Developmental Neuropsychology

Implementation of False Discovery Rate for Exploring Novel Paradigms and Trait Dimensions With ERPs

Michael J. Crowley [a] , Jia Wu [a] , Scott McCreary [a] , Kelly F. Miller [b] &
Linda C. Mayes [a]

[a] Yale Child Study Center, Yale University, New Haven, Connecticut,
USA

[b] Department of Psychology, Pomona College, Claremont, California,
USA

PLEASE SCROLL DOWN FOR ARTICLE

Ψ Psychology Press
Taylor & Francis Group

Implementation of False Discovery Rate for Exploring Novel Paradigms and Trait Dimensions With ERPs

Michael J. Crowley, Jia Wu, and Scott McCreary

Yale Child Study Center, Yale University, New Haven, Connecticut

Kelly F. Miller

Department of Psychology, Pomona College, Claremont, California

Linda C. Mayes

Yale Child Study Center, Yale University, New Haven, Connecticut

False discovery rate (FDR) is a multiple comparison procedure that targets the expected proportion of false discoveries among the discoveries. Employing FDR methods in event-related potential (ERP) research provides an approach to explore new ERP paradigms and ERP-psychological trait/behavior relations. In Study 1, we examined neural responses to escape behavior from an aversive noise. In Study 2, we correlated a relatively unexplored trait dimension, ostracism, with neural response. In both situations we focused on the frontal cortical region, applying a channel by time plots to display statistically significant uncorrected data and FDR corrected data, controlling for multiple comparisons.

In the last 50 years event-related potential (ERP) research has yielded a broad number of insights about the neural underpinnings of cognitive, perceptual, and emotional processes (Cacioppo, Tassinary, & Berntson, 2000; Hajcak, MacNamara, & Olvet, 2010; Polich, 1993; Zani, Proverbio, & Posner, 2002). As with many technical fields, including ERP research, scientific advances can be followed by the gradual accumulation of knowledge during periods of what Kuhn (Kuhn, 1996) called "normal science." Broadly speaking, this type of process has led many ERP researchers to follow in the footsteps of those before them, pursuing specific ERP responses and paradigms familiar to the labs they were trained in and examining specific scalp regions where these ERP responses were typically observed.

For instance, after identification of the P3/P300 (Chapman & Bragdon, 1964; Sutton, Braren, Zubin, & John, 1965), a gradual accretion of data about the P300 led to an understanding of its time course, the stimulus and paradigmatic properties that elicit it, and the putative cognitive

and neural processes that support it (Hansenne, 2000a; Polich, 2007; Polich & Criado, 2006). Armed with this knowledge, investigators from a variety of fields applied P300 paradigms to their areas of research, yielding a broad range of phenomenological and clinical applications for the P300 (Hansenne, 2000b; Polich, 2004; Polich & Ochoa, 2004). Now, if an investigator wants to study novelty and attentional salience with ERPs, he or she may look to some variant of the P300 paradigm. Alternatively, if a positive peak ERP response is observed at ∼300 msec in central/central posterior electrode regions under a context of novelty or salience, a discussion of the P3 or P300 is likely to follow. Generally speaking, we know where to look on the scalp for this response, when to look for it, and under what conditions it is likely to emerge.

On the other hand, what if ERP researchers are faced with a different type of problem? For instance, investigators may have a process they would like to study which is amenable to ERP research (Luck, 2005), but which uses a novel paradigm. They may be unsure where to look for ERP differences across conditions, both in terms of timing and scalp location. Of course, to the extent that the investigation resembles previously studied paradigms, prior research may guide the analysis. Another option is to look at a broad range of scalp regions across a wide ERP time course in order to empirically identify condition differences. However, this creates a multiple comparison problem. For instance, if researchers focus only on the frontal region, even within the traditional 10–20 system (F7, F3, Fp1, Fz, F4, F8, Fp2), they could be making many comparisons. These seven electrodes, sampled at 250 Hz within a 900 msec time-window, create a voltage comparison every 4 msec (225 comparisons per channel), yielding 1,575 comparisons, while a high-density array could yield upward of 8,000–16,000 comparisons in the frontal region alone, depending on the EEG electrode array density.

For some time now, functional magnetic resonance imaging (fMRI) research has had to deal with the multiple comparison problem, relying on some combination of a priori comparisons (regions of interest, ROI), masking procedures, and control for multiple comparisons with approaches such as false discovery rate (FDR) procedures (Benjamini & Hochberg, 1995; Benjamini, Krieger, & Yekutieli, 2006; Benjamini & Yekutieli, 2001). The FDR is the expected proportion of false discoveries among observed significant effects (the discoveries). FDR procedures are statistical methods applied to correct for multiple comparisons in multiple hypothesis testing situations. They are expressly directed toward mitigating the increased likelihood of false discoveries that result from conducting many tests. These procedures now allow for mass univariate analyses (i.e., *t*-tests, correlations), providing a more comprehensive examination of ERP data by time and location.

There are a growing number of mass univariate approaches amenable to ERP research including permutation-based procedures (Blair & Karniski, 1993; Bullmore et al., 1999; Maris & Oostenveld, 2007) and several false discovery approaches (Benjamini & Hochberg, 1995; Benjamini et al., 2006; Benjamini & Yekutieli, 2001; Efron, 2004). Because mass univariate procedures are relatively new, most of this work on ERPs has focused on simulation studies (Groppe, Urbach, & Kutas, 2011; Lage-Castellanos, Martinez-Montes, Hernandez-Cabrera, & Galan, 2010) and one recent examination of the P300 (Lage-Castellanos et al., 2010).

To illustrate a potential approach for the use of FDR in ERP research, we applied the original method developed by Benjamini and Hochberg, (B-H FDR, 1995) for independent and positively dependent testing situations. Benjamini and Yekutieli (2001) show that the B-H method controls FDR at the nominal level for dependent tests. Specifically, the technical condition for B-H is known as positive regression dependency (PRD) on a subset. PRD is a technical condition, such

as the simple case with multivariate normal data in which the covariance matrix has all positive elements. For a comprehensive justification of the FDR method under positive dependence see Benjamini and Yekutieli (2001).

It is useful to discuss the B-H FDR procedure within the context of traditional management of multiple comparisons. The Bonferroni correction is probably the most commonly used multiple comparison procedure. With a significance threshold set at α (e.g., .05), and n separate tests performed, the Bonferroni correction sets the threshold for significance of observed p-values at $\leq \alpha/n$. Thus, for a collection of n statistical tests, the Bonferroni correction controls the "family-wise error rate" using an overall threshold of α among all tests, regardless of whether or not the null is actually rejected.

By contrast, the FDR controls the expected proportion of false positives (incorrect rejections of the null hypothesis) among all tests for which the null hypotheses *was* rejected (i.e., the individually significant tests). The B-H procedure assumes p-values will be uniformly distributed under the null hypothesis. For instance, with uniformly distributed p-values in a sorted list, the value 3% of the way down the list should be ~0.03, and so on. Relying on this assumption, the B-H procedure controls FDR by sorting observed p-values (smallest to largest) from a family of observed test statistics and sequentially comparing these to a list of computed B-H critical values. In this way, the larger observed p-values are compared to larger FDR estimates and the smaller observed p-values are compared to smaller FDR estimates.

The computation of B-H values can be easily computed by hand or in a spreadsheet (Benjamini, Drai, Elmer, Kafkafi, & Golani, 2001). Drawing on the example provided by Benjamini et al. (2001), with 17 tests ranked 1 to i for a family of m null hypotheses tested in a study, and α at .05, the largest observed p-value (rank 17) is evaluated at a B-H threshold of .05 $\times$ i/m or .05 $\times$ 17/17 or .05. The next to largest observed p-value (rank 16) is evaluated at .05 $\times$ 16/17 or .047, and so on.

We chose the B-H FDR approach for several reasons. First, it appears to provide a good balance in controlling for both type I and type II errors (Lage-Castellanos et al., 2010). Second, it has relatively good power for detecting narrowly and broadly distributed effects (Groppe et al., 2011). Third, this correction approach has received the most practical application to date in fMRI research (Genovese, Lazar, & Nichols, 2002). Lage-Castellanos et al. (2010) observe that this may be because the B-H FDR is probably less stringent than the FDR method, which accounts for both positive dependence and negative dependence (Benjamini & Yekutieli, 2001).

Although the B-H procedure continues to be the most widely used FDR approach, it is not without shortcomings. For instance its level of conservativeness increases with the number of hypotheses tested, reducing statistical power with large numbers of statistical tests (Qian & Huang, 2005; Storey, 2002).

To illustrate exploratory uses of FDR, we draw on two data sets. The first presents a new paradigm measuring ERP responses following escape behaviors in an adult sample (Study 1). This example illustrates FDR applied to examine ERP differences across conditions with paired-sample t-tests. The second example (Study 2) uses a virtual ball-toss game called Cyberball in which child participants experience social exclusion. ERP social rejection events were then correlated with an ostracism-distress questionnaire across the sample (Crowley, Wu, Molfese, & Mayes, 2010).

Mass univariate approaches still leave the investigator with the problem of summarizing their findings across many tests. We adopted two approaches, Approach A and Approach B to contrast

potential methods for employing FDR along with data aggregation. In Approach A, prioritizes the mass univariate effects under investigation. Accordingly, we first conducted our statistical tests across the entire frontal region with data sampled at 4 msec (Study 1: paired *t*-test; Study 2: Pearson correlation) and then applied FDR. After conducting FDR on individual data points, the investigator is faced with how to describe the overall findings temporally. To address this issue, we used principal component analysis (PCA), windowing the data into temporal regions that reflect common variability in terms of FDR corrected significant effects over time. Temporal PCA allows the investigator to decide where an effect "begins" and "ends" across the channels where the effect was observed. Approach B prioritizes common variability in the ERP waveform. Here we first conducted a temporal PCA across the frontal region, computing mean amplitude for each temporal window identified. We next conducted our statistical tests across all temporal components for all frontal channels and then applied FDR.

Thus, across the two approaches, PCA served related but distinct purposes. In Approach A, PCA was used to aggregate effects found to be significant across windows of time. In Approach B, PCA was used to window ERP data with temporally common variability before applying FDR. Across Study 1 and Study 2, we focus on electrodes in the frontal region.

STUDY 1

Behaviors learned and maintained in order to minimize or eliminate the occurrence of aversive events are negatively reinforced. As a basic mechanism, negative reinforcement (NR) touches many aspects of life. We scratch an itch, we take an aspirin to mitigate a headache and we soothe a crying baby, all to remove an aversive state and experience relief. Similarly anxiety, self-injury, coercive family processes, and addictive behaviors all are supported by negative reinforcement. Functional neuroimaging studies have begun to examine NR processes. Neuroimaging work in humans suggests that both appetitive and aversive prediction errors are represented in the striatum, a key basal ganglia structure (Delgado, Li, Schiller, & Phelps, 2008; Schiller, Levy, Niv, LeDoux, & Phelps, 2008; Seymour et al., 2005). We found that cues reflecting failures in expected avoidance of an aversive sound produces a negative deflection in the fronto-central EEG event-related potential (feedback error-related negativity, fERN) consistent with a prediction error (Crowley, Wu, Bailey, & Mayes, 2009). To date, however, few if any studies employ ERP methods to examine *escape* behaviors putatively supported by negative reinforcement processes. With minimal a priori information regarding the timing and topography of ERP responses for escape behavior, we apply the B-H FDR approach to isolate relevant neural activity.

Methods: Study 1

Participants included 35 undergraduate students, ages 18–22 yrs. (12 male) who participated for course credit. They had normal or corrected-to-normal visual acuity and were screened for current or historical brain injury or disease. The mean age was 18.92 ($SD = 1.10$). Two participants were excluded from study analyses because of excessive data artifact (one male). All procedures were conducted with the understanding and written consent of participants and with approval of the Human Investigation Committee at the Yale School of Medicine.

Procedure

Each participant sat 60 cm before a 19 in. LCD monitor in a dimly lit (60-w bulb), sound-attenuated room. They were instructed to minimize blinks while responding with their right index finger on a 4-button response pad.

Cue Avoidance Task (CAT).   The CAT is a negative reinforcement task for studying escape responses for aversive sounds. The participant's task was to press a button to terminate an aversive noise when a cue appeared or when a cue appeared without noise. For 135 trials, one of three trial types (1 sec silence, 2 sec noise, or 10 sec noise) preceded and mapped onto one of three images (a blue square, a yellow triangle, or a green circle, ∼5 cm by 5 cm). For each participant, a specific noise duration (silence, 2 sec, 10 sec) always preceded the same image. To reduce possible habituation, three possible sounds were used including white noise (30% of trials), an alarm (30% of trials), and a baby cry (40% of trials). The paring of images and conditions was counterbalanced across participants, and within participants the noises were randomly assigned. Thus, duration of sound is the independent variable, and brain response is the dependent variable. For the purposes of our FDR illustration, we compare only the silence to the 10 sec noise condition.

A blank screen with a duration of 900 to 1,000 msec (ITI) began each trial. A visual representation of static (5 cm × 5 cm square) then appeared on screen while the sound (noise or silence) played for the assigned duration (1 sec, 2 sec, or 10 sec). The escape cue (colored shape) appeared after the specified duration and the participant could press a response pad button to escape the noise. Upon the participant's click, the noise and the cue image remained for 750 msec and then disappeared, and the screen went blank for 1,000 msec.

Sounds were played at 75 dB from a speaker positioned 193 cm from the floor and $70 + 5$ cm above the participant. At random points throughout the session, the participant was asked, "How relieved were you, just now, after you pressed the button?" The participant responded on a scale of 0–10, where 0 is "Didn't bother me" and 10 is "Extremely relieved." At the end, participants are asked to "Please rate how relieved you feel when you see the following image" for each of the three cue images, on the same scale as previously mentioned. ERPs were derived for all trials, time-locked to the participant's response (45 trials for each of the three cues).

Electrophysiological methods.   Each participant's head circumference was measured to determine the appropriate net size and to mark Cz as the juncture of the halfway point between nasion to inion and left to right preauricular notches. A Hydrocel high-density 128 Ag/AgCl electrode net array (Geodesic Sensor Net, EGI Inc.) was placed on the participant's head using standard procedures. Before this, the net was soaked in warm potassium chloride solution (KCl) serving as the electrolyte. Amplifier hardware filters were set at .1–100 Hz for data acquisition. Brain wave data were recorded through the Netstation v.4.4 software package (EGI, Inc.) and EGI high impedance amplifiers, sampling at 250 hz (EGI, Inc. Series 200 amplifier). All impedances remained at or under 40 k ohms as indicated by impedance measures made immediately before and after the test session. The E-prime v.2.0 (PST, Inc.) software package controlled the stimulus presentation. The filters were set at .1–30 Hz.

Each ERP epoch included a 100 msec pre-stimulus baseline and a 750 msec post-stimulus interval. Artifact rejection was carried out to eliminate ERPs contaminated by movement and eye artifacts. Segments exceeding the threshold of 200 μV were marked bad. Epochs with any eye blink or eye movement (threshold 150 μV) were rejected. Epochs with more than 10 bad channels

(40% or more segments marked bad) were also rejected. Rejection rates were comparable across stimulus conditions. Averaged data were baseline-corrected by subtracting the average microvolt value across the 100 msec pre-stimulus interval from the post-stimulus segment. After artifact rejection, the single trial data were re-referenced from the vertex (Cz) to an average reference of all electrodes. The trial-by-trial data were then averaged separately for each of the 128 electrode sites and each of the three stimulus conditions (the three shapes). Twenty-three participants had sufficient number of trials after average. Twelve participants with fewer than 25 trials in the average were brought through ocular artifact removal (Blink slope threshold 14 μV/msec) during the preprocessing (Gratton, Coles, & Donchin, 1983). Eleven of these participants had sufficient number of trials in the average. For the purposes of FDR analyses, we focus on 39 channels in the frontal region, which cover the full frontal area, 30% of the channels (see Figure 1, Geodesic Sensor Net channel numbers: 12 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 32, 33, 34, 38, 128, 1, 2, 3, 4, 5, 8, 9, 10, 14, 116, 117, 118, 121, 122, 123, 124, 125, 17, 15, 16, 11, 6).

Results: Study 1

Paired samples *t*-test on each data point and each channel were conducted. *T*-tests compared neural response segmented on escape behavior for the silence condition versus the 10 sec noise condition. Because the data were sampled at 250 Hz, there were 187 samples within the 750 msec epoch, 4 msec per sample. The total number of tests was 187 * 39 (channels) = 7,293. Of these, 1,624 tests were significant without FDR across 33 frontal channels. Uncorrected effects are

FIGURE 1   128 electrode array with frontal electrodes demarcated.
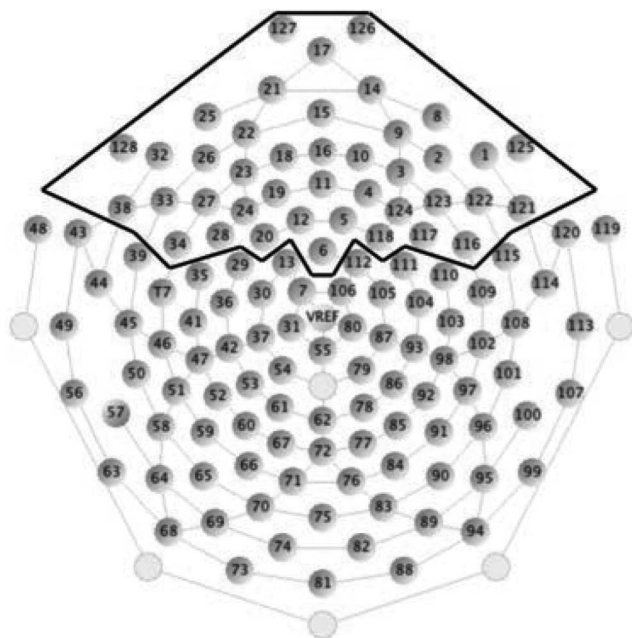
displayed on Figure 2A, plotted *p*-value (y-axis) by time (x-axis) for all channels. Within this cluster, one channel (19) is enlarged for viewing, plotted as the base 10 logarithm, which measures the order of the *p*-value (Figure 2C). At a at false discovery rate of .05 (independent or positively dependent tests), the adjusted alpha (B-H critical *p*) was .0038, meaning that all observed *p*-values
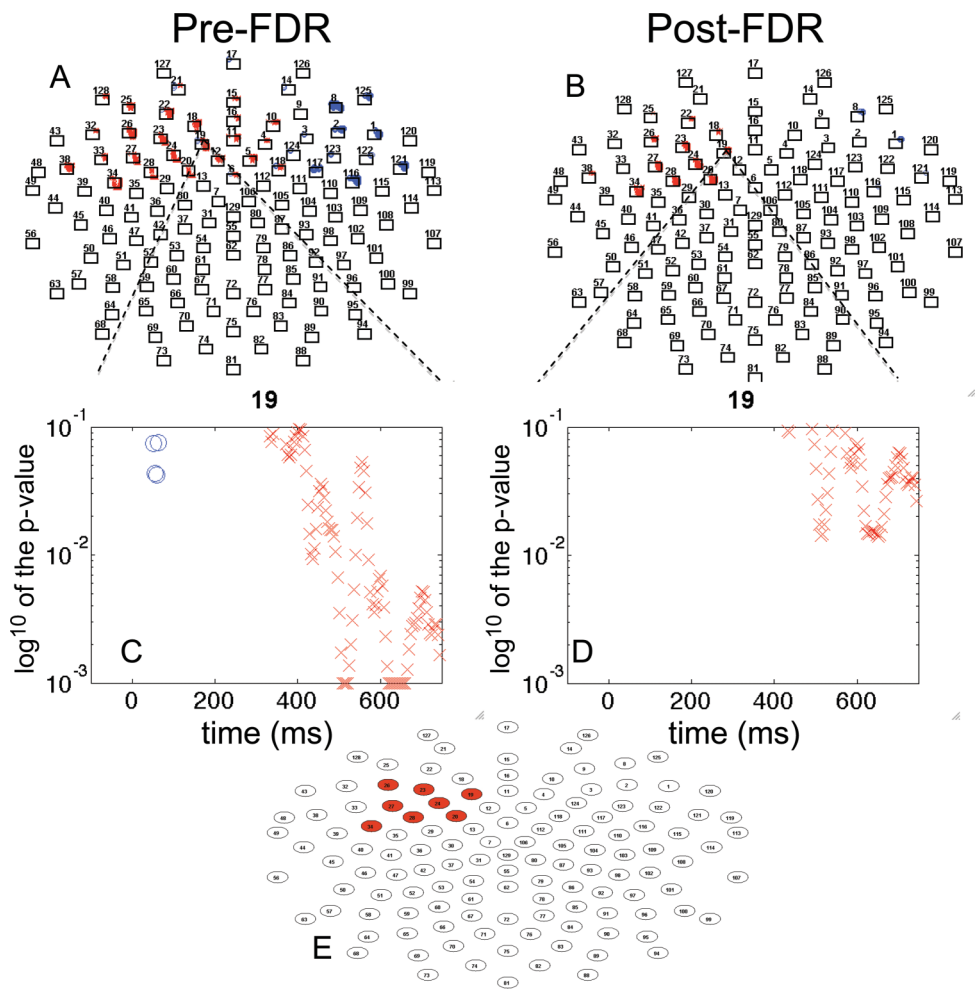
FIGURE 2 Analysis of the Cue Avoidance Task for frontal channels. (A) significant mass univariate *t*-tests plotted by *p*-value x time, $p < .05$; (B) false discovery rate (FDR) corrected univariate *t*-tests plotted by *p*-value by time; (C) single channel mass univariate *t*-tests plotted by *p*-value x time; (D) single channel FDR-corrected univariate *t*-tests plotted *p*-value x time; (E) channels remaining significant post-FDR (Approach A). [blue "o" indicates positive *t*-value; red "x" indicates negative *t*-value; *p*-value higher than 0.1 is not plotted; *p*-value lower than 0.001 is plotted on the 0.001 line]. (color figure available online)

less than this value were significant. With FDR, 555 tests were significant and 10 out of 39 frontal channels had a significant difference (for localization, see plot Figure 2B and Figure 2D for an enlarged view). To construct a temporal window for averaging, we conducted a PCA on the window of significant effects (channels 18, 19, 20, 22, 23, 24, 26, 27, 28, 34), "1" significant and "0," non-significant. This analysis resulted in one factor (Variance: 60.69%, Time Interval: 384–748 msec, Peak Time: 700 msec). Based on the factor score, we included only those channels that had the most influence on the factor and excluded the channels with a loading less than .5 (Table 1). In addition to having sub-threshold factor scores, channels 18 and 22 only had a short period of time significant (15–20 msec), these channels were excluded. The remaining 8 channels (19, 20, 23, 24, 26, 27, 28, 34) had at least a 200–250 msec period significant, during the late slow wave region for which where was a reliable, statistically controlled association between neural response and condition (Figure 2E). Across the effective time range for the 8 channels (384–748 msec), the paired samples t-test on the two responses was $t(33) = -5.38$, $p < .001$ (silence, mean $= 0.53$ μV, SE $= 0.43$ vs. 10 sec noise, mean $= 2.21$ μV, SE $= 0.43$). The average waveforms for silence versus escape, averaged across the eight left frontal channels, are presented in Figure 3.

Our second approach (Approach B) first involved conducting a temporal PCA to window the data in time. We conducted a temporal PCA on the 39 channels, which resulted in 4 temporal factors (Factor 1 Variance: 67.21%, 220– 748 msec, Peak 616 msec; Factor 2 Variance: 17.25%, 88–392 msec, Peak 204 msec; Factor 3 Variance: 7.12%, Time Interval: 4–208 msec, Peak 52 msec).

Next, we conducted t-tests on the mean voltages for each window across the silence condition and the 10 sec noise condition, separately for each channel. This yielded 117 tests (3 windows $\times$ 39 channels). Eighteen tests were significant without FDR across the three temporal windows (see Figure 1 for channel locations, Factor 1: channels 1, 8, 12, 18, 19, 20, 22, 23, 24, 26, 27, 28, 34, 38, 116; Factor 2: channels 34, 8; Factor 3: channel 8). After FDR correction, 8 significant channels (19, 20, 23, 24, 26, 27, 28, 34) were identified for the temporal factor 1 (220–748 msec) using a B-H false discovery rate of 0.05 the adjusted B-H critical value was $p = .0058$. None of the of the other factors had significant results. Across the effective time range (220–748 msec)

TABLE 1
Cue Avoidance Task PCA Factor Scores Used to Window Significant Channels

Channel	Factor Score
18	0.0853
19	0.9655
20	2.5236
22	0.0532
23	1.4864
24	2.4594
26	0.9131
27	1.8049
28	2.3366
34	2.2969
8	0.0546

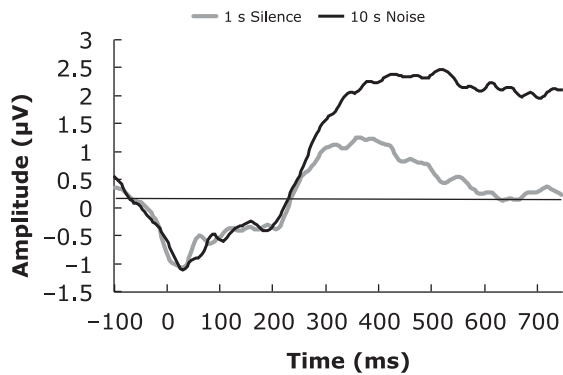*Note.* PCA = principal component analysis.

FIGURE 3 Cue avoidance task waveforms generated from channels selected via FDR selection procedures, which converged across Approach A and Approach B.

for the 8 channels, the paired samples *t*-test on the two response was $t(33) = -4.57$, $p < .001$ (silence, mean $= 0.62 \mu V$, SE $= 0.39$ vs. 10 sec noise, mean $= 1.94 \mu V$, SE $= 0.41$).

Exploratory Whole Head Analyses

We repeated the FDR portion of Approach A on the whole head for purposes of comparison with our examination of the frontal cortical region. The CAT task yielded many effects across the anterior and posterior regions (Figure 4A). Following FDR we see results comparable to those

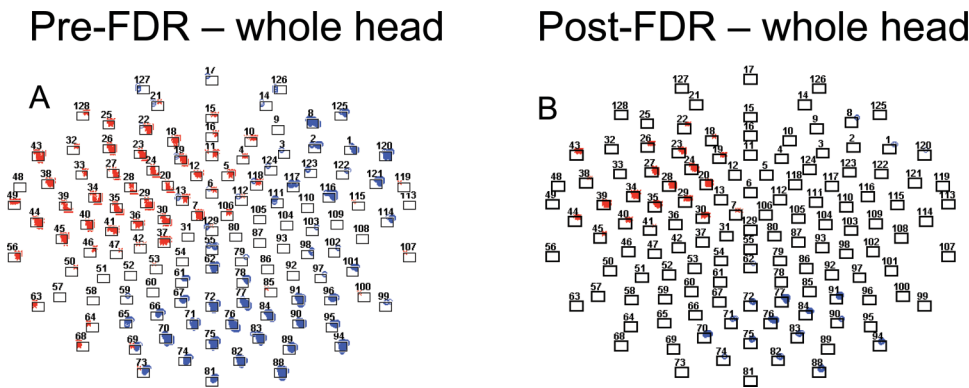FIGURE 4 Analysis of the Cue Avoidance Task for the whole head. (A) significant mass univariate *t*-tests plotted by *p*-value x time, $p < .05$; (B) FDR-corrected univariate *t*-tests plotted by *p*-value by time; blue "o" indicates positive *t*-value; red "x" indicates negative *t*-value. The shaded area indicates those electrodes that emerged from the analysis of the frontal region only. (color figure available online)

identified when analyses were isolated to the frontal region. The overlap of effects from the earlier analysis (frontal region) is highlighted in grey (Figure 4B).

Summary and Comparison: Study 1

Across Approach A and Approach B, we arrived at a comparable differentiation in terms of location for the difference of the silence and 10 sec noise escape conditions. Both localized to the left frontal region, comprising identical channels (34, 26, 27, 28, 23, 24, 19, 20). In terms of temporal window, the approaches yielded similar results, (Approach A: 384–748 msec vs. Approach B: 220–748), though Approach A, led to a more narrow temporal window. In terms of effect size for the aggregate effect across channels, Approach A (FDR first) yielded an effect size of .94, slightly larger than the effect size for Approach B, .79. We more broadly compare Approach A versus Approach B in the discussion.

STUDY 2

The purpose of this study was to illustrate the application of FDR to identify ERP-questionnaire associations for the frontal cortical region (Figure 1) for social rejection events and self-reported ostracism distress (Need Threat Scale). In childhood, peer rejection and exclusion directly impact a child's academic and social functioning (Buhs & Ladd, 2001; Ladd, Herald-Brown, & Reiser, 2008; Ladd & Troop-Gordon, 2003). Recently, social neuroscientists have begun to examine the neural correlates of social exclusion with a simple interactive game called Cyberball (Williams & Jarvis, 2006). Eisenberger et al. (2003) observe that social distress in response to exclusion on the Cyberball task is linearly related to fMRI blood-oxygen-level dependent (BOLD) signal in neural circuitry, common to both physical and social pain (Eisenberger, Jarcho, Lieberman, & Naliboff, 2006; Eisenberger & Lieberman, 2004; Panksepp, 2003). While neuroimaging studies to date average fMRI BOLD signal over an entire exclusion block in Cyberball (Eisenberger, Lieberman, & Williams, 2003; Eisenberger, Way, Taylor, Welch, & Lieberman, 2007; Masten et al., 2009; Sebastian et al., 2011), everyday experience would suggest that these interactions involve perception of rapidly unfolding events, occurring in a fraction of a second. ERPs hold the potential to move the discussion from overall rejection experiences as studied in fMRI, to rejection events and the timing of neural responses within these events as they track subjective experience.

Methods: Study 2

Participants included forty-nine healthy 8–13-year-old children ($M = 10.58$, $SD = 1.47$, 26 female) recruited from the community via a mass mailing. Children had no history of neurological impairment and had normal vision. One participant did not have sufficient trials (final $n = 48$). This sample includes a subset ($n = 33$) who were reported on previously (Crowley et al., 2010), but with a different data analytic approach. All procedures were conducted with the understanding and written consent of participants and with approval of the Yale School of Medicine Human Investigation Committee.

Procedure

Each child sat 24 inches before a 19-inch computer LCD monitor. Procedures for preparation and application of the Geodesic Sensor Net (GSN) of 128 Ag/AgCL electrodes (EGI, Inc.) and the parameters for EEG acquisition were identical to those used in Study 1. For the interactive Cyberball task used here, each child's electroencephalogram (EEG) and behavior were continuously monitored across the session such that stimulus presentation occurred only when the child was sitting still and looking at the monitor.

The Cyberball Social Exclusion Task.   In Cyberball, the participant ostensibly plays a computer-based ball-toss game with two other players who regularly toss the ball to the participant. Abruptly, the others exclude the participant, only throwing to one another. This exclusionary experience is distressing to participants, as per their self-reports of distress on a Need Threat Scale, described below (Crowley et al., 2010; Eisenberger et al., 2003; Williams, 2007). In this paradigm, participants were led to believe that they would be playing with two other children over the Internet. Children were presented with pictures of the two simulated players, who were matched for each participant's gender and ethnicity. Participants used a response pad to control to which player they threw the ball. (For a full description of the instructions, see Crowley et al., 2010). After the study, the child and parent were debriefed and informed that the other players were not real.

This ERP version of Cyberball consisted of 155 trials across two blocks, a fair play block (108 trials) and then an exclusion block (47 trials). During the 108-trial fair play block, the cyber-players threw to the participant 36 times. Throws to the participant were pseudorandom and predetermined within a list such that the participant waited for either 0, 1, 2, or 3 throws by the other players before receiving the ball again (frequency 12, 12, 10, and 2, respectively). During fair-play cyber-players threw to each another and not to the participant 36 times ("*not my turn*" events). The participant threw back to the other "players" for the remaining 36 trials. Seamlessly, fair play merged into a 47-trial exclusion block. This block represented 96% exclusion. Of the 47 rejection trials (exclusion block), the ball only came to the participant three times to maintain attention, Only 35 rejection events from this block were used in ERP analyses.

Immediately after the game, children completed the Need Threat Scale (Crowley et al., 2010; van Beest & Williams, 2006), a reliable and valid ostracism distress measure (Masten et al., 2009; Sebastian, Viding, Williams, & Blakemore, 2009; van Beest & Williams, 2006), which has been related to fMRI BOLD signal in previous research (Eisenberger et al., 2003). Children responded on the computer while still wearing the EEG cap. The 21-item Need Threat Scale, with simplified language for children (Crowley et al., 2010) gauges feelings of distress along four dimensions: belonging ("I felt rejected"), self-esteem ("I felt liked"), meaningful existence ("I felt invisible."), control ("I felt powerful"), on a 5-point choice, from "Not at all" to "Extremely." A majority of the research on the neural correlates of social exclusion relies on the sum of these four scales as an index of ostracism distress (higher scores indicated greater distress).

The EEG data was first processed through a 0.1 Hz first order high-pass filter and a 30 Hz low-pass filter. The data were then segmented into epochs containing a 100 msec pre-stimulus baseline and a 900 msec post-stimulus interval. Bad eye channels were manually marked and interpolated by surrounding channels. Next, artifact rejection was applied, in which bad segments (threshold 200 μV) were marked. Epochs with any eye blink or eye movement (threshold 150 μV) were rejected. Epochs with more than 10 bad channels (40% or more segments marked bad) were also

rejected. The remaining bad segments were replaced using spherical spline interpolation. Single trial data were re-referenced from the vertex (Cz) to an average reference of all electrodes because the latter was thought to be a better representation of a true zero (Junghofer, Elbert, Tucker, & Braun, 1999). The data was baseline corrected to the 100 msec pre-stimulus interval and single trial data was averaged respectively for rejection events over 900 msec (Figure 5). Participants
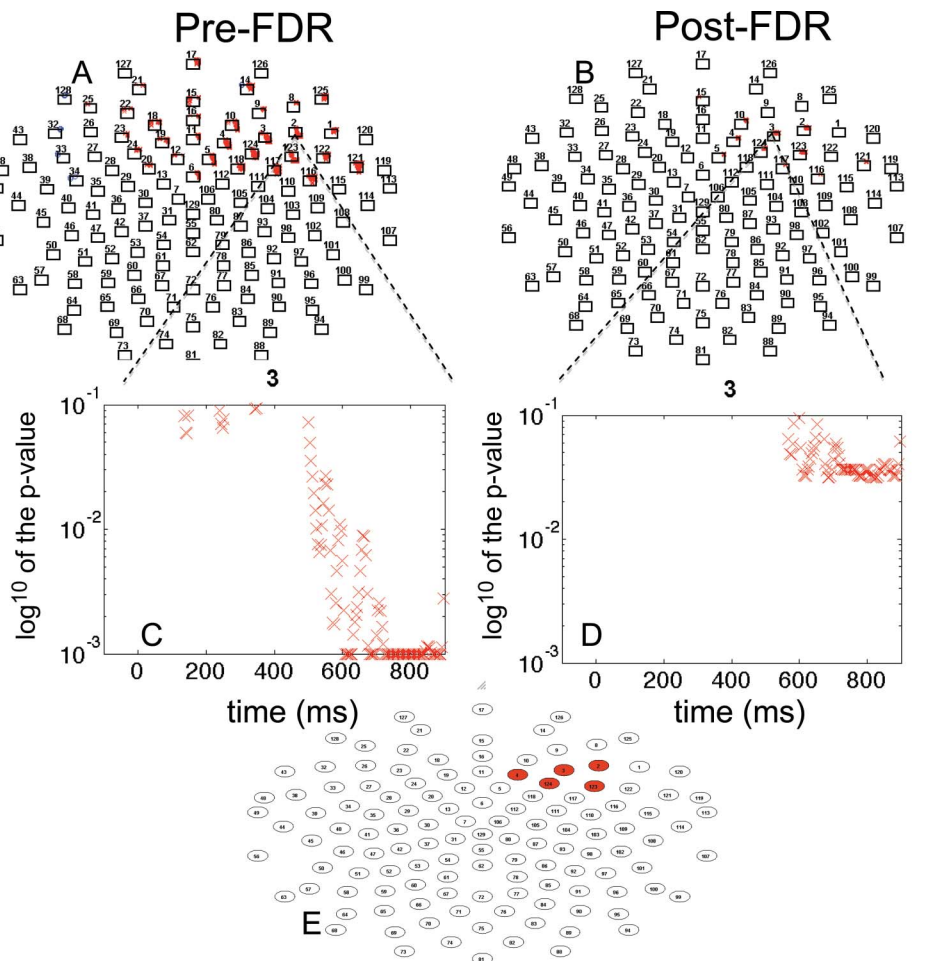
FIGURE 5 Cyberball rejection Events correlated with ostracism distress for frontal channels. (A) significant correlations plotted by *p*-value x time, *p* < .05; (B) FDR-corrected correlations plotted by *p*-value x time; (C) single channel correlation plotted by *p*-value x time; (D) single channel FDR-corrected correlations plotted *p*-value x time; (E) channels remaining significant post-FDR, (Approach A). [blue "o" indicates positive *t*-value; red "x" indicates negative *t*-value; *p*-value higher than 0.1 is not plotted; *p*-value lower than 0.001 is plotted on the 0.001 line]. (color figure available online)

providing at least 15 artifact free trials per condition were included ($n = 33$). Participants with fewer than 15 artifact free trials per condition received additional preprocessing with statistical eye-blink removal (blink threshold 14 μV/msec) (Gratton et al., 1983). Participants whose data yielded 15 good trials per condition with this additional approach ($n = 15$) were then included in the overall statistical analysis ($n = 48$).

Results: Study 2

To understand the relations between social rejection events and self-reported ostracism distress we employed two approaches that used FDR. As described earlier, in Approach A, we computed Pearson product moment correlations for each 4 msec data point and ostracism distress. There are 39 channels in the frontal region (Figure 1). Each channel had 225 data points (900 msec/4 msec per sample). Thus, the total number of tests was 8,775 (39*225). Before FDR there were 1,414 correlations with p-values below .05. Uncorrected effects are displayed on Figure 5A, plotted p-value (y-axis) by time (x-axis) for all channels. Within this cluster, one channel (3) is enlarged for viewing, for uncorrected (Figure 5C) and corrected effects (Figure 5D). After FDR correction, 326 tests were significant using a B-H false discovery rate of 0.05 (Figure 5B). The adjusted B-H critical value was $p = .0018$. We conducted a PCA on the window of significant effects, "1" significant and "0" non-significant. This provided a time window we could average for an ERP slow wave. A PCA on the window of the significant channels (2, 3, 4, 5, 10, 117, 121, 123, 124) yielded a single factor (Factor 1 Variance: 53.08%; Time Interval: 672–896 msec; Peak Time: 580 msec). Based on the factor score, we included only those channels that had the most influence on the factor and excluded the channels with a loading less than .5 (Table 2). This approach yielded five channels for which there were reliable, statistically controlled associations between ERP rejection events and self-reported ostracism (channels 2, 3, 4, 123, 124; Figure 5, E). As expected, the correlation between the average of these electrodes reflects the FDR-identified association between neural response to rejection and self-reported distress ($r = -0.55$, $p < .0001$, Figure 6A). We illustrate the ERP differences for the top 10 participants in the sample and the bottom 10 participants in Figure 6B.

TABLE 2
Cyberball Rejection Event PCA Factor Scores Used to Window Significant Channels

Channel	Factor Score
2	2.4629
3	2.1235
4	0.7287
5	0.3816
10	0.4578
117	0.1431
121	0.2007
123	2.5781
124	0.7535

Note. PCA = principal component analysis.
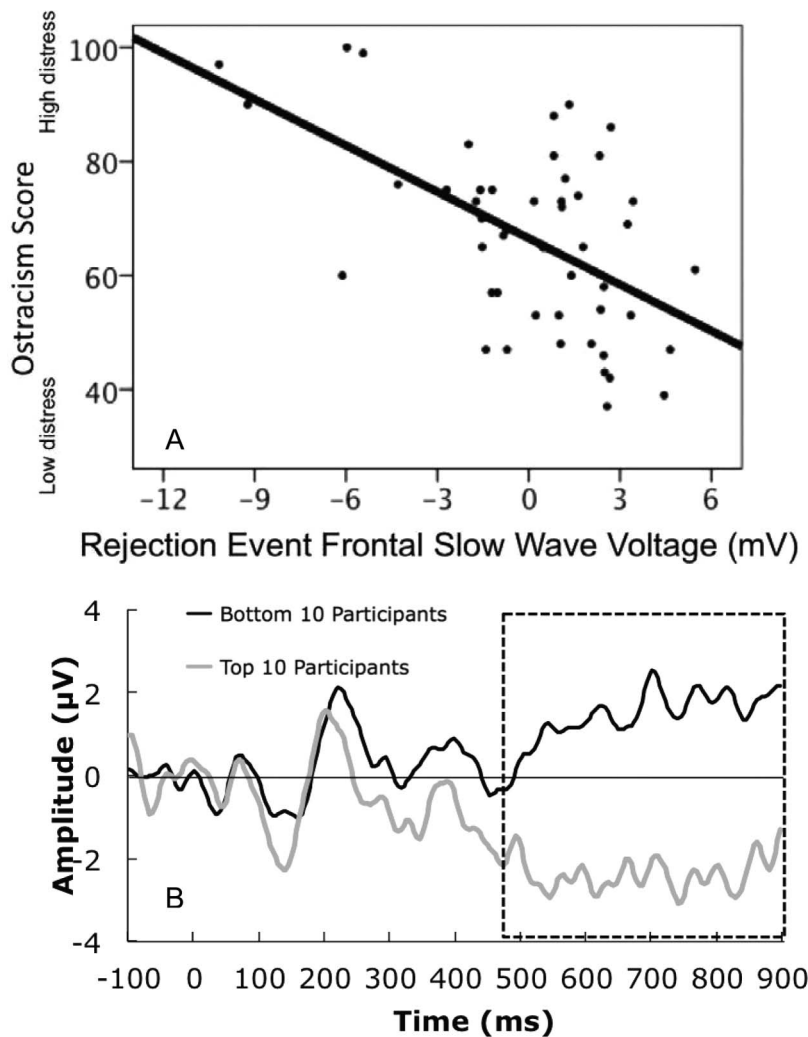
FIGURE 6 (A) plot of ostracism distress by rejection event frontal slow wave; (B) Plot of event-related potentials (ERPs) for participants with the bottom and top 10 scores for ostracism distress. (Data averaged over frontal channels identified with approach A.)

Our second approach (Approach B) first involved conducting a temporal PCA to window the data. Then we conducted Pearson correlations with the mean voltages for each window and each channel with the ostracism score. We conducted a temporal PCA on the 39 channels that resulted in 4 temporal factors (Factor 1 Variance: 48.07%, 448–900 msec, Peak 816 msec; Factor 2 Variance: 22.80%, 256–624 msec, Peak 412 msec, Factor 3 Variance: 7.56%, Time Interval: 172–332 msec, Peak 228 msec, Factor 4 Variance: 6.29%, 24–212 msec. Peak 144 msec).

For each temporal factor, we correlated the mean amplitude on each channel with the ostracism distress questionnaire. This yielded 156 correlations (4×39). Twenty-three tests were significant without FDR across three of four factors (Factor 1: channels 1, 2, 3, 4, 5, 10, 14, 17, 19, 116, 117, 118, 121, 122, 123, 124, 125; Factor 2: channels 2, 3, 121, 123, 124; Factor 3: channel 19). After FDR correction, six significant channels (2, 3, 4, 117, 123, 124) were identified in temporal factor 1 (448–900 msec) using a B-H false discovery rate of .05. The adjusted B-H critical value was $p = .0058$. None of the other factors had significant results. To confirm a comparable pattern with ostracism distress, we took the mean amplitude for factor 1 averaged on the over significant channels and correlated this measure with ostracism distress, yielding a nearly identical pattern to Approach A, $r = -0.51$, $p < .001$, $n = 48$.

Summary and Comparison: Study 2

Across Approach A and Approach B, we arrived at a comparable association between neural response to a rejection event and perceived ostracism distress. Approach A identified right frontal channels comparable to Approach B (2, 3, 4, 123, 124 vs. 2, 3, 4, 117, 123, 124). In terms of temporal window, the approaches yielded similar results, (672–896 vs. 448–900), reflecting a slow wave. Lastly, the association between neural responses and ostracism distress, an estimate of effect size, was also comparable ($r = -.55$ vs. $-.51$). Of note, we did conduct a massive screening of the whole head with the data from study 2 but no comparisons survived and FDR correction. The relative merits of Approach A versus Approach B are elaborated on in the discussion below.

DISCUSSION

We applied false discovery to guide data exploration in two studies, the first employing a new paradigm and the second linking a psychological trait to neural response. To this end, we provided examples of an FDR method for data mining. In each of these studies we compared the two approaches. In Approach A, tests were first made for all data points (*t*-test, correlation), followed by a temporal PCA to window the data across significant effects. In Approach B, we first conducted a temporal PCA, followed by tests for all PCA-derived windows. As expected, temporally windowing the data first with a PCA led to fewer tests overall (Approach B). Yet the localization of the observed effects were identical across approaches in Study 1 (left frontal, 8 of 8 channels) and nearly identical in Study 2 (right frontal, 5 of 6 channels). Approach A favored the comparison of interest (condition in Study 1, trait-neural response association in Study 2), resulting in a more narrow temporal ERP window of effect across both studies (164 msec difference in Study 1, 228 msec difference in Study 2). Despite the larger temporal window observed with Approach B, the size of the aggregate effect was similar across approaches for both studies (condition Study 1; correlation Study 2).

We also applied a massive screening of the whole head in an exploratory fashion with B-H FDR under Approach A only. Here the effects diverged across Study 1 and Study 2. The data from Study 1 yielded whole-head effects that encompassed those observed in the frontal region analyses. We cannot ascertain the extent to which negative dependence in the data contributed to the results. Strikingly, the data from Study 2 failed to reach statistical significance for any effects

when BH-FDR was employed globally. This outcome shows that BH-FDR is not immune to the multiplicity problem and may warrant a compromise as we started within only looking at the frontal cortical region.

Approach A Versus Approach B

In comparing Approach A and Approach B, there appears to be a slight advantage in terms of aggregate effect, for Approach A. This makes sense in light of the fact that comparisons or correlations happen first in Approach A. Only the neural data that yields FDR-corrected significant differences has a chance to be included in the PCA. Whereas when neural data are aggregated first, some neural data will be included which covaries temporally in the ERP but does not relate to a comparison difference or correlation, yielding a slightly smaller effect overall. In our view, both approaches are valid, and the preference of the investigator in relation to the goal of the study should determine which method is employed. If the investigator is mainly interested in detecting condition differences or relating neural response to other external criteria, then Approach A may be more appropriate. In particular, two conditions may not differ along the full range of a component, or a psychological process may not relate to a neural response along the full range of a component. Approach A affords the investigator the opportunity to account for this possibility. On the other hand, if the investigator prefers to let covarying temporal properties of the neural signal drive their data-aggregation, which *may* be more replicable and certainly more comparable across different external correlates (behavior vs. questionnaire), Approach B may be preferred. It is important to note that we provide only two data examples here. Other data sets may yield different results, possibly with greater divergence across approaches. For instance, the reliability of an external criterion (questionnaire or behavior) could affect the results differentially across Approaches A and B.

Brief Discussion of Study Findings

Study 1 documented a difference in left frontal slow wave neural activity for a behavior that terminated an aversive noise. Terminating an aversive stimulus produced a greater left frontal positive potential compared to the same behavior executed during silence. We designed this paradigm to examine neural responses to escape behavior as a negative reinforcement process. Given this initial pilot project, we can now begin to examine the extent to which the magnitude of this neural response tracks degree of perceived relief and the extent to which this neural response tracks individual differences in anxiety and addiction vulnerability (Zvolensky, Gonzalez, Bonn-Miller, Bernstein, & Goodwin, 2008).

Study 2 documented an association between a late positive neural response for rejection events and self-reported distress. Individuals reporting greater distress had a more negative frontal slow wave for rejection events during an exclusion experience. We believed that rejection events in exclusion would track ERP rejection events, but we were unsure as to where on the scalp these effects would emerge. Having employed FDR to document this association we are now directing our efforts toward youth with vulnerabilities such as social anxiety, who are likely to respond differently to social exclusion (Zadro, Boland, & Richardson, 2006).

Future Directions

We began our discussion by employing the B-H false discovery method when examining a new paradigm or trait-neural response relation. A similar approach could also be valuable for examining widely used paradigms over extended scalp regions. It is not uncommon for an ERP study to collect data from many more electrodes than are analyzed and/or reported in a manuscript—interesting aspects of the data may therefore be going unanalyzed and unreported. For instance, face processing studies of emotion often report on the N170 response posterior temporal electrode regions, yet frontal responses may also provide useful information about stimulus conditions and individual differences.

Our investigations yield several implications for which data sets are best suited to FDR analysis. For instance, the B-H FDR method we employed is thought to be less effective for negative dependent relationships. We therefore used FDR to examine only the frontal region, as posterior and frontal effects commonly emerge as dipoles resulting in a negative dependent relationship.

Many ERP researchers will recall reading a manuscript that compares neural responses across clinical groups in which the ERP waveform morphology looks different, yet these differences fail to meet thresholds for statistical significance. Or, because the manuscript examines well-known ERP responses, some other aspect of the ERP record that does show a group differences is left untested because there is no experimental precedent to do so. In these cases, an FDR approach analogous to that detailed above could be used to window the data where the group difference is maximal or to explore the data more thoroughly. Admittedly, such an endeavor can extend manuscript length, but with the widespread inclusion of supplementary materials, this should not deter investigators, reviewers, or editors from encouraging such practices.

Limitations

Our method and findings should be qualified by several limitations, but also optimism for discovery approaches. First, we observed effects only for slow wave neural responses. These tend to be effects with longer response windows, requiring fewer trials to produce effects. In practice, we do not know how our approach would perform with more traditional ERP peaks. However, the real data examination of the P300 with FDR provided by Lage-Castellanos et al. (2010) suggests the approach would work here as well.

Second, we employed the FDR approach only for the frontal region. From the perspective of the B-H FDR method, posterior and frontal effects commonly emerge as dipoles resulting in a negative dependent relationship across tests. The B-H FDR approach we chose is thought to be less effective for negative dependency situations. When we did apply a whole head analysis in Study 1, we observed that the frontal channels we initially identified were still among those identified. We also observed a set of posterior channels consistent with a dipole effect. Thus, while negative dependency may be of concern from a theoretical perspective, in practice our data from study 1 suggest we would come to a similar conclusion about the frontal region effect.

Third, we applied only one FDR approach. While the general method we employ here could be adapted to other FDR approaches and other multiple comparison control approaches, employing similar plots and temporal windowing with PCA, we cannot generalize past the B-H FDR method employed here. FDR and related discovery approaches hold great promise to take a step toward getting more from our data and moving our science in new directions.

REFERENCES

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, *125*(1–2), 279–284.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188.

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

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: An examination of mediating processes. *Developmental Psychology*, *37*, 550–560.

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.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology* (2nd ed.). Cambridge, England ; New York, NY: Cambridge University Press.

Chapman, R. M., & Bragdon, H. R. (1964). Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature*, *203*, 1155–1157.

Crowley, M. J., Wu, J., Bailey, C. A., & Mayes, L. C. (2009). Bringing in the negative reinforcements: The avoidance feedback-related negativity. *Neuroreport*, *20*, 1513–1517.

Crowley, M. J., Wu, J., Molfese, P. J., & Mayes, L. C. (2010). Social exclusion in middle childhood: Rejection events, slow-wave neural activity, and ostracism distress. *Social Neuroscience*, *5*, 483–495.

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society London B Bioliological Science*, *363*(1511), 3787–3800.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, *99*(465), 96–104.

Eisenberger, N. I., Jarcho, J. M., Lieberman, M. D., & Naliboff, B. D. (2006). An experimental study of shared sensitivity to physical pain and social rejection. *Pain*, *126*(1–3), 132–138.

Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: A common neural alarm system for physical and social pain. *Trends in Cognitive Sciences*, *8*(7), 294–300.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science*, *302*(5643), 290–292.

Eisenberger, N. I., Way, B. M., Taylor, S. E., Welch, W. T., & Lieberman, M. D. (2007). Understanding genetic risk for aggression: Clues from the brain's response to social exclusion. *Biological Psychiatry*, *61*(9), 1100–1108.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15*(4), 870–878.

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.

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

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

Hansenne, M. (2000a). The p300 cognitive event-related potential. I. Theoretical and psychobiologic perspectives. *Clinical Neurophysiology*, *30*(4), 191–210.

Hansenne, M. (2000b). The p300 cognitive event-related potential. II. Individual variability and clinical application in psychopathology. *Clinical Neurophysiology*, *30*(4), 211–231.

Junghofer, M., Elbert, T., Tucker, D. M., & Braun, C. (1999). The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*, *110*(6), 1149–1155.

Kuhn, T. S. (1996). *The structure of scientific revolutions*, 3rd edition. Chicago, IL: University of Chicago Press.

Ladd, G. W., Herald-Brown, S. L., & Reiser, M. (2008). Does chronic classroom peer rejection predict the development of children's classroom participation during the grade school years? *Child Development*, *79*, 1001–1015.

Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development*, *74*, 1344–1367.

Lage-Castellanos, A., Martinez-Montes, E., Hernandez-Cabrera, J. A., & Galan, L. (2010). False discovery rate and permutation test: An evaluation in ERP data analysis. *Statistics in Medicine*, *29*(1), 63–74.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

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

Masten, C. L., Eisenberger, N. I., Borofsky, L. A., Pfeifer, J. H., McNealy, K., Mazziotta, J. C., & Dapretto, M. (2009). Neural correlates of social exclusion during adolescence: Understanding the distress of peer rejection. *Social Cognitive and Affective Neuroscience*, *4*, 143–157.

Panksepp, J. (2003). Neuroscience. Feeling the pain of social loss. *Science*, *302*(5643), 237–239.

Polich, J. (1993). Cognitive brain potentials. *Current Directions in Psychological Science*, *2*(6), 175–179.

Polich, J. (2004). Clinical application of the P300 event-related brain potential. *Physical Medicine and Rehabilitation Clinics of North America*, *15*(1), 133–161.

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

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185.

Polich, J., & Ochoa, C. J. (2004). Alcoholism risk, tobacco smoking, and P300 event-related potential. *Clinical Neurophysiology*, *115*(6), 1374–1383.

Qian, H. R., & Huang, S. (2005). Comparison of false discovery rate methods in identifying genes with differential expression. *Genomics*, *86*(4), 495–503.

Schiller, D., Levy, I., Niv, Y., LeDoux, J. E., & Phelps, E. A. (2008). From fear to safety and back: Reversal of fear in the human brain. *Journal of Neuroscience*, *28*(45), 11517–11525.

Sebastian, C., Viding, E., Williams, K. D., & Blakemore, S. J. (2009). Social brain development and the affective consequences of ostracism in adolescence. *Brain and Cognition*, *72*(1), 134–145.

Sebastian, C. L., Tan, G. C., Roiser, J. P., Viding, E., Dumontheil, I., & Blakemore, S. J. (2011). Developmental influences on the neural bases of responses to social rejection: Implications of social neuroscience for education. *Neuroimage*, *57*(3), 686–694.

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*, *8*, 1234–1240.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *64*, 479–498.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(700), 1187–1188.

van Beest, I., & Williams, K. D. (2006). When inclusion costs and ostracism pays, ostracism still hurts. *Journal of Personality and Social Psychology*, *91*(5), 918–928.

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, *58*, 425–452.

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, *38*(1), 174–180.

Zadro, L., Boland, C., & Richardson, R. (2006). How long does it last? The persistence of the effects of ostracism in the socially anxious. *Journal of Experimental Social Psychology*, *42*(5), 692–697.

Zani, A., Proverbio, A. M., & Posner, M. I. (Eds.). (2002). *The cognitive electrophysiology of mind and brain*. New York, NY: Academic Press.

Zvolensky, M. J., Gonzalez, A., Bonn-Miller, M. O., Bernstein, A., & Goodwin, R. D. (2008). Negative reinforcement/negative affect reduction cigarette smoking outcome expectancies: Incremental validity for anxiety focused on bodily sensations and panic attack symptoms among daily smokers. *Experimental and Clinical Psychopharmacology*, *16*(1), 66–76.