses in a 100-ms time window. The ERN is plotted over the scalp (left) and is maximal at the frontal-central midline of the scalp.

Hajcak 2008



Replication

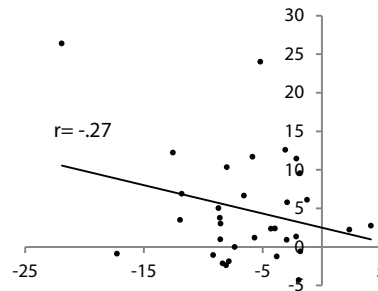
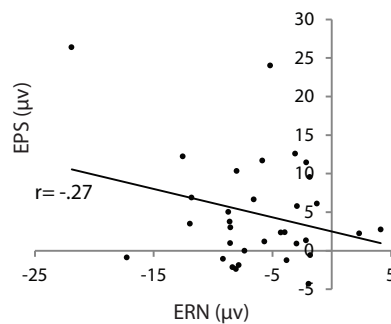


Figure 1.3: Left: original scatterplot depicting the magnitude of the error-potentiated startle as a function of ERN amplitude. Right: replication findings, not significant ( $p > .05$ )

Replication



High ERN Amplitude

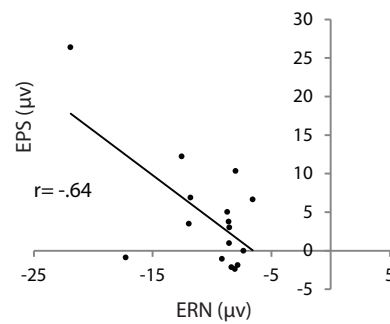
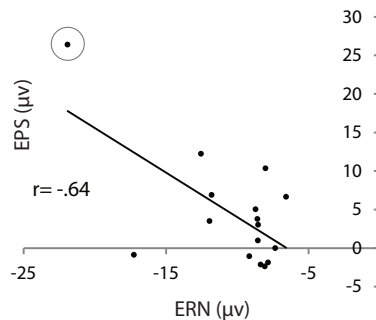


Figure 1.4: Left: Replication scatterplot depicting the magnitude of the error-potentiated startle as a function of ERN amplitude,  $r = -.27$ ,  $p > .05$ . Right: correlation plot including only those participants whose ERN amplitudes were above median  $5.85\mu V$ ).

High ERN Amplitude



Without Outlier

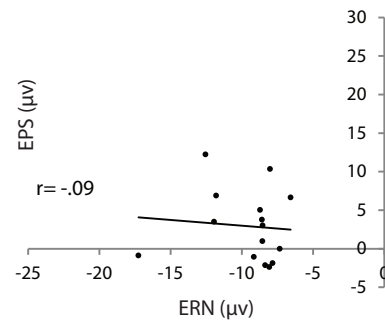


Figure 1.5: Left: Scatterplot depicting the magnitude of the error-potentiated startle as a function of ERN amplitude in participants whose ERN was of relatively high amplitude,  $r = -.60$ ,  $p > .05$ , outlier circled. Right: Same as left, but without outlier.



# Chapter 2

## The Extension

### 2.1 Methods

Following the replication, participants were asked to complete the state-trait anxiety inventory (STAI) (see [State versus Trait Anxiety](#)). The task and stimuli were identical to that of the replication.

Additionally, participants performed the task in another 8 blocks of 30 trials for each of two conditions. In the low anxiety condition, subjects received +2 points for correct responses faster than their average RT, +1 points for correct responses slower than their average RT, and -1 points for incorrect responses. In the high anxiety condition, +1 point was received for correct/fast trials, -1 point for correct/slow trials, and -2 points for incorrect trials. The order in which each of these conditions was presented was counterbalanced between subjects. By manipulating points, we sought to manipulate state anxiety. Subjects were informed at the outset that the fastest, most accurate participant would win \$75. Participants filled out an identical STAI form following the extension trials as well.

### 2.2 Results

#### 2.2.1 Anxiety

There was no significant difference between trait scores reported by participants before and after anxiety manipulation  $t(24) = -.49, p > .63$ . There was also no significant difference in state anxiety before and after the anxiety manipulation,  $t(24) = 1.09, p > .28$ . Trait and state anxiety were not significantly correlated before anxiety manipulation,  $r=.28, p > .14$ . Following anxiety manipulation, however, trait and state anxiety are significantly correlated,  $r=.62, p < .01$ .

### 2.2.2 Anxiety and Electrophysiological Measures

There was no effect on ERN of anxiety manipulation ( $F(1,25) = 1.10, p > .30$ ), order ( $F(1,25) = 0.40, p > .54$ ) or interaction between the two ( $F(1,25) = 1.60, p > .22$ ). There was no effect on EPS of anxiety manipulation ( $F(1,25) = 1.71, p > .20$ ), order ( $F(1,25) = 4.05, p > .05$ ) or interaction between the two ( $F(1,25) = 0.49, p > .49$ ).

Because the EPS magnitude is calculated by subtracting the average startle magnitude of correct trials from error trials, it may be obscuring differential effects of anxiety manipulation on startle magnitude within these trial types. An ANOVA examining effects on startle magnitude within trial type (error, correct predictable, or correct unpredictable) and anxiety condition (low or high) and between the order of anxiety conditions (low to high or high to low) found main effects of anxiety condition ( $F(1,25) = 5.39, p < .01$ ) and trial type ( $F(2, 50) = 4.96, p < .01$ ) (see Figure 2.1), and an interaction between order and anxiety condition ( $F(1, 25) = 12.29, p < .01$ ). This ANOVA found no effect of order ( $F(1, 25) = 0.90, p > .35$ ). Additionally, interactions between order and trial type ( $F(2, 50) = 2.94, p = .06$ ); trial type and anxiety condition ( $F(2, 50) = 0.90, p > .41$ ); and order, trial type, and anxiety condition ( $F(2, 50) = 2.18, p > .12$ ) were not significant. See Table 2.1.

Table 2.1: Overall Startle Magnitude Effects

Effect	<i>df</i>	<i>F</i>	<i>p</i>
order	1, 25	0.90	.35
anxiety condition	1, 25	5.39	.03*
trial type	2, 50	4.96	.01*
order:anxiety condition	1, 25	12.29	.001*
order:trial type	2, 50	2.94	.06
anxiety condition:trial type	2, 50	0.90	.41
order:anxiety condition:trial type	2, 50	2.18	.12

A final series of ANOVAs examined the effects of order and anxiety condition and their interaction for each trial type. See Table 2.2 for a summary.

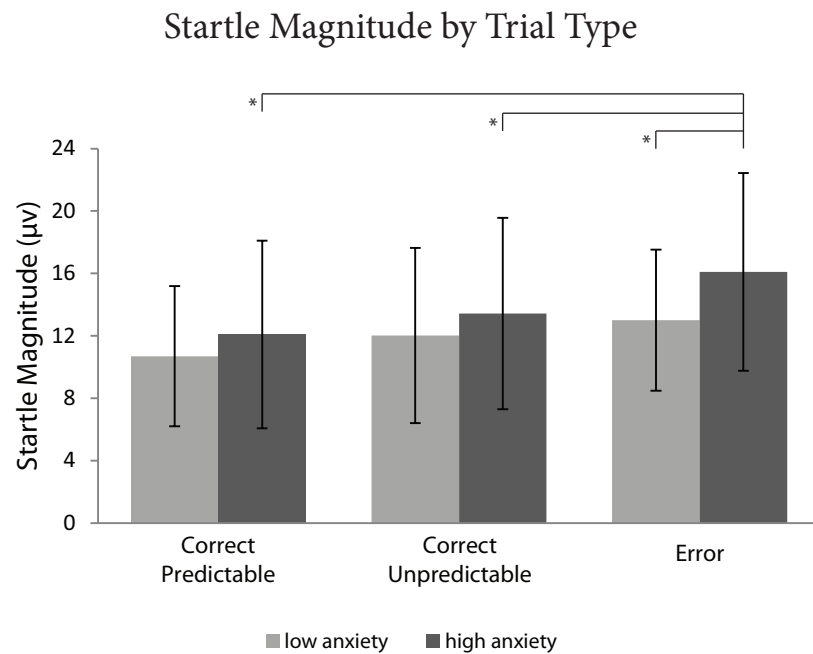
Analysis of variance in error startle magnitude within anxiety condition and between order found no significant effect of order ( $F(1, 25) = .23, p > .63$ ). However, there was an effect of anxiety condition,  $F(1, 25) = 5.30, p < .03$ , and an interaction between order and anxiety condition,  $F(1, 25) = 6.90, p < .02$ . Error startle magnitude was significantly higher in the high anxiety condition,  $M = 15.85\mu V, SD = 17.36$ , than of the low condition,  $M = 12.63\mu V, SD = 12.26, t(26) = -2.08, p < .05$ . See Figure 2.2.



Table 2.2: Startle Magnitude ANOVAs by Trial Type

Effect	Error	Correct Pre- dictable	Correct Unpredictable
order	.23	1.27	1.39
anxiety condition	5.30*	1.47	1.74
order:anxiety condition	6.90*	10.76**	1.63

*F* values, \* $p < .05$ , \*\* $p < .01$

Figure 2.1: Mean startle magnitude ( $\mu$ V) by trial type for low and high anxiety conditions.

Analysis of variance in correct predictable startle magnitude within anxiety condition and between order found no significant effect of order,  $F(1, 25) = 1.27$ ,  $p > .26$ , or anxiety condition,  $F(1, 25) = 1.47$ ,  $p > .23$ . However, there was an interaction between order and anxiety condition,  $F(1, 25) = 10.76$ ,  $p < .01$ . See Figure 2.3.

Analysis of variance in correct unpredictable startle magnitude within anxiety condition and between order found no significant effect of order,  $F(1, 25) = 1.39$ ,  $p > .24$ , anxiety condition,  $F(1, 25) = 1.74$ ,  $p > .19$ , or interaction therein,  $F(1, 25) = 1.63$ ,  $p > .21$ .

There was no correlation between error startle magnitude and state anxiety either before or after the extension trials. There was also no correlation between ERN and state anxiety or between ERN and trait anxiety, before or after anxiety manipulation.

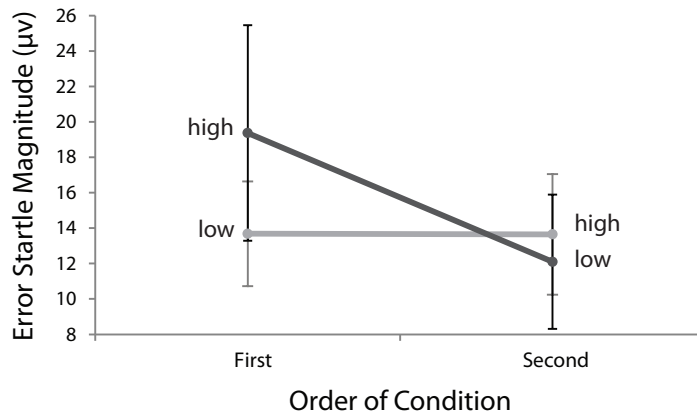


Figure 2.2: Mean error startle magnitude ( $\mu\text{V}$ ) between low and high anxiety conditions, split by order in which anxiety manipulations were presented.

## 2.3 Discussion

The finding of no significant difference between trait scores before and after anxiety manipulation corroborates the internal validity of the STAI. However, the insignificant difference between state anxiety before and after the manipulation may be more complex: perhaps state anxiety changes were brief enough not to be reported by the end of the experiment, or the manipulation was not effective at any point. There may have been other effects from having ended the experiment, because the participant is able to eat, drink, and communicate with experimenter, and anticipates leaving. Though other research has found a relationship between the ERN and higher ratings of anxiety, such a relationship did not manifest in this study.

The significant effect of anxiety condition on startle magnitude for error trials suggests the feedback had an effect on defensive motivation. That the STAI was administered only before both and after both conditions, but not between them, precludes an analysis of the effect of each anxiety condition on state anxiety independently. This may be why no correlation between state anxiety and error startle magnitude was found.

The interaction effect of order and anxiety condition may be attributable to a downward trend of startle response from first to second conditions (whether low to high or high to low anxiety), but that flattens for the low to high condition because whatever decrease attributable to order is matched by what would have been an increase caused by the high anxiety condition itself. See Figure 2.2 (right). This is consistent with the order by extension interaction effect found for correct predictable startle magnitude; irrespective of order, startle decreases in the second condition seen. See Figure 2.3.

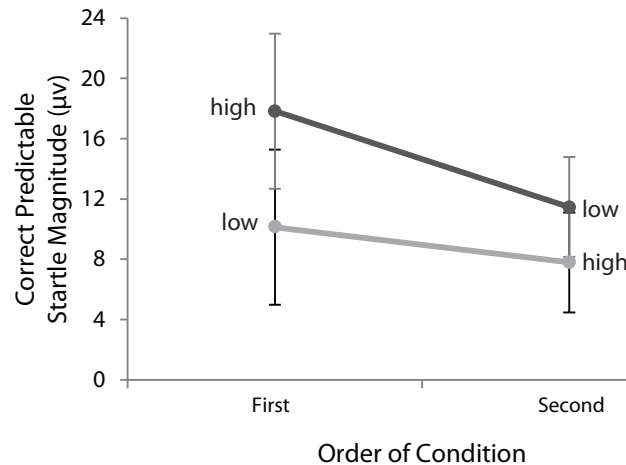


Figure 2.3: Mean correct predictable startle magnitude ( $\mu V$ ) between low and high anxiety conditions, split by order in which anxiety manipulations were presented. Participants in the high-to-low anxiety condition order appeared to both begin with a higher startle magnitude and to decrease more.

## 2.4 General Discussion

This study replicated the original study's finding that startle magnitude was significantly greater for error than for correct trials. The ERN was significant as well. However, the ERN failed to predict the EPS until, by virtue of a median split, those participants with relatively low-amplitude ERNs were excluded (Riesel et al., 2013). When the participant whose ERN and EPS each nearly count as outliers ( $z(\text{ERN}) = -2.98$ ,  $z(\text{EPS}) = 2.95$ ) is excluded, the correlation between ERN amplitude and EPS magnitude is no longer significant. See Figure 1.5.

That the magnitude of startle response following errors was larger than that of correct trials held in the extension, as well. Error startle was also greater in the higher anxiety condition. Taken together, these results corroborate the relationship between error monitoring and defensive motivation and also the significance of reward incentives and punishment described in the introduction (Leue et al., 2012; Weinberg et al., 2011). However, whether this result points to an effect of anxiety per se remains unclear because no significant relationship between STAI and electrophysiological results emerged.

Although it was just short of statistical significance, a modest correlation existed between relatively high-amplitude ERNs defined by the median split and trait anxiety. Further research, perhaps using other measures of trait anxiety, may find a third variable effect of anxiety driving the relationship between ERN and EPS. Indeed, evidence exists for a relationship between ERN enhancement and a variety of disorders associated with sensitivity to error commission, including Obsessive-Compulsive Disorder and Generalized Anxiety

Disorder (Riesel et al., 2013).

That the STAI largely failed to relate meaningfully to physiological measures might be accounted for by a variety of potential confounds for anxiety. Length of time required for preparation of experiment ranged from 90 minutes to 3 hours, and familiarity between participant and experimenter and participant and assistants varied widely. Additionally, some subjects were run during the semester when they had work and school obligations, while some were run over their college winter vacation. Finally, time of day experiments were run varied between 8am and 11pm. Future experiments would hold more of these factors constant. To reduce effects of interaction between the experimenter and the participant, the experimenter could adhere to a script. Participants would be run either exclusively while on a break or when school is in session. Participants would be run during a narrower window of time during the day.

### 2.4.1 On the Practice of Replication

Differences known between the circumstances of the original study and those of the replication include number of electrodes used (64 and 96, respectively), decibel level of aversive auditory stimulus (105dB and maximum measured at 99dB, respectively). Additionally, the relationship between peak latencies of startle within each experiment was identical in that differences between latencies were on the order of 2 to 4 milliseconds, and found not to be significant, but on average latencies were much shorter for this experiment than for the original. However, all peak latencies in the replication were 30 milliseconds earlier than those in the original study. This difference may be accounted for by our having tested the difference between timing of Presentation port codes sent from the computer presenting stimuli to the computer recording EEG data against the timing reported by a microphone recording directly from speakers. Because the actual presentation of sounds was 20ms slower than that reported by Presentation port codes sent, experiments in this lab are programmed to send port codes 20ms after the program presents auditory stimuli.

## Chapter 3

# The Future of Open Science

The introduction identified several conflicts pervading science today and provided a cursory description of the role the Center for Open Science intends to play in addressing these problems. However, it remains important to discuss what specific technologies will make a more open scientific process possible and how they will do so. The success of an open science movement depends not only on the potential of these technologies but on their adoption, which is both a function of their usability and of the priorities of their users.

Thomas Kuhn wrote over 50 years ago in *The Structure of Scientific Revolutions* that scientific progress had been long and inaccurately considered a simple aggregation of knowledge and theory. He attributed this in part to how insulated professional science had typically been from the laity, as the professional scientist was typically writing for his colleagues and not beholden to commercial interests. The tradeoff for efficiency in work that took first principles as given was a kind of ahistoricism and a retrospectively linear narrative of progress. He called for a new perspective of scientific progress that held paradigm shifts at its crux. His examples of such shifts were typically incited by new theories incommensurate with those foundational previously, e.g. the Copernican revolution. However, they all rest on his assertion that inasmuch as science progresses with paradigm shifts, it is also subject to contemporary culture: “Lifelong resistance, particularly from those whose productive careers have committed them to an older tradition of normal science, is not a violation of scientific standards but an index to the nature of scientific research itself.”

More recently, authors have extended this argument not only to an assumption of the relationship between science and the culture in which it is practiced, but projection of an imminent revolution in the culture and practice of science itself (Nielsen, 2012; Björk et al., 2010; Open Science Collaboration, 2012). Many potential origins are driving this revolution, but failure to replicate is a prominent convergence of many of these. Regardless of whether the replication fails due to chance, fraud, or poor methods in data collecting and/or anal-

ysis, current methods of sharing information make replication difficult (Gad-el Hak, 2004; Ioannidis, 2013) and retraction of questionable research opaque (Van Noorden, 2011).

The results of academic research have been shared in the form of journals since the mid-17th century (Nielsen, 2012), and though that method began as a radically open means of sharing information for the time, it has since evolved with changing cultural pressures on science as a profession. Because publishing is today the primary means by which a researcher can build reputation, it is fraught with competition both on the publisher's side to privilege novel, positive findings (Nielsen, 2012) and on the researcher's side to produce those findings (Gad-el Hak, 2004). It is suggestive, for example, that a 2011 study of 4600 papers from all scientific disciplines published between 1990 and 2007 found a 22% increase in positive findings (Fanelli, 2011).

Meta-analyses suggest this selection bias has driven most published research findings to be false (Ioannidis, 2005) or otherwise untenable for replication, either because their data analyses are inconsistent (Prinz et al., 2011), otherwise incorrect because of improper data collection (John et al., 2012), or even fabricated (Van Noorden, 2011).

Whatever solution there will be to the problem of replicability will also need to confer advantages on adoptees of that solution. Many scientists report being too busy to go through the extra steps to share their data, or having difficulty with the infrastructure being used (Nelson, 2009). Another prominent concern is ownership, in that publishing data prior to publishing a complete article might allow other researchers to steal their ideas or otherwise better them. However, founder of the Center for Open Science Brian Nosek says the answer to this is that precedence is established when research is put out in the open rather than when it is formally published (Bohannon, 2013).

Nosek established the Open Science Framework as a web-based infrastructure to facilitate reputation-building based on shared materials and methods rather than simply results, and he is not alone in envisioning a future that incentivizes more varied contributions to scientific progress. Figshare, for example, is a file sharing platform initiated to allow scientists to publish every component of their research, from datasets to videos. Each object will have a digital object identifier (DOI) assigned to it so it can be cited in other research. Additionally, each will be made available under a Creative Commons license under which it is explicitly allowable for a potential user to share and adapt the work as long as it is attributed to the original author and kept in the public domain. This not only allows negative or otherwise unpublished results to be made available, but the possibility for citation of work other than a published finding provides authors incentive to participate. This is due in part to access to published findings becoming increasingly difficult for consumers to maintain, including well-funded university libraries (Sample, 2012).

Subscription prices have outpaced inflation by over 250% in the last 30 years (Dingley, 2005). While the costs of commercial journals have increased in that time, these do not account for the rate at which prices are increasing, particularly in science, technology, and medicine (Panitch & Michalak, 2005). As an example, the largest publisher of academic journals Elsevier saw a profit margin of 28% in 2012 (Elsevier, 2013), greater than the record fourth quarter profit recorded by Apple of the same fiscal year (23%) (Dowling, 2012). Elsevier additionally supported the Research Works Act (H.R. 3699) (Lin, 2012) introduced in 2011 that called for prohibition of open-access mandates for federally funded research, and reversion of the National Institutes of Health (NIH) public access policy making such research available online, i.e. open access (OA) (Issa & Maloney, 2011).

The non-profit project Public Library of Science (PLOS) seeks to create an online library of OA journals. PLOS content is funded by a publication fee paid by the author or the author's funder. The NIH includes funds specifically allocated to such charges in its own grants, and the increase in OA influence (Laakso et al., 2011) in the form of OA scientific journals or authors posting manuscripts of articles in open repositories on the web has been suggested a meaningful reason for other research funders like the NIH to require OA availability as well (Björk et al., 2010).

Ultimately, it seems the drive toward openness will need to be spearheaded by such influential entities in professional science, i.e. funding agencies, scientific societies, and journals (Nelson, 2009), who have the leverage to make such openness a necessary condition for publication and reputation building.

As with science, the drive toward openness in software saw its viability yoked to how well it coalesced with traditional models. GNU, the first open answer to the proprietary and already widely adopted version of the Unix operating system developed by AT&T, was distributed with a manifesto. It proclaimed the necessity of the open source movement but also addressed the need for GNU to be compatible with Unix despite its being closed, because that integration made GNU more convenient to adopt (Stallman, 1993). FOSS itself is widespread today. It is even adopted widely among companies providing proprietary software, a model synthesizing proprietary and open source into a "mixed source" model (Casadesus-Masanell & Llanes, 2011). One economist proposes that endeavors like FOSS are successful not in displacing but in being allowed to prosper alongside traditional models thanks to what he calls a networked information economy, "an economy of information, knowledge, and culture that flow [sic] through society over a ubiquitous, decentralized network," (Benkler, 2003).

FOSS and open science share the potential benefit of distributing information alongside traditional models like proprietary software and closed-access publication, respectively. The

pragmatism of an approach to open science that searches for this compromise is confirmed by the seeming success of that compromise in software, largely facilitated by the internet (Casadesus-Masanell & Llanes, 2011).

Contra this optimism regarding the possibilities afforded by the internet, however, is the argument that the Internet is just the next in a series of communications industries to cycle from a decentralized to a consolidated model (Wu, 2011). This vulnerability suggests the imprudence of relying on the internet as an unchecked foundation for FOSS, open science, and the intersection of the two.

The futures of open science and the FOSS are cognate. However the particulars of material practice in open science will evolve, e.g. whether the print journal will be replaced by electronic alternatives, advocacy of open information sharing must be the common driver.



# Appendix A

## Comparison Tables

Table A.1: This is a complete table of results for comparison between the original and replication studies.

Result	Original	Replication
<b>Startle</b>		
ANOVA: Mean Startle		
Magnitude by Trial Type	$F(2,60)=5.64, p < .05$	$F(2,56) = 9.36, p < .001$
$t$ test: Startle (Error) versus		
Startle (Correct Predictable)	$t(30) = 2.51, p < .01$	$t(28) = 3.79, p < .001$
$t$ test: Startle (Error) versus		
Startle (Correct Unpredictable)	$t(30) = 2.61, p < .01$	$t(28) = 2.90, p < .01$
$t$ test: Startle (Correct Predictable)		
versus Startle (Correct Unpredictable)	$t(30) = 0.99, p > .30$	$t(28) = -1.72, p > .05$
<b>ERN</b>		
$t$ test: ERN difference from 0	$t(30) = 3.83, p < .001$	$t(29) = -6.85, p < .001$
<b>EPS-ERN correlation</b>		
Correlation between EPS and ERN	$r = .38, p < .05$	$r = .28, p > .05$
Correlation After Median Split	$r = -.60, p < .05$	$r = -.56, p < .03$



# Appendix B

## Free and Open Source Software Used

The only licensed software used was Presentation, for presentation of stimuli, Brain Vision Recorder, and Analyzer. The electrodes amplifiers converted the signal from analog to digital, Brain Vision Recorder collected raw EEG and data trigger codes from Presentation, and Analyzer was used for EEG preprocessing.

See The Future of Open Science for a discussion of FOSS.

Table B.1: Free (F) and Open Source (OS) Software Used.

Purpose	Software	License	Status
typesetting	LaTeX	LaTeX Project Public License	F/OS
reference management	BibDesk	Berkeley Software Distribution	F/OS
bibliographic data	Zotero	Affero General Public License	F/OS
data analysis	R	GNU General Public License	F/OS
version control	Git	GNU General Public License	F/OS

Code for this experiment can be found at [github.com/meli-lewis/](https://github.com/meli-lewis/).



# Appendix C

## Report to the Reproducibility Project

Replication of “Errors Are Aversive”

by Greg Hajcak & Dan Foti

(2008, *Psychological Science*)

Melissa Lewis and Michael Pitts

mlewis@reed.edu

mpitts@reed.edu

### C.1 Introduction

The error-related negativity (ERN) is a negative-going event-related potential component that peaks very quickly after a commission error, and owing to its modulation by affective state and its maximal location being over the anterior cingulate cortex, could be functionally related to a defensive motivational state. That is, it may have some relationship to physiological responses associated with the organism being in danger. The human startle response is a specific measure of defensive state and is often measured by blink response at the orbicularis muscle just below the eye. If such a stable measure of defensive motivation is correlated to the ERN in magnitude, it suggests the latter’s functional role in defensive motivation.

The original experiment demonstrated both that people’s startle response to a loud auditory stimulus is greater when they have just committed an error and that the magnitude of this response is correlated to the amplitude of the ERN.

## C.2 Methods

### C.2.1 Power Analysis

Because the main finding of the original study was a correlation of .38,  $p < .05$ , one-tailed, and I am to meet a power of .8, I will run 42 subjects.

### C.2.2 Planned Sample

Sample will be 18-35 years old, with normal or corrected-to-normal vision and no history of neurological disorders.

### C.2.3 Materials

“Continuous electroencephalographic (EEG) and electromyographic (EMG) activity was recorded using an ActiveTwo head cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, The Netherlands).”

We used the EasyCap electrode cap and BrainAmps Standard system (Brain Products, Gilling, Germany).

“Recordings were taken from 64 scalp electrodes based on the ten-twenty system, as well as from two electrodes placed on the left and right mastoids. The electrooculogram (EOG) generated from blinks and eye movements was recorded from four facial electrodes: two approximately 1 cm above and below the participant’s right eye, one approximately 1 cm to the left of the left eye, and one approximately 1 cm to the right of the right eye.”

We used caps with 96 scalp electrodes at equidistant positions as well as from two reference electrodes placed on the left and right mastoids. EOG was recorded identically.

“The startle response was measured with two electrodes placed approximately 12 mm apart under the participant’s left eye on the orbicularis muscle. As per BioSemi’s design, the ground electrode during acquisition was formed by the Common Mode Sense active electrode and the Driven Right Leg passive electrode. All bioelectric signals were digitized on a laboratory microcomputer using ActiView software (BioSemi). Sampling was at 1024 Hz.”

EMG electrodes were placed under the eye and 12mm apart. Though the original article did not describe their placement except in reference to the other, the medial electrode was placed under the pupil and the lateral electrode was placed 12mm to the right of it, as per Hajcak’s directions in an email correspondence. His directions about vertical placement of the electrodes were to feel the orbicularis muscle for where the blink is maximal. All

bioelectric signals were digitized on a laboratory microcomputer using Recorder software (Brain Products). Sampling was at the rate allowable by the software, 1000 Hz.

#### C.2.4 Procedure

“The task was an arrow version of the flankers task (cf. Hajcak et al., 2005). On each trial, five horizontally aligned arrowheads were presented, and participants had to respond to the direction of the central arrowhead by pressing the left or right mouse button. On compatible trials, all five arrowheads pointed in the same direction (either left or right), and on incompatible trials, the central arrowhead pointed in the direction opposite the direction of the flanking arrowheads. Compatible and incompatible trials were equally frequent, and all stimuli were presented for 200 ms with an intertrial interval that varied randomly from 500 to 1,000 ms. Participants performed eight blocks of 30 trials. At the end of each block, participants received performance feedback designed to encourage fast and accurate responding. If performance was 75% correct or lower, the message ‘Please try to be more accurate’ was displayed; performance above 90% correct was followed by ‘Please try to respond faster’; if performance was between these levels, the message ‘You’re doing a great job’ was displayed. All participants performed one practice block of 30 trials.”

I was unable to obtain the Presentation code from Hajcak, but wrote my own and when I sent it to him, he approved it.

#### C.2.5 Analysis Plan

Analysis was identical to that of the original study except in the software used and in the sampling rate (1000 Hz rather than 1024 Hz).

Offline analysis was performed using Brain Vision Analyzer software (Brain Products, Gilching, Germany). Data was referenced to the numeric mean of the mastoids and band-pass filtered with cutoffs of 0.1 and 30 Hz. EEG was segmented for each trial, beginning 200 ms before response and continuing for 800 ms. It was corrected for blinks and eye movements using EMCP. The ERN was defined as the average activity in a 0- to 100-ms window following response onset on error trials, but it generally peaks approximately 50 ms following commission of an error. The ERN was statistically evaluated using R, with Greenhouse-Geisser correction applied to p values associated with multiple degrees of freedom, repeated measures comparisons.

Startle response EMG data will be band-pass filtered (28–512 Hz; 24 dB/ octave roll-off), rectified, then low-pass filtered at 30 Hz (24 dB/ octave) and baseline-corrected. Response magnitude and latencies will be quantified according to a peak in the 20- to 120-ms win-

dow after the startle probe was presented. It will be statistically evaluated with Greenhouse-Geisser correction applied to p values associated with multiple degrees of freedom, repeated measures comparisons. However, it was studied with R rather than with SPSS as in the original study. Participants were excluded only if their averages following correction contain artifacts of high reference or ground impedance and/or facial movement disrupting a potential signal.

### **C.2.6 Differences from Original Study**

Differences in materials were outlined in the above materials section. The original article did not specify an age range, but the age range for this study was 18-35. The original author analyzed data using SPSS, while this data was analyzed with R. The original article described the auditory stimulus as being 105dB, where in this study the speakers could only go up to 99dB.



# References

- Andrews, P. W., & Thomson, J. A. (2009). The bright side of being blue: Depression as an adaptation for analyzing complex problems. *Psychological Review*, 116(3), 620–654.  
URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0016242>
- Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170–177.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S1364661304000531>
- Benkler, Y. (2003). Freedom in the commons: Towards a political economy of information. *Duke Law Journal*, 52(6), 1245–1276.  
URL <http://www.jstor.org/stable/1373170>
- Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The state–trait anxiety inventory, trait version: structure and content re-examined. *Behaviour Research and Therapy*, 36(7-8), 777–788.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0005796798000230>
- Björk, B.-C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guðnason, G. (2010). Open access to the scientific journal literature: Situation 2009. *PLoS ONE*, 5(6), e11273.  
URL <http://dx.plos.org/10.1371/journal.pone.0011273>
- Bohannon, J. (2013). Psychologists launch a bare-all research initiative. <http://news.sciencemag.org/scienceinsider/2013/03/psychologists-launch-a-bare-all-.html>.
- Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, affective & behavioral neuroscience*, 7(4), 356–366. PMID: 18189009.
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, 4(6), 215–222.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S1364661300014832>

- Carpenter, S. (2012). Psychology's bold initiative. *Science*, 335(6076), 1558–1561.  
URL <http://www.sciencemag.org/cgi/doi/10.1126/science.335.6076.1558>
- Casadesus-Masanell, R., & Llanes, G. (2011). Mixed source. *Management Science*, 57(7), 1212–1230.  
URL <http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.1110.1353>
- Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and biobehavioral reviews*, 28(3), 317–332. PMID: 15225974.
- Dingley, B. (2005). U.s. periodical prices - 2005. <http://www.ala.org/alcts/sites/ala.org.alcts/files/content/resources/collect/serials/ppi/05usppi.pdf>.
- Dowling, S. (2012). Apple reports fourth quarter results. <http://www.apple.com/pr/library/2012/10/25Apple-Reports-Fourth-Quarter-Results.html>.
- Elsevier, R. (2013). Reed elsevier 2012 results. <http://www.reedelsevier.com/mediacentre/pressreleases/2013/Pages/reed-elsevier-2012-results.aspx>.
- Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Journal of Anxiety Disorders*, 15(3), 231–245.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0887618501000603>
- Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. error processing in choice reaction tasks. *Electroencephalography and clinical neurophysiology*, 78(6), 447–455. PMID: 1712280.
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891–904.  
URL <http://link.springer.com/10.1007/s11192-011-0494-7>
- Fang, F. C., & Casadevall, A. (2011). Retracted science and the retraction index. *Infection and Immunity*, 79(10), 3855–3859.  
URL <http://iai.asm.org/cgi/doi/10.1128/IAI.05661-11>
- Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, 7(6), 600–604.  
URL <http://pps.sagepub.com/lookup/doi/10.1177/1745691612460686>
- Gad-el Hak, M. (2004). Publish or Perish—An ailing enterprise? *Physics Today*, 57(3), 61.  
URL <http://link.aip.org/link/PHTOAD/v57/i3/p61/s1&Agg=doi>

- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, 55(4), 468–484. PMID: 6187540.
- Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (STICSA): comparison to the state-trait anxiety inventory (STAI). *Psychological Assessment*, 19(4), 369–381.  
URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/1040-3590.19.4.369>
- Hajcak, G. (2012). What we've learned from mistakes: Insights from error-related brain activity. *Current Directions in Psychological Science*, 21(2), 101–106.  
URL <http://cdp.sagepub.com/lookup/doi/10.1177/0963721412436809>
- Hajcak, G., & Foti, D. (2008). Errors are aversive: defensive motivation and the error-related negativity. *Psychological science*, 19(2), 103–108. PMID: 18271855.  
URL <http://www.ncbi.nlm.nih.gov/pubmed/18271855>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124.  
URL <http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A. (2013). This i believe in genetics: discovery can be a nuisance, replication is science, implementation matters. *Frontiers in Genetics*, 4.  
URL [http://www.frontiersin.org/Statistical\\_Genetics\\_and\\_Methodology/10.3389/fgene.2013.00033/full](http://www.frontiersin.org/Statistical_Genetics_and_Methodology/10.3389/fgene.2013.00033/full)
- Issa, D., & Maloney, C. B. (2011). Research works act, h.r. 3699, 112th cong., 1st sess. (2011).  
<http://hdl.loc.gov/loc.uscongress/legislation.112hr3699>.
- Jennings, R. G., & Horn, J. D. (2011). Publication bias in neuroimaging research: Implications for meta-analyses. *Neuroinformatics*, 10(1), 67–80.  
URL <http://www.springerlink.com/index/10.1007/s12021-011-9125-y>
- 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.  
URL <http://pss.sagepub.com/lookup/doi/10.1177/0956797611430953>
- Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS ONE*, 6(6), e20961.  
URL <http://dx.plos.org/10.1371/journal.pone.0020961>

- Leue, A., Lange, S., & Beauducel, A. (2012). Modulation of the conflict monitoring intensity: The role of aversive reinforcement, cognitive demand, and trait-BIS. *Cognitive, Affective, & Behavioral Neuroscience*, 12(2), 287–307.  
URL <http://www.springerlink.com/index/10.3758/s13415-012-0086-x>
- Lin, T. (2012). Cracking open the scientific process. <http://www.nytimes.com/2012/01/17/science/open-science-challenges-journal-tradition-with-web-collaboration.html>.
- Luck, S. J. (2005). *An Introduction to the event-related potential technique*. Cambridge, Mass.: MIT Press.
- Luck, S. J., & Kappenman, E. S. (2012). *The Oxford handbook of event-related potential components*. New York: Oxford University Press.
- Nelson, B. (2009). Data sharing: Empty archives. *Nature*, 461(7261), 160–163.  
URL <http://www.nature.com/doifinder/10.1038/461160a>
- Nielsen, M. A. (2012). *Reinventing discovery : the new era of networked science*. Princeton, N.J.: Princeton University Press.
- O’Beirne, G. A., & Patuzzi, R. B. (1999). Basic properties of the sound-evoked post-auricular muscle response (PAMR). *Hearing research*, 138(1-2), 115–132. PMID: 10575120.
- Open Science Collaboration (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6), 657–660.  
URL <http://pps.sagepub.com/lookup/doi/10.1177/1745691612462588>
- Panitch, J. M., & Michalak, S. (2005). The serials crisis. <http://www.unc.edu/scholcomdig/whitepapers/panitch-michalak.html>.
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9), 712–712.  
URL <http://www.nature.com/doifinder/10.1038/nrd3439-c1>
- Riesel, A., Weinberg, A., Moran, T., & Hajcak, G. (2013). (in press). time course of error-potentiated startle and its relationship to error-related brain activity. *International Journal of Psychophysiology*.
- Sample, I. (2012). Harvard university says it can’t afford journal publishers’ prices. <http://www.guardian.co.uk/science/2012/apr/24/harvard-university-journal-publishers-prices>.

- Stallman, R. (1993). The gnu manifesto. <http://www.gnu.org/gnu/manifesto.html>.
- Taubes, G. (1995). Epidemiology faces its limits. *Science (New York, N.Y.)*, 269(5221), 164–169. PMID: 7618077.
- Van Noorden, R. (2011). Science publishing: The trouble with retractions. *Nature*, 478(7367), 26–28.  
URL <http://www.nature.com/doifinder/10.1038/478026a>
- Weinberg, A., Riesel, A., & Hajcak, G. (2011). Integrating multiple perspectives on error-related brain activity: The ERN as a neural indicator of trait defensive reactivity. *Motivation and Emotion*, 36(1), 84–100.  
URL <http://www.springerlink.com/index/10.1007/s11031-011-9269-y>
- Wu, T. (2011). *The master switch: the rise and fall of information empires*. New York: Vintage Books.