

The ERP, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study

Chad C. Williams  | Thomas D. Ferguson | Cameron D. Hassall | Wande Abimbola | Olave E. Krigolson 

Centre for Biomedical Research, University of Victoria, Victoria, BC, Canada

Correspondence

Chad C. Williams, Centre for Biomedical Research, University of Victoria, P.O. Box 17000 STN CSC, Victoria, BC V8W 2Y2, Canada.

Email: ccwillia@uvic.ca

Funding information

Chad Williams would like to acknowledge support from Natural Sciences and Engineering Research Council of Canada (CGSD3-518118-2018) and all authors would like to acknowledge support from the Natural Sciences and Engineering Research Council of Canada (RGPIN 2016-094).

Abstract

Human learning, at least in part, appears to be dependent on the evaluation of how outcomes of our actions align with our expectations. Over the past 23 years, electroencephalography (EEG) has been used to probe the neural signatures of feedback processing. Seminal work demonstrated a difference in the human event-related potential (ERP) dependent on whether people were processing correct or incorrect feedback. Since then, these feedback evoked ERPs have been associated with reinforcement learning and conflict monitoring, tied to subsequent behavioral adaptations, and shown to be sensitive to a wide range of factors (e.g., Parkinson's disease). Recently, research has turned to frequency decomposition techniques to examine how changes in the EEG power spectra are related to underlying learning mechanisms. Although the literature on the neural correlates of feedback processing is vast, there are still methodological discrepancies and differences in results across studies. Here, we provide reference results and an investigation of methodological considerations for the ERP (reward positivity) and frequency (delta and theta power) correlates of feedback evaluation with a large sample size. Specifically, participants ($n = 500$) performed a two-armed bandit task while we recorded EEG. Our findings provide key information about the data characteristics and relationships that exist between the neural signatures of feedback evaluation. Additionally, we conclude with selected methodological recommendations for standardization of future research. All data and scripts are freely provided to facilitate open science.

KEYWORDS

cognitive control, EEG, ERPs, error processing, learning, oscillation/time frequency analyses

1 | INTRODUCTION

1.1 | Electroencephalographic measures of feedback processing

The initial examination of the electroencephalographic (EEG) correlates of feedback processing began with work by Falkenstein and colleagues (1991) and concurrently by Gehring and colleagues (1990, 1993). They reported

a negative deflection in the human electroencephalogram evoked by errors committed during the performance of a speeded response task—an event-related potential (ERP) component now named the error-related negativity (ERN). Based on these data, the authors concluded that the ERN reflected the online monitoring of one's performance.

Following from this seminal work, studies began to examine feedback processing—the adaptation of behavior from external feedback. The ERP correlates of feedback

processing were first reported by Miltner et al. (1997) who found a frontal component differentially sensitive to positive and negative feedback—a component that was eventually named the feedback-related negativity (FRN)¹. They concluded that this “new” component was the feedback evoked version of the aforementioned ERN and reflected a generic neural system for performance evaluation and learning.

An abundance of later work suggested that the FRN should be more aptly named the reward positivity. Specifically, the initial supposition made by Miltner et al. (1997) was that the FRN reflected modulation of the incorrect (or negative) outcome ERP waveform. However, subsequent work instead demonstrated that the difference in the FRN time range appeared to be driven by a modulation of the correct (or positive) outcome waveform and was thus renamed the reward positivity (Holroyd et al., 2008; Proudfit, 2015).

The functional significance of the reward positivity² (or FRN), however, is still a matter of debate. Early interpretations of the reward positivity were divided between conflict monitoring and reinforcement learning accounts (although see Holroyd & Yeung, 2012, and Holroyd & McClure, 2015). The conflict monitoring account proposed by Botvinick and colleagues (e.g., Botvinick et al., 2004) posited that the reward positivity (and the ERN) were sensitive to the processing of conflict as the brain decides whether or not top-down control is required. Conflict may arise, for example, when responses in a given task compete with each other. Response conflict can be seen in frontal-central neural signals such as the reward positivity, which is modulated by the extent of top-down influence that is required by a given response.

An alternate theoretical account of the reward positivity proposes that it reflects a reinforcement learning prediction error—the computational difference between one's expectations of an outcome and the actual outcome (Holroyd & Coles, 2002). Specifically, if a discrepancy between an expectation and outcome exists, the signal generated by this computation is used to update representations and adapt subsequent behavior (Holroyd & Krigolson, 2007; Luft, 2014; Walsh & Anderson, 2012; Williams et al., 2020). Although changes in the amplitude of the reward positivity were originally theorized to reflect a neural signal that outcomes are worse-than-expected (e.g., a reward was smaller than expected; Holroyd & Coles, 2002; Miltner et al., 1997), reinforcement learning accounts have been modified to propose

that the signal is observed when outcomes are better-than-expected (e.g., a reward was bigger than expected; Foti et al., 2011; Holroyd et al., 2008; Krigolson, 2018; Krigolson et al., 2014; Proudfit, 2015).

More recent developments, however, have led to several additional theories including ones that propose that the reward positivity reflects independent aspects of both the processing of better-than-expected and worse-than-expected outcomes—propositions derived from the use of frequency decomposition methods as opposed to the ERP technique (Bernat et al., 2008, 2011, 2015; Foti et al., 2015). The initial reason for turning to frequency decomposition techniques to examine the reward positivity was primarily due to its temporal overlap with the P300 (Bernat et al., 2008, 2011, 2015), a component that has also been shown to play a role in feedback processing (Yeung & Sanfey, 2004). Using frequency decomposition techniques, Bernat and colleagues (2011, 2015) concluded that changes in the amplitude of the reward positivity were associated with changes in theta power and were driven by loss feedback, whereas changes in the amplitude of the P300 were associated with changes in delta power and were driven by gain feedback. These findings demonstrated a dissociation of overlapping feedback processing ERP components and the independent processing of better-than-expected and worse-than-expected events.

Adopting this perspective would mean that the “reward positivity” naming convention is misleading and may bias interpretations of findings as only reflective of gain feedback processing. Although we will not here contrast this perspective with the conceptual framework of the reward positivity, it may be necessary to be cautious when reading this article by remembering that the reward positivity may be thought of as an ERP component reflecting both gain and loss feedback processing.

1.2 | Methodological considerations of neural feedback processing

To date, Sambrook and Goslin (2015) provide the most rigorous look at methodological considerations and differences in results for the reward positivity in a meta-analysis. Specifically, they included 55 data t-sets that investigated the effects of valence, reward magnitude, and reward likelihood on the ERP correlates of feedback processing. Interestingly, the amplitude of the reward positivity across these studies was strikingly variable. For example, when investigating the effects of likelihood, grand averaged amplitudes ranged from ~ -0.60 to $\sim +4.75 \mu V^3$. The inconsistency of amplitudes

¹It is important to note that the FRN has also been called the feedback negativity (FN), the medial frontal negativity (MFN), the feedback ERN (fERN), and more recently the RewP or reward positivity.

²For the remainder of this manuscript we will solely use the term reward positivity.

³Sambrook and Goslin (2015) measured a feedback related negativity thus actually found a range of $\sim -4.75 \mu V$ to $\sim +0.60 \mu V$, however, we flipped the polarity to be consistent with the reward positivity. These values are approximated via visual inspection of their Figure 5.

across studies is rarely addressed and often construed as noise derived by the EEG method. In reality, much of this variation is caused by the unstandardized practices of quantifying the amplitude of the reward positivity and also independent variables that have an impact on component amplitude but are not examined.

There are three main methods to quantify an ERP component: base-to-peak, maximum peak, and mean peak (Luck, 2014). For the reward positivity, base-to-peak refers to measuring a difference between the P200 component (a component that precedes the reward positivity) and the N200 component (a component that is modulated to elicit the reward positivity). A maximum peak instead determines the maximal amplitude of the reward positivity peak within a designated time window. A mean peak reflects the mean voltage within a window surrounding the maximal amplitude of the reward positivity peak. It is evident that no consensus on which method to use exists as Sambrook and Goslin (2015) found nearly equal occurrences of each method across the studies included in their meta-analysis.

The lack of consensus with regard to peak detection is surprising given that researchers have previously described the limitations of base-to-peak and maximum peak measures (Luck, 2014; Luck & Gaspelin, 2017; Picton et al., 2000). Specifically, the base-to-peak method measures the peak of interest against the preceding peak and thus actually measures the difference between the two (Picton et al., 2000). If the preceding peak is also affected by the experimental manipulation, then this method does not adequately capture the component of interest. Alternatively, the maximum peak method is susceptible to noise (e.g., unrelated biological and environmental noise), which artificially inflates effects (Luck, 2014; Luck & Gaspelin, 2017). Neither of these methods should then be common in the literature, yet they are. The mean peak, in contrast, is less susceptible to confounding effects of other peaks and noise (Luck, 2014; Luck & Gaspelin, 2017). Here, we pit these three peak quantification methods—base-to-peak, maximum peak, and mean peak—against each other to provide a firm statement about the superiority of the mean peak method and resolve this source of variability in ERP research. Unfortunately, even confirming the superiority of mean peak measures over the other peak methods does not remove all sources of variation across research. For example, the time window used for peak detection has also been demonstrated to impact peak quantification (Sambrook & Goslin, 2015).

When measuring the reward positivity, it is also essential to adequately capture the peak of the component. Thus, perhaps some variability around component timing is warranted, but not to the degree as seen in past studies. For example, measuring the reward positivity amplitude ranges from using a 200–250 ms time window to a 240–400 ms time window (Sambrook & Goslin, 2015). Not only do some of these time

windows not overlap at all, but the size of the windows vary from 50 to 160 ms. Fortunately, Sambrook and Goslin (2015) do put forth a potential window timing of 270 to 300 ms but caution that the variability across studies may rather warrant a time window of 240 to 340 ms. Although a good guide, their caution demonstrates that there is still no consensus on the time window that should be used to quantify the reward positivity.

A potential cause for inconsistencies in both the amplitudes and the time windows of reward positivity analyses may be due to variability in component scalp topography or, more precisely, the electrode site of investigation. Again considering those studies included within Sambrook and Goslin (2015), three electrode sites were used when investigating the reward positivity: FCz, Cz, and Fz (or a pool of these and other electrodes). For those familiar with the literature of the reward positivity, it is not surprising that electrode site FCz is the most commonly used. It has not become the standard, however, as is demonstrated by the frequent use of other electrode sites. With that said, these three electrode sites are all proximal and thus there is some agreement as to where the reward positivity peaks topographically. Yet, to ensure that the methodology of research going forth is consistent, it is still necessary to reach a consensus on which electrode site to analyze.

One final point to note, although not a main focus of this article, is whether to measure conditional waveforms (gain and loss) or the difference waveform (gain-loss). Especially with the reward positivity, a common practice is to measure the difference waveform but this is far from convention. Indeed, some of the discrepancies in reported results may rise as there are certain cases where measuring individual waveforms are more telling or appropriate than measuring the difference waveform. For example, when considering the learning effects of the reward positivity, research has focused solely on the gain waveform (Krigolson et al., 2014; Williams et al., 2020, 2018). With that said, Luck (2014) proposed that difference waveforms are more informative than individual conditional waveforms. In line with Luck, we will focus our discussion on difference waveform findings, but to be thorough we still provide results for conditional waveforms as well.

1.3 | Methods of measuring neural feedback processing

There exists a variety of methods for assessing neural activity with EEG—in this manuscript, to ensure a full complement of analyses, we will consider ERP, fast Fourier transforms (FFT), and time–frequency transforms. ERP methods extract neural activity by observing the peaks and troughs of averaged EEG signals time-locked to events of interests

(see Luck, 2014 for more). FFT and time–frequency analyses rather dissect these EEG signals into underlying brain oscillations, which are used for analysis. Whereas FFT analyses extract frequency activity collapsed across time, time–frequency analyses extract frequency activity extended over time (see Cohen, 2014 for more).

ERP methods have long been the staple in EEG research and have been the vehicle for significant discoveries and achievements. The original widespread adoption of ERP methods over frequency analyses was, in part, because they were intuitive and easy to conduct (especially as EEG was discovered before computers; Luck, 2014). Now, however, a multitude of toolboxes and tutorials have made frequency analyses easy and accessible (e.g., Cohen, 2014). The rise in adopting frequency analyses over ERP methods was driven, in part, by their ability to disentangle underlying cognitive components that are confounded within ERP research (e.g., Bernat et al., 2008). There are different approaches to frequency analyses, and deciding between FFTs and time–frequency analyses, for example, involves whether researchers are regarding the timing of oscillations. Whereas FFT methods ignore the dimension of time, time–frequency analyses do not.

The removal of time as a factor has two advantages: first, it reduces researchers' degrees of freedom in that the timing of oscillations do not need to be decided, and second, it is computationally faster. The former allows for easier standardization across research, while the latter allows for more diverse applicability of research—for example, with real-time monitoring of brain activity. With an interest in component timing, however, researchers may adopt time–frequency analyses or short-time Fourier transforms (the application of FFTs with sliding time windows)—methods that have been shown as equivalent when parameters are equal (Bruns, 2004). Thus, either frequency decomposition technique may be appropriate depending on the research design and goals of the researcher. Here, we conduct ERP, FFT, and time–frequency analyses to assess the activity of neural learning systems.

1.4 | Current study

The study of neural signals associated with feedback processing is of interest across theoretical, applied, and clinical neuroscience; thus it is important to define their characteristics. Here, we (1) determine normative neural data of feedback processing using a large sample in a single task, (2) compare different EEG processing methods (ERP, FFT, and time–frequency), and (3) assess selected methodological standards. Although researchers know that the neural signals associated with feedback evaluation (i.e., the reward positivity, delta power, and theta power) are not present in everyone, to what extent is this true? Here, we seek to answer this key

question. Indeed, normative data facilitate research, the understanding of individual differences, and the assessment of clinical populations (Hill et al., 2018; Mitrushina et al., 2005; Zimmerman, 2011). Understand normative scores in these neural components can lead to better interpretation of our findings and aid in designing future research. Large sample studies afford us the knowledge to properly understand and interpret the variability we see in our data—mitigating the biased assumption that uncharacteristic scores reflect noise. Furthermore, describing normative patterns in our data can inform the interpretation of individual factors of learning across contexts and populations (Hill et al., 2018), and the assessment of clinical populations via biomarkers—for example, Proudfit (2015) has described that neural measures of feedback processing can gauge depression. Understanding and interpreting data, however, moves beyond simply describing it. There exists a variety of methods for processing EEG data, each with advantages and disadvantages but none with standardized methods. Here, we (1) present normative data of ERP, FFT, and time–frequency methods, (2) contrast EEG analysis methods to determine how they relate, and (3) conclude with selected methodological considerations.

In sum, we sought to provide an accurate definition of the reward positivity and related frequency components (delta and theta power) based on a large sample size of participants ($n = 500$) performing the same experimental task. We chose to use a two-armed bandit task as it is the most commonly used paradigm to evoke the reward positivity and has been demonstrated to have clinical relevance (e.g., Proudfit, 2015). Moreover, this article focuses on the descriptive measures of the reward positivity, delta power, and theta power, on recommendations for standardized practices, and on comparisons across processing methods, rather than theoretical interpretations of our findings. For theoretical interpretations of the reward positivity, we would point readers to a recent journal special edition on reward and feedback processing (Foti & Weinberg, 2018) and a series of review articles (Proudfit, 2015; Sambrook & Goslin, 2015; Walsh & Anderson, 2012). For theoretical interpretations of delta power and theta power in feedback processing, see Bernat et al. (2015) and Sambrook and Goslin (2016).

2 | TASK AND DESIGN

2.1 | Task

Participants were seated in front of a 19" LCD computer monitor in a sound dampened, dark room. They responded to the task (originally written in MATLAB version R2016a using Psychophysics Toolbox extension version 3.0.12; Brainard, 1997) using either a 5-button handheld RESPONSEPixx controller or a dual handheld

RESPONSEPixx controller (VPixx, Vision Science Solutions, Quebec, Canada).

Participants completed a simple gambling task. On each trial, they saw a black fixation cross for 500 ms, followed by two colored squares for 500 ms, and then, the fixation cross turned gray (go cue) and participants were to select one of the two squares (square locations—left, right—were randomized on each trial) within a 2,000 ms time limit. They were then presented with a black fixation cross for 300 to 500 ms, and then, simple feedback as to their performance (“WIN” for gain, “LOSE” for loss) for 1,000 ms in black font. If the participants responded before the go cue they were instead delivered “TOO FAST” feedback and if they did not respond before the 2,000 ms time limit, it would be considered a loss. The goal of the participants was to accumulate wins by determining which of the two squares would more often lead to gains (60% vs. 10%). In this task, participants accumulated wins; however, were not paid money. They would see the same pair of colors for one block of 20 trials. They conducted six blocks of unique color pairs.

Within this task, we found that responses were correct 75% [73%, 76%] of the time which, due to the probabilistic nature of the feedback, resulted in 47% [46%, 48%] gain feedback⁴. Correct responses (383 ms [372, 394 ms]) were statistically quicker than incorrect responses (406 ms [393, 418 ms]), $M_d = -23$ ms [-30, -15 ms], $t(499) = 5.84$, $p < .0001$, $d = -0.26$.

2.2 | Participants

Five hundred undergraduate students (Age: 490 responses, $M = 21.71$ years old [21.24 years old, 22.18 years old]; Gender: 495 responses, Female = 341, Male = 154; Handedness: 492 responses, Right Handed = 440, Left Handed = 52) were included within the analyses of this manuscript and were recruited via the University of Victoria psychology participant pool. In addition, birthplace as self-reported by participants were collected beginning mid-way through data collection (221 Responses: Canada: 170, China: 11, USA: 7, and the following had 1 to 4 participants each: Bolivia, Brazil, Columbia, England, Europe, Iran, Israel, Japan, Malaysia, Mexico, Morocco, New Zealand, Philippines, Russia, Saudi Arabia, Scotland, South Africa, South Korea, Taiwan, and United Arab Emirates). To accurately capture the neural correlates of feedback processing, we used higher standards than we normally would to ensure data quality. Thus, we collected data until we had 500 participants that were not characterized by one of the following a priori criteria: trial counts

after artifact rejection were less than 15 per condition, total artifact rejection exceeded 40% of trials rejected, FCz (electrode of interest) specific artifact rejection exceeded 40% of trials rejected, or independent component analysis based blink correction failed. These criteria were extremely strict to ensure we included clean data in our analyses, and as such we analyzed a total of 637 participants before reaching our goal of 500 clean participants. All participants had normal or corrected-to-normal vision and volunteered to take part in the experiment for extra course credit in a psychology course. All participants provided informed consent approved by the University of Victoria's Human Research Ethics Board.

2.3 | Data and script availability

In line with open science policies, all data and scripts (processing, analysis, plotting, and statistics) used for this manuscript can be found at www.osf.io/65x4v/.

3 | EVENT-RELATED POTENTIALS

3.1 | Data acquisition, processing, and analyses

EEG data were recorded from either a 64 or 32 electrode (Ag/AgCl) EEG system (ActiCAP, Brain Products, GmbH, Munich, Germany) using Brain Vision Recorder (Version 1.21.0004, Brain Products GmbH, Munich, Germany). During recording, electrodes were referenced to a common ground, impedances were kept below 20 k Ω on average, data were sampled at 500 Hz, and an antialiasing low-pass filter of 245 Hz was applied via an ActiCHamp amplifier (Revision 2, Brain Products GmbH, Munich, Germany). Stimuli and EEG markers were temporally synced using a DataPixx stimulus synchronization unit (VPixx, Vision Science Solutions, Quebec, Canada).

All EEG data were processed in MATLAB using custom code (specifically, the following packages: MATLAB-EEG-fileIO, MATLAB-EEG-preProcessing, MATLAB-EEG-icaTools, MATLAB-EEG-time-Frequency Analysis; reference to all code can be found at www.github.com/NeuroTools). Many functions in these packages have dependencies on EEGLAB (Delorme & Makeig, 2004). First, any participant data that were collected with a 64 electrode EEG system was reduced to a 32 electrode system according to Figure 1. Data were processed twice wherein the first pass was used to identify noisy or damaged electrodes, and the second pass was to process data for analysis. Data were re-referenced to an average mastoid reference and filtered using a 0.1 to 30 Hz passband (Butterworth, order 4) and a

⁴Correct and incorrect correspond to whether participants selected the higher valued stimulus while gain and loss correspond to the presented feedback stimulus.

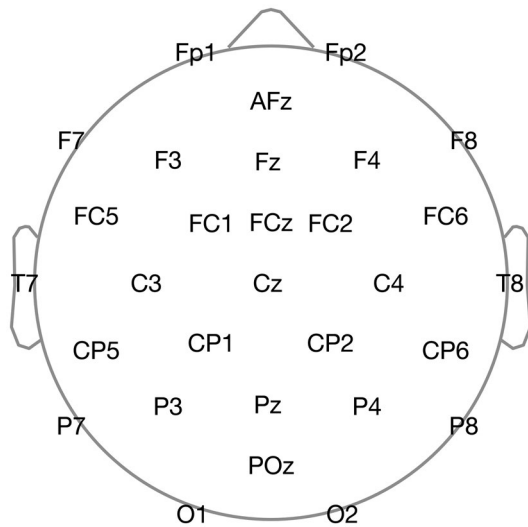


FIGURE 1 Topographic map of electrode placements and labels

60 Hz notch filter. To correct eye blinks, data were run through EEGLAB's independent component analysis (ICA), components reflective of blinks were manually identified via topographic maps and component loadings, and data were reconstructed without components reflective of eye blinks. Data were then segmented from -500 to $1,500$ ms relative to feedback stimulus onset, baseline corrected using a -200 to 0 ms window, and run through artifact rejection with $10 \mu\text{V}/\text{ms}$ gradient and $100 \mu\text{V}$ maximum–minimum criteria. Artifact rejection trial removal rates for each electrode were used to determine noisy or damaged electrodes in that any electrode with a trial rejection rate greater than 40% were tagged for removal. The second pass consisted of the same procedures with the addition of initially removing electrodes identified as noisy or damaged and interpolating them using the spherical spline method after the ICA back transformation was implemented. Additionally, we here used a segment window of -500 to $1,300$ ms for artifact rejection, and then, trimmed it to -200 to $1,000$ ms for the ERP analysis. Final data had a trial rejection rate due to artifact rejection criteria of 19% [$SD: 13\%$]⁵.

Data were processed to result in event-related potentials (ERPs). Specifically, each condition (gain and loss) within each participant were averaged to create conditional waveforms. Difference ERPs were created as the gain condition minus the loss condition. Grand average ERPs were created for each condition (gain and loss) and the difference (gain–loss) by averaging across participants.

All analyses were conducted at electrode site FCz; however, parallel analyses at electrode site Cz can be found

in Appendix A of Supporting Information. First, the peak time of the reward positivity was determined by finding the peak amplitude of the entire difference waveform for each participant. We used three methods to then extract reward positivity amplitudes: mean peak, maximum peak, and base-to-peak. For the mean peak, we averaged the voltages surrounding the peak location for each participant. As there is no standardized window size, we examined a series of widths, from 1 to 500 ms, and found the inflection point of the resulting curve to be at a width of ± 46 ms surrounding the peak (see 3.2. ERP Results). Thus, we averaged data within this window for each participant. To extract the maximum peak, we identified the largest amplitude within the 200 to 400 ms time window, as this window size is most common from past research (Sambrook & Goslin, 2015). To extract the base-to-peak measures, we extracted the minimum voltage of the trough immediately prior to the reward positivity (i.e., the P200; within 150–250 ms from feedback stimulus onset) and subtracted it from the maximum peak measure. For each measure (mean, maximum, and base-to-peak), descriptive statistics (i.e., mean and standard deviation) and estimation statistics (i.e., 95% confidence intervals and Cohen's d) were computed and reported. Although null-hypothesis significance testing statistics are not here provided, all repeated-measures t tests comparing gain from loss conditions across all measures can be computed with the statistics provided (this statement additionally applies to Fast Fourier Transform and time–frequency analyses).

We also examined the number of participants who had a reversed, or flipped, reward positivity. Specifically, although we would expect a positive amplitude when measuring the reward positivity, this was not the case for all participants. The amplitude of the reward positivity was negative for some participants. Correspondingly, we computed the percentage of participants with a negative deflecting reward positivity and additionally analyzed them separately from those with a typical reward positivity. Finally, we computed sample size requirements to detect the reward positivity for desired power of 80%, 85%, 90%, 95%, and 99% using the “pwr” R package (Champely, 2018).

3.2 | ERP results

Feedback stimuli elicited a reward positivity with a large effect size when using a mean peak measure, a maximum peak measure, and a base-to-peak measure, see Table 1 and Figures 2 and 3. Additionally, we provide both mean and maximum amplitude scores with different window sizes when considering both the reward positivity peak and baseline activity (see Figure 4). The reward positivity peaked at a latency of 310 ms [$SD: 26$ ms, 95% CI: 308 ms,

⁵Here, the high rejection rate is because we used a wide time window and strict rejection criteria to ensure clean data in our analyses.

TABLE 1 Descriptive statistics and effect sizes for the reward positivity, delta, and theta across ERP, FFT, and time–frequency analysis methods

	Mean [95% CI]	Standard deviation	Cohen's <i>d</i> [95% CI]
<i>ERP</i>			
Mean	3.70 μV [3.34 μV , 4.07 μV]	4.11 μV	0.90 [0.77, 1.03]
Maximum	7.82 μV [7.42 μV , 8.23 μV]	4.59 μV	1.71 [1.56, 1.85]
Base-to-Peak ^{vii}	10.52 μV [10.12 μV , 10.91 μV]	4.49 μV	2.34 [2.18, 2.50]
<i>FFT</i>			
Delta	1.19 μV^2 [0.86 μV^2 , 1.51 μV^2]	3.70 μV^2	0.45 [0.32, 0.58]
Theta	-0.48 μV^2 [-0.54 μV^2 , -0.42 μV^2]	0.73 μV^2	-0.93 [-1.08, -0.78]
<i>Time–frequency</i>			
Delta	0.82 μV^2 [0.67 μV^2 , 0.97 μV^2]	1.67 μV^2	0.69 [0.56, 0.83]
Theta	-1.70 μV^2 [-1.90 μV^2 , -1.51 μV^2]	2.25 μV^2	-1.07 [-1.23, -0.91]

Note: All values calculated as the difference between gain and loss conditions (gain-loss). 95% CI = 95% confidence intervals.

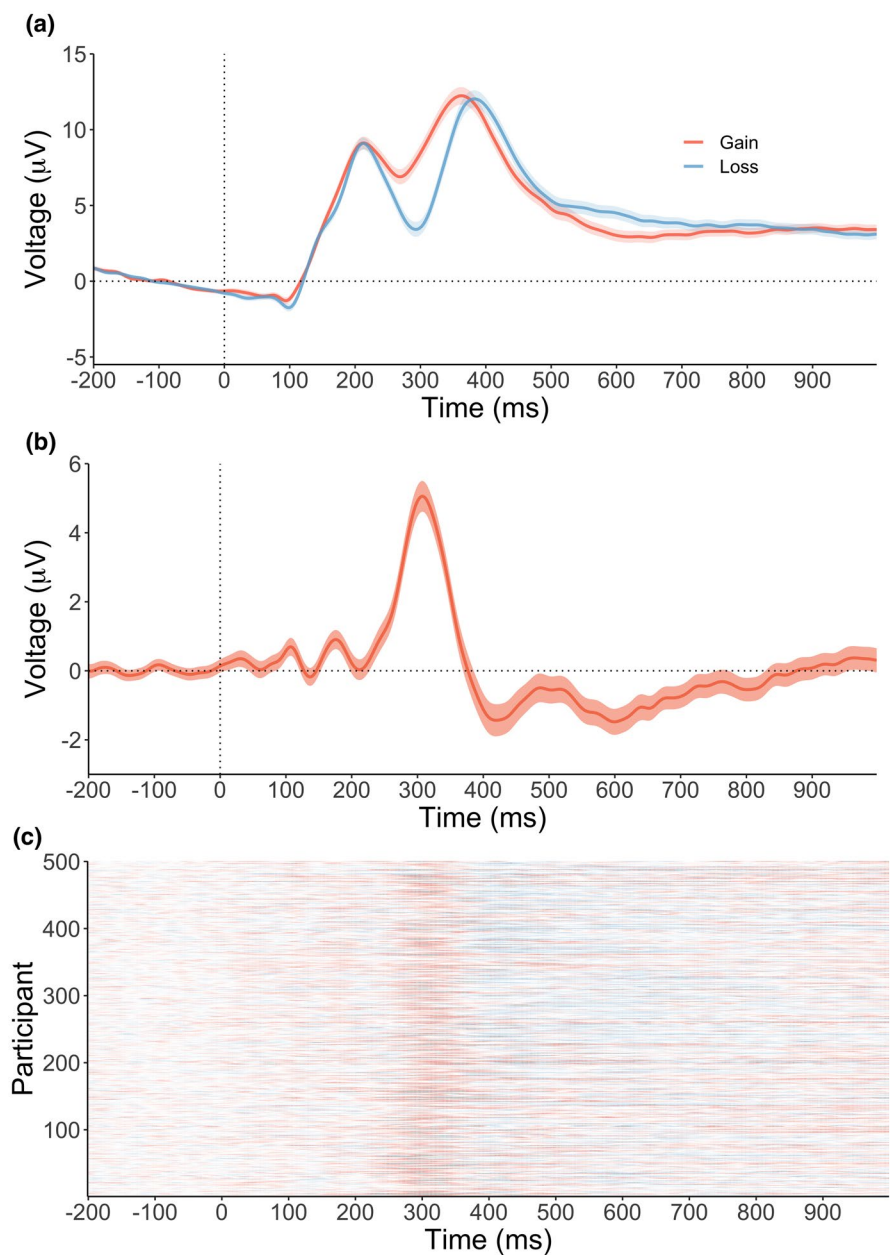


FIGURE 2 ERP waveforms at electrode FCz illustrating the reward positivity. (a) grand averaged conditional waveforms with 95% confidence intervals, (b) grand averaged difference waveform with 95% confidence intervals, and (c) heatmap (two-dimensional representation where color corresponds to magnitude) of amplitude for each participant across time

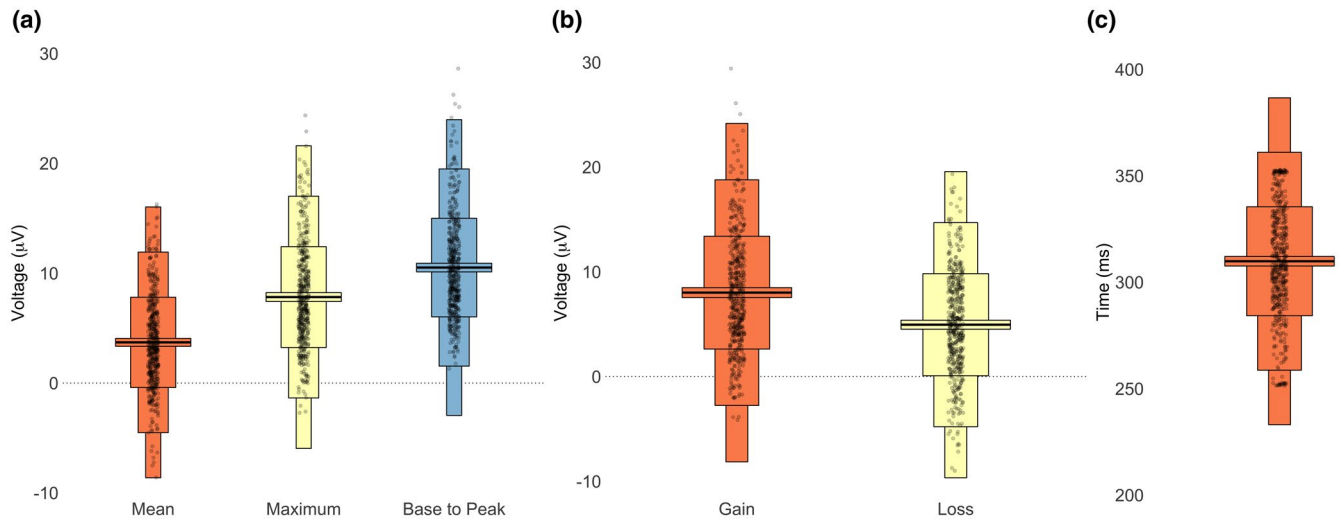


FIGURE 3 Illustration of individual data and error terms for the reward positivity. (a) Difference amplitudes for each amplitude extraction technique, (b) conditional amplitudes for the mean peak measure, and (c) peak timing of the reward positivity. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, and 3 standard deviations

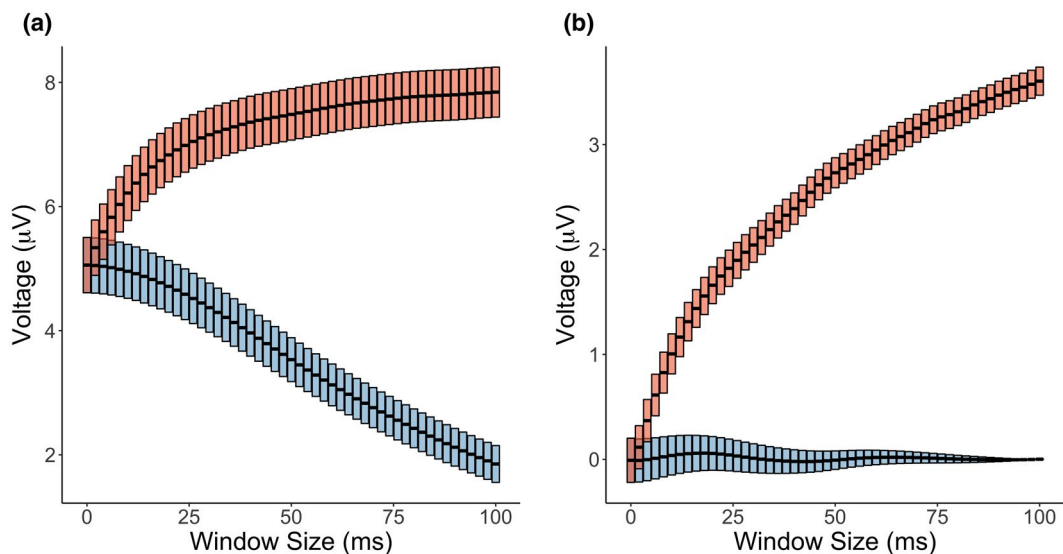


FIGURE 4 EEG amplitudes using mean (blue) and maximal (red) measures with different window sizes around the reward positivity peak (a) and within baseline activity (b). Crossbar represents mean and 95% confidence intervals

312 ms] (see Figure 3) and was centered on FCz and Cz (see Figure 5). Moreover, the gain waveform had a large negative deflection at peak latency 277 ms [*SD*: 37 ms, 95% CI: 274 ms, 280 ms] and the loss waveform had a medium negative deflection at peak latency 292 ms [*SD*: 30 ms, 95% CI: 290, 295 ms], see Table 2 and Figure 2. For mean peak, maximum peak, and base-to-peak measures, 17.2%, 2.0%, and 0.0% of participants, respectively, demonstrated a flipped reward positivity, see Figure 3 and S1. Lastly, power calculations for prospective sample size requirements to detect a reward positivity are depicted in Table 3.

4 | FAST FOURIER TRANSFORMS

4.1 | Data acquisition, processing, and analyses

Data were acquired and processed in the same manner as was described for ERPs (see Section 3.1.), except using a segment of -500 – $1,000$ ms. Fast Fourier transforms (FFT) for each trial were conducted on a segment of -500 – $1,000$ ms from feedback stimulus onset, with 500 ms Hanning taper at the beginning and end of the segment, and with normalization. Note that this method results in a resolution of 0.67 Hz, thus making

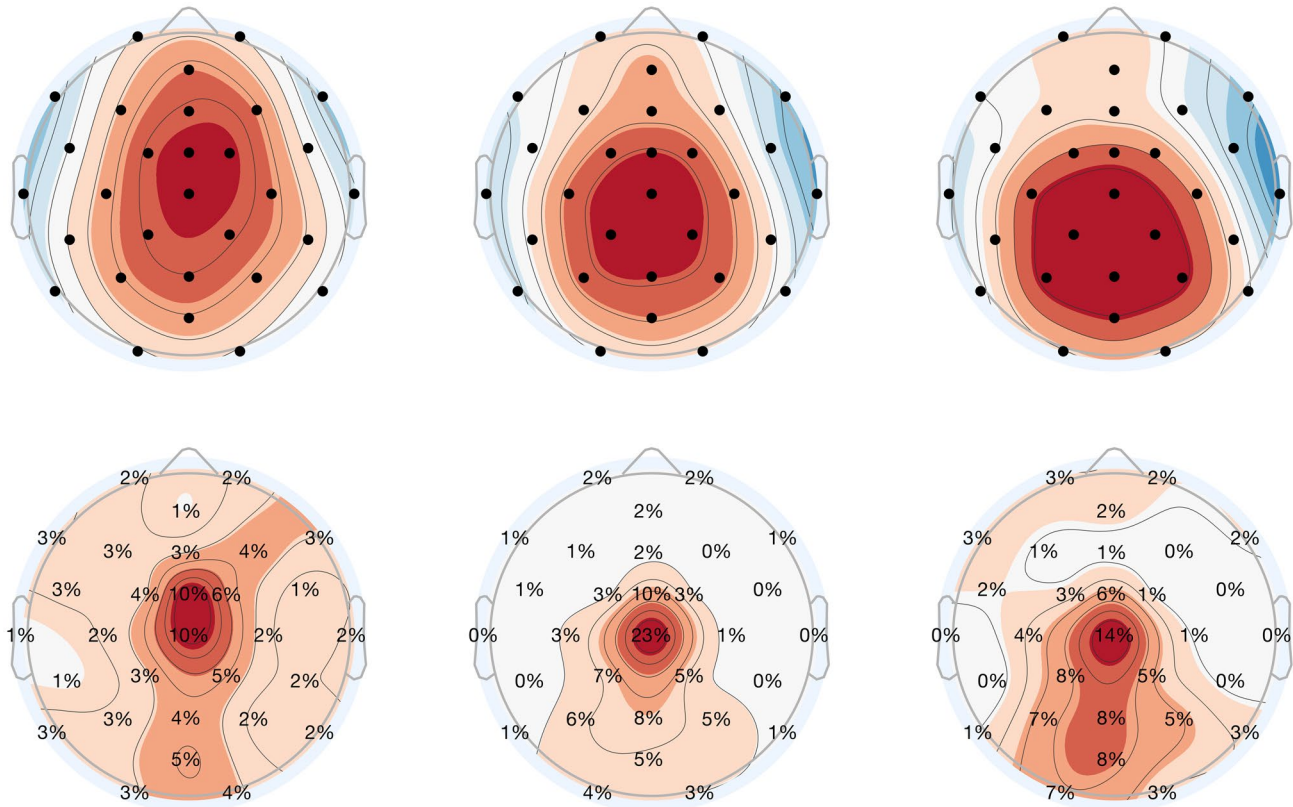


FIGURE 5 Topographic maps of grand averaged data for the reward positivity. Left: amplitudes from the difference waveform, middle: amplitudes from the gain waveform, right: amplitudes from the loss waveform. Top: traditional topographic maps, Bottom: percentage (%) of participants where the reward positivity peaked at the respective electrode

TABLE 2 Descriptive statistics and effect sizes for conditional (gain and loss) amplitudes of the reward positivity, delta, and theta across ERP, FFT, and time–frequency analysis methods

	Mean [95% CI]	Standard deviation	Cohen's <i>d</i> [95% CI]
<i>ERP</i>			
Gain	8.02 μ V [7.54 μ V, 8.49 μ V]	5.38 μ V	1.49 [1.35, 1.63]
Loss	4.96 μ V [4.53 μ V, 5.38 μ V]	4.87 μ V	1.02 [0.89, 1.15]
<i>FFT</i>			
Gain delta	14.33 μ V ² [13.82 μ V ² , 14.85 μ V ²]	5.86 μ V ²	3.46 [3.13, 3.79]
Loss theta	3.58 μ V ² [3.45 μ V ² , 3.71 μ V ²]	1.46 μ V ²	3.46 [3.13, 3.79]
<i>Time–frequency</i>			
Gain delta	0.10 μ V ² [0.10 μ V ² , 0.11 μ V ²]	0.04 μ V ²	3.31 [3.00, 3.63]
Loss theta	0.46 μ V ² [0.44 μ V ² , 0.48 μ V ²]	0.22 μ V ²	2.92 [2.64, 3.21]

Note: 95% CI = 95% confidence intervals. All measures are a comparison of the conditional activity versus zero.

the lower frequencies less represented within analyses that the higher frequencies. The output was then averaged for each condition and participant. Difference FFTs were computed as the gain condition minus the loss condition, and grand averaged FFTs for both conditions (gain and loss) and the difference were constructed by averaging data across participants.

All analyses were conducted at electrode site FCz and parallel analyses at electrode site Cz can be found in Appendix A of Supporting Information. Descriptive (i.e., mean and standard deviation) and estimation statistics (i.e., 95% confidence

intervals and Cohen's *d*) were computed for both delta (<3 Hz) and theta (3–7 Hz) power (Bernat et al., 2008)⁶. Additional descriptives and estimation statistics were computed for each frequency. Finally, we computed and analyzed the percentage of participants with reversed effects and conducted power analyses for both averaged and individual frequencies.

⁶Please note that although we are here using a cut-off between delta and theta at 3 Hz, we rather suggest having your own data determine the frequency ranges of interest. To facilitate this, we also include individual frequency analyses.

TABLE 3 Required within-subject sample size calculations for different levels of desired power for reward positivity, delta, and theta measures

	Desired power				
	80%	85%	90%	95%	99%
	Difference				
<i>ERP</i>					
Mean	12	13	15	18	25
Maximum	5	5	6	7	9
Base-to-Peak	4	4	4	5	6
<i>FFT</i>					
Delta	40	46	53	65	91
Theta	11	12	14	17	23
<i>Time-frequency</i>					
Delta	18	21	24	29	40
Theta	9	10	11	13	18

Note: Values calculated from Cohen's *d* effect sizes as determined using difference data (gain-loss).

4.2 | FFT results

FFT analyses revealed that gain feedback elicited more delta power than loss feedback with a small effect, see Table 1 and Figures 6 and 7. In contrast, loss feedback elicited a medium effect of theta power related to gain feedback, see Table 1 and Figures 6 and 7. We further investigated gain-specific delta power and loss-specific theta power (i.e., within each respective condition) in Table 2. Further descriptive and estimation statistics for each frequency can be found in Figure 8. Whereas delta power was centered on electrodes FCz and Cz, theta activity peaked on electrode FCz (see Figure 9). For delta and theta activity, 39.4% and 25.6% of participants, respectively, demonstrated reversed neural power, see Figure 7 and S2. Moreover, 7.6% of participants had both reversed delta and theta power (i.e., 13.2% of those with reversed power had both delta and theta reversed). Prospective required sample sizes to detect delta and theta power effects are presented in Table 3 and prospective required sample sizes for each frequency individually are presented in Table 4.

5 | TIME-FREQUENCY WAVELET ANALYSES

5.1 | Data acquisition, processing, and analyses

Data were acquired and processed in the same manner as was described for ERPs (see Section 3.1.). The following time-frequency technique was applied to a segment size

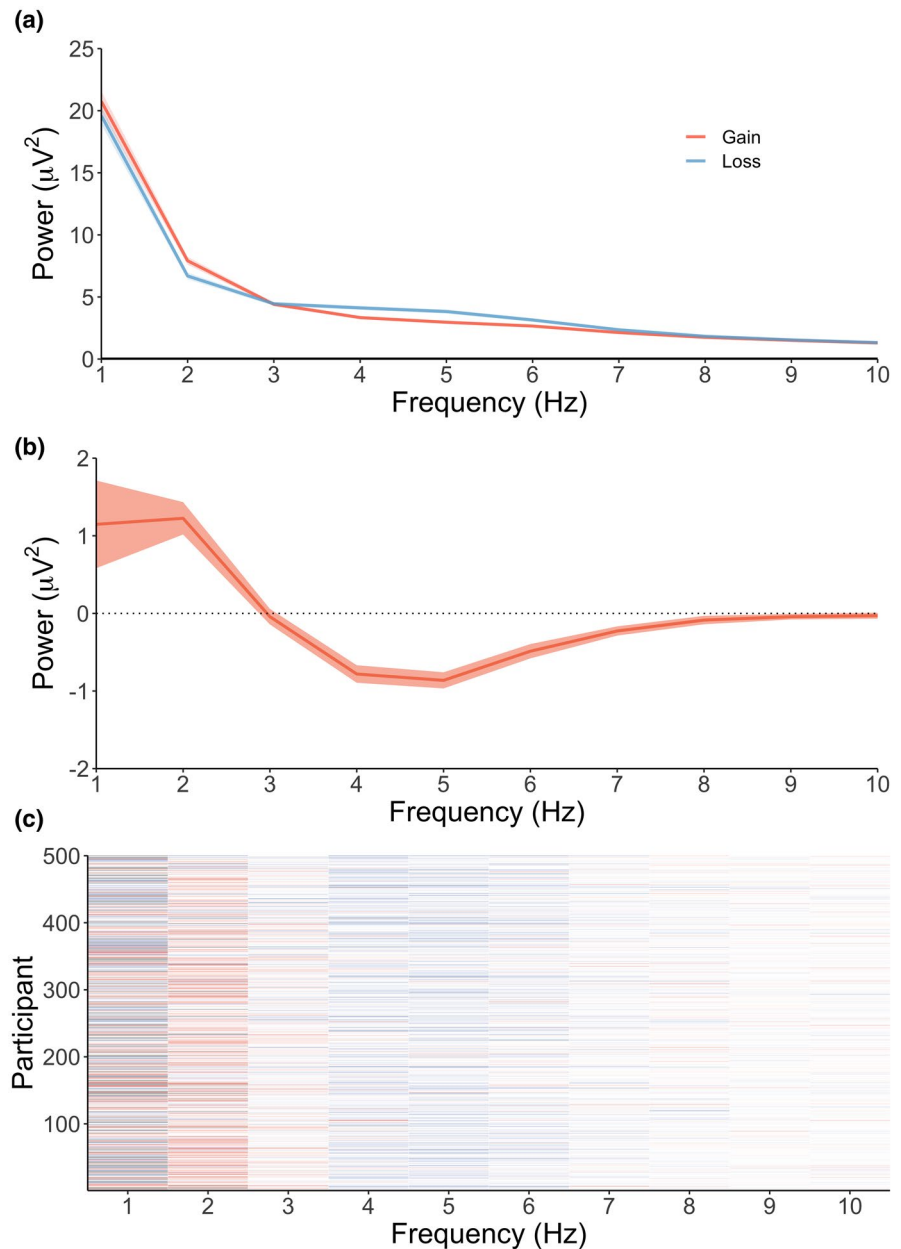
of -500 to $1,300$ ms; however, analyses were conducted on, and plots were restricted to, a window size of -200 to $1,000$ ms. The additional 300 ms prior to and following the segment size was to ensure that time-frequency edge artifacts (artifacts that emerge at the beginning and end of the time window) would not be included within the time window of interest.

In this article we chose to use the complex Morlet wavelet convolutions method of time-frequency decomposition because it is the most popular method in the neuroscience literature (Cohen, 2014). It is important to note, however, that there are many other methods of time-frequency decomposition that we could have used—for example, the Bernat group (2008, 2011, 2015) use the reduced inference distribution method (Bernat et al., 2005). Indeed, there also exists the filter-Hilbert method, the short-time Fourier transforms method, the multi-taper method, and many more (Cohen, 2014). Each method can have unique advantages in contrast to the others—for example, the reduced inference distribution method has an enhanced temporal resolution, circumventing a limitation of some other methods that contain a “smearing” of activity across time.

Time-frequency analyses were conducted as the product of FFT-transformed EEG data and a normalized complex Morlet wavelet (Gaussian-windowed complex sine wave, Morlet parameter of 6, frequency range of $1-30$ Hz in 30 linear steps; in line with Cohen, 2014). Output was averaged for each condition and participant. Difference time-frequency wavelets were created by subtracting the loss condition from the gain condition, and grand average time-frequency wavelets were created for both conditions (gain and loss) and the difference by averaging data across participants.

All analyses were conducted at electrode site FCz; however, Appendix a of Supporting Information contain parallel analyses for Cz. First, we used the collapsed localizer method for extracting clusters of activity within time-frequency analyses (Luck & Gaspelin, 2017). The collapsed localizer method involves first averaging all time-frequency data across conditions and participants and then, using a power threshold (here $5.5 \mu\text{V}^2$) to isolate cluster activity of interest. This process resulted in two clusters of interest: one within the delta band and one within the theta band. We then used the clusters to extract delta and theta power from the time-frequency conditional and difference (gain-loss) wavelets for each participant. Specifically, for the delta band we extracted any cell (time and frequency) within the clusters below 3 Hz and averaged the output for each participant (Bernat et al., 2008); as no activity within the 1 Hz band reached our threshold, delta power simply corresponded to 2 Hz activity. We conducted the same process for the theta band but restricted to the 3–7 Hz band (Bernat et al., 2008). Descriptive (i.e., mean and standard deviation) and estimation statistics (i.e., 95%

FIGURE 6 FFT line plots at electrode FCz illustrating delta and theta effects. (a) grand averaged conditional line plots with 95% confidence intervals, (b) grand averaged difference line plot with 95% confidence intervals, and (c) heatmap (two-dimensional representation where color corresponds to magnitude) of power for each participant across frequencies



confidence intervals and Cohen's d) were then conducted on these data at electrode FCz. Also, descriptive and estimation statistics were computed for each frequency, the percentage of individuals with reversed effects were computed and separately analyzed, and power calculations for both averaged and individual frequencies were conducted.

5.2 | Time–frequency wavelet results

The collapsed localizer method for cluster extraction on time–frequency wavelet data determined that delta power was larger in the gain condition and that theta power was larger in the loss condition, see Table 1 and Figures 10 and 11. Whereas feedback valence (gain and loss) elicited a

medium effect of delta power, they elicited a large effect of theta power. The mean peak latencies (delta power: 358 ms [SD : 279 ms, 95% CI: 334, 383 ms]; theta power: 420 ms [SD : 161 ms, 95% CI: 405, 433 ms]) with descriptive and estimation statistics for averaged delta and theta bands are also plotted in Figure 11. The mean power and latency with descriptive and estimation statistics for each frequency are plotted in Figures 12 and S3, respectively. Furthermore, analyses specifically tying delta power to the gain condition and theta power to the loss condition can be found in Table 2. Delta power peaked at electrode Cz and theta power peaked at electrode FCz, see Figure 13. For delta and theta power, 30.8% and 19.6% of participants, respectively, demonstrated reversed neural effects, see Figures 11 and S4. Moreover, 5.4% of participants had both reversed delta and theta activity

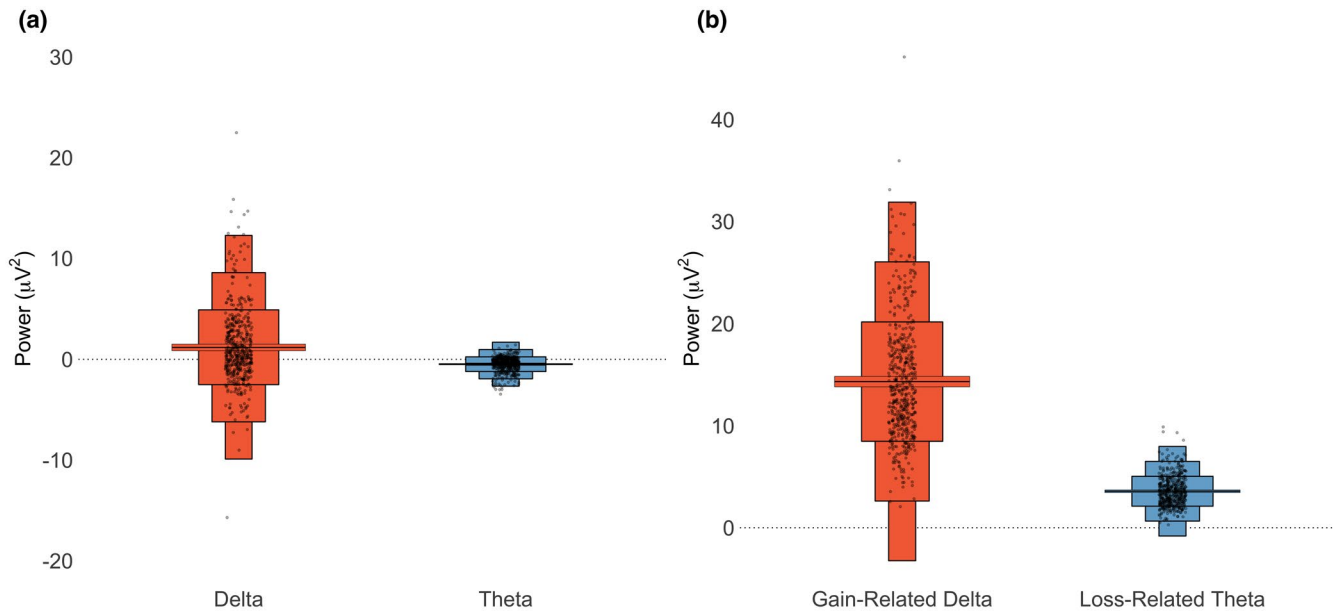


FIGURE 7 Illustration of individual data and error terms for delta and theta EEG power. (a) Difference amplitudes for delta and theta and (b) conditional amplitudes for delta and theta. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, and 3 standard deviations

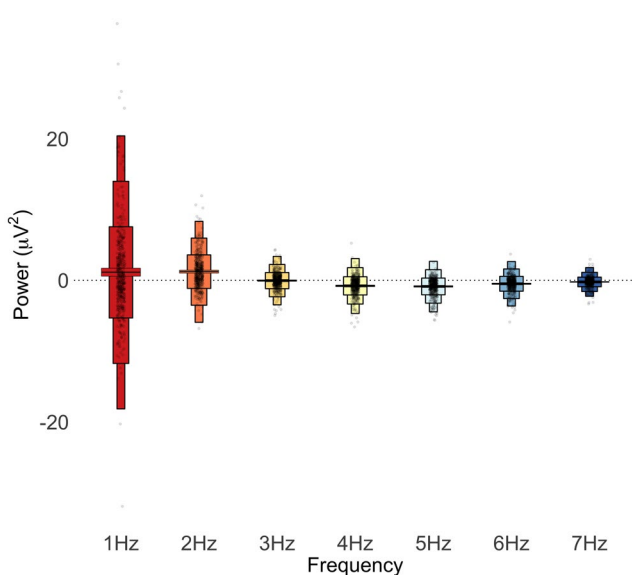


FIGURE 8 Illustration of individual power data and error terms for FFT derived analyses. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, and 3 standard deviations

(or in other words, 12.0% of participants with reversed activity had both reversed delta and theta power). Prospective sample sizes for future research are described in Table 3 for both delta and theta power and prospective required sample sizes for each frequency are presented in Table 4.

6 | COMPARISONS BETWEEN MEASURES

6.1 | Data analyses

Next, we explored associations between different neural measures with a series of correlational analyses. All of these analyses only considered neural data as determined by the difference measure approach.

6.2 | Comparison results

For FFT measures, there was a medium positive relationship between the reward positivity and delta power, no association between the reward positivity and theta power, and a small positive association between delta and theta power, see Table 5 and Figure 14a-c. For time–frequency analyses, there was a medium-to-large positive association between the reward positivity and delta power, a small negative association between the reward positivity and theta power, and no association between delta and theta power, see Table 5 and Figure 14d-f.

7 | INDIVIDUAL DIFFERENCES

7.1 | Data analyses

Finally, we explored the associations of behavioral and demographic data with neural data. For gender effects

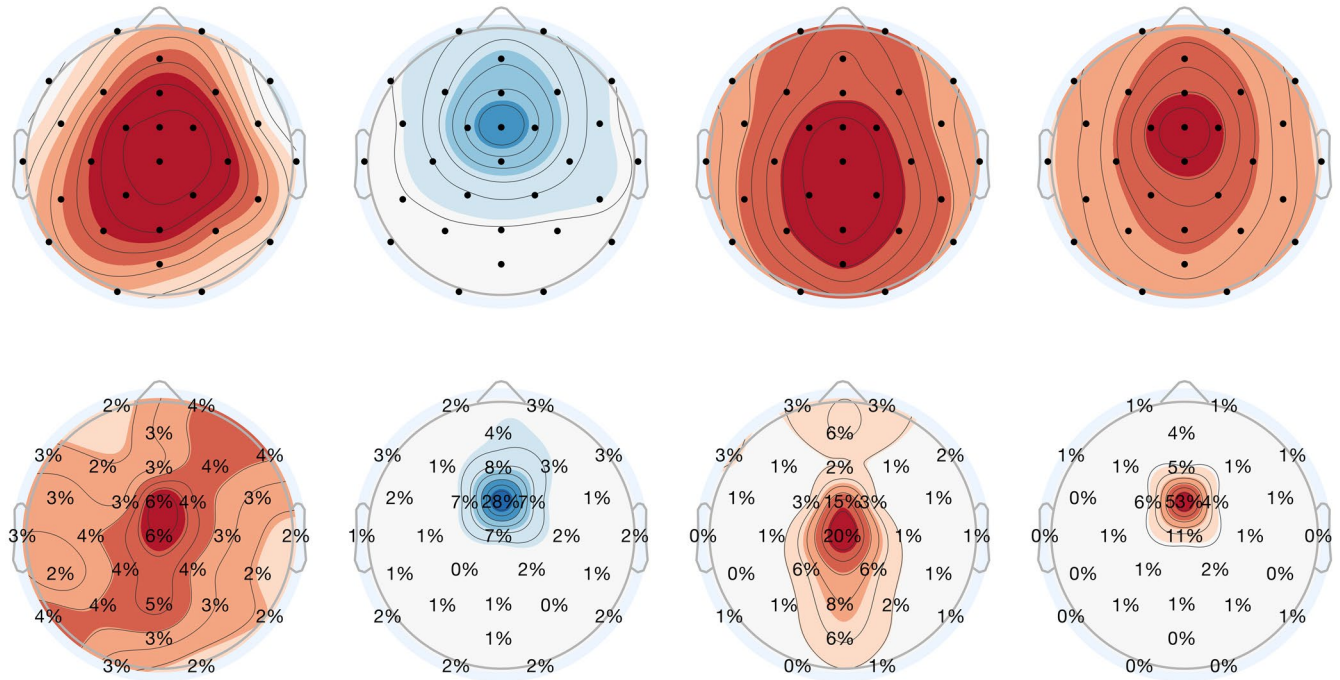


FIGURE 9 Topographic maps of grand averaged data for delta and theta EEG power from FFT analyses. From left to right: delta power of the difference between gain and loss, theta power of the difference between gain and loss, delta power for the gain condition, and theta power for the loss condition. Top: traditional topographic maps, Bottom: percentage (%) of participants where delta or theta power peaked at the respective electrode

TABLE 4 Required within-subject sample size calculations for different levels of desired power for each frequency

	Cohen's <i>d</i>	Desired power				
		80%	85%	90%	95%	99%
Difference						
<i>FFT</i>						
1 Hz	0.25	125	143	167	206	290
2 Hz	0.73	17	19	22	26	36
3 Hz	-0.05	2,886	3,301	3,863	4,777	6,753
4 Hz	-0.86	13	14	16	20	27
5 Hz	-1.03	9	11	12	14	19
6 Hz	-0.66	20	23	26	32	44
7 Hz	-0.47	37	42	49	60	84
<i>Time-frequency</i>						
2 Hz	0.69	18	21	24	29	40
3 Hz	-0.17	269	308	360	444	628
4 Hz	-0.96	11	12	13	16	22
5 Hz	-1.10	9	10	11	13	17
6 Hz	-0.92	11	13	14	17	24
7 Hz	-0.75	16	18	21	25	35

Note: Values calculated from Cohen's *d* effect sizes as determined using difference data (gain-loss).

on neural activity, we compared females versus males using independent-samples *t* tests. For age and accuracy effects on neural activity, we conducted correlational analyses.

7.2 | Associations results

Analyses determined small effects of gender in the reward positivity and time-frequency delta activity but not in any

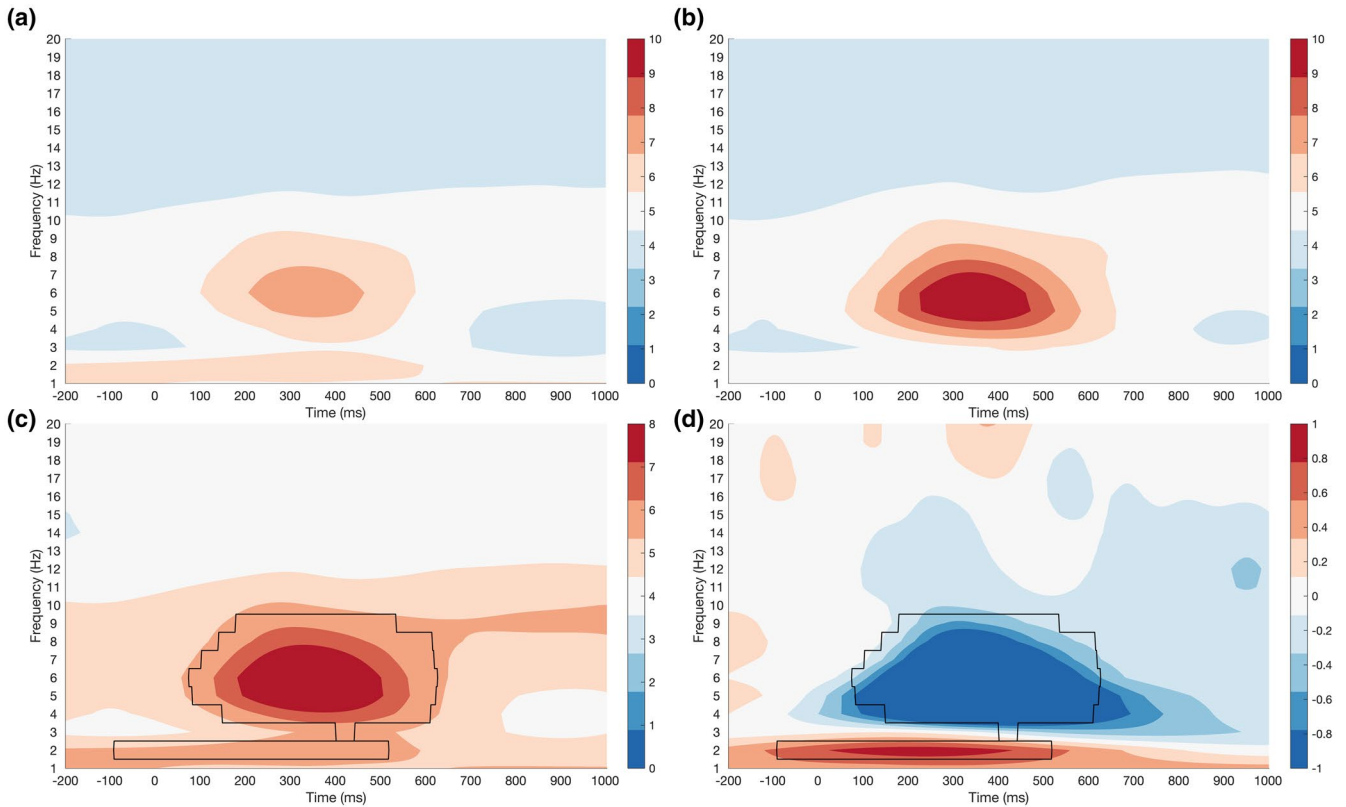


FIGURE 10 Time–frequency wavelets at electrode FCz illustrating delta and theta power. (a) grand averaged conditional wavelet for the gain condition, (b) grand averaged conditional wavelet for the loss condition, (c) wavelets grand averaged across both the gain and loss conditions, and (d) grand averaged difference wavelet. Contour lines were determined using the collapsed isolator method

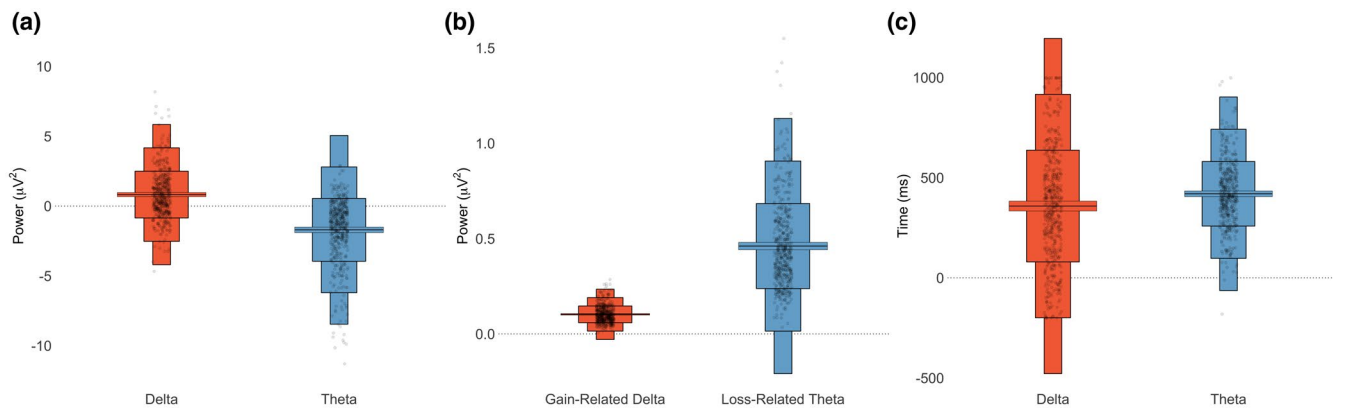


FIGURE 11 Illustration of individual data and error terms for delta and theta power. (a) Difference amplitudes for delta and theta power, (b) conditional amplitudes for delta and theta power, and (c) peak amplitude times for delta and theta power. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, and 3 standard deviations

other neural measures, see Table 6. Specifically, females had higher reward positivity amplitudes and delta power in comparison to males. In addition, we report the same analyses but when focusing on conditional data in Table 7. Correlational analyses determined small negative associations of age for the reward positivity and small positive associations for theta activity (for both FFT and time–frequency measures), see

Table 8 and Figure 15a-e. Increases in age corresponded to smaller reward positivity amplitudes and less pronounced theta activity (as theta power effects are negatively peaked). We also provide the same analyses but for conditional data in Table 9 and corresponding plots in Figure 16a-f.

Correlational analyses also determined small to moderate negative associations between accuracy rates and all neural

measures, see Table 8 and Figure 15f-J. Increases in accuracy corresponded to smaller reward positivity amplitudes and delta activity. Additionally, increases in accuracy corresponded to more pronounced theta activity. These associations

were seen in both FFT and time–frequency measures. Parallel analyses with conditional data are also reported in Table 9 and Figure 16G-L.

8 | DISCUSSION

Here, we sought to provide normative descriptives of the EEG correlates of feedback processing (i.e., ERP vs. frequency vs. time–frequency), examinations of how EEG processing methods are related, and selected recommendations for EEG methodology.

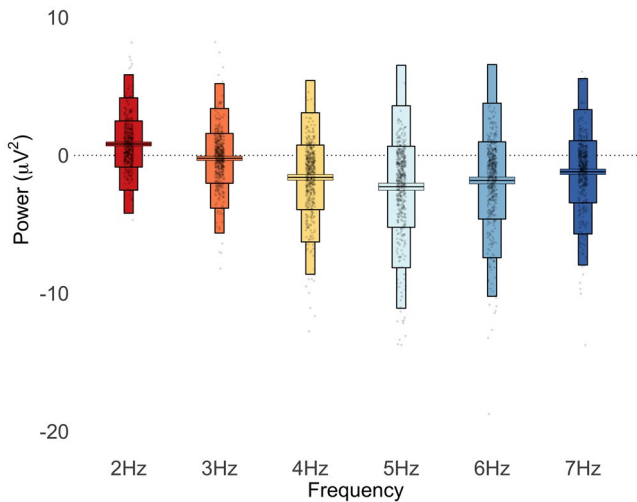


FIGURE 12 Illustration of individual power data and error terms for time–frequency derived analyses for frequencies with effects. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, and 3 standard deviations

TABLE 5 Correlational results between the reward positivity, delta activity, and theta activity

	Pearson <i>r</i> [95% CI]	<i>p</i> value
<i>FFT</i>		
Reward positivity and delta	.54 [.48, .60]	<.0001
Reward positivity and theta	-.09 [-.17, .00]	.0509
Delta and theta	.19 [.11, .28]	<.0001
<i>Time–frequency</i>		
Reward positivity and delta	.40 [.32, .47]	<.0001
Reward positivity and theta	-.16 [-.24, -.07]	.0004
Delta and theta	-.03 [-.11, .06]	.5411

Note: 95% CI = 95% confidence intervals.

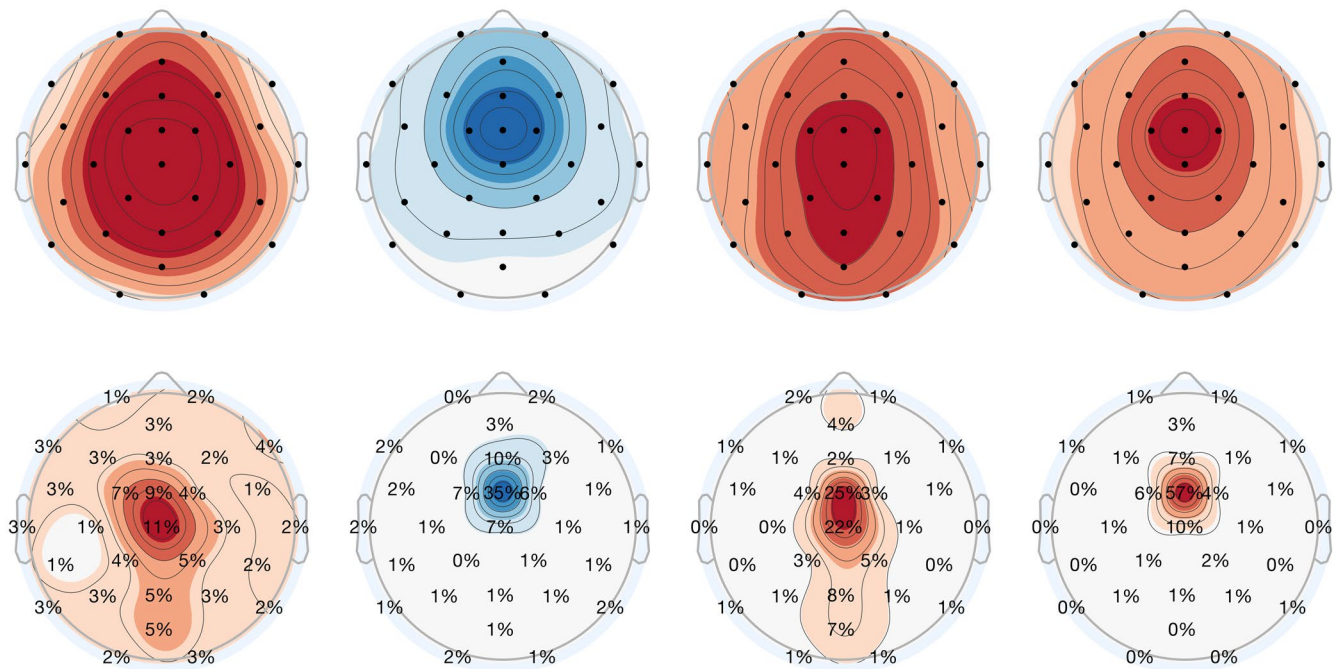


FIGURE 13 Topographic maps of grand averaged data for delta and theta power from time–frequency analyses. From left to right: delta power of the difference between gain and loss, theta power of the difference between gain and loss, delta power for the gain condition, and theta power for the loss condition. Top: traditional topographic maps, Bottom: percentage (%) of participants where delta or theta power peaked at the respective electrode

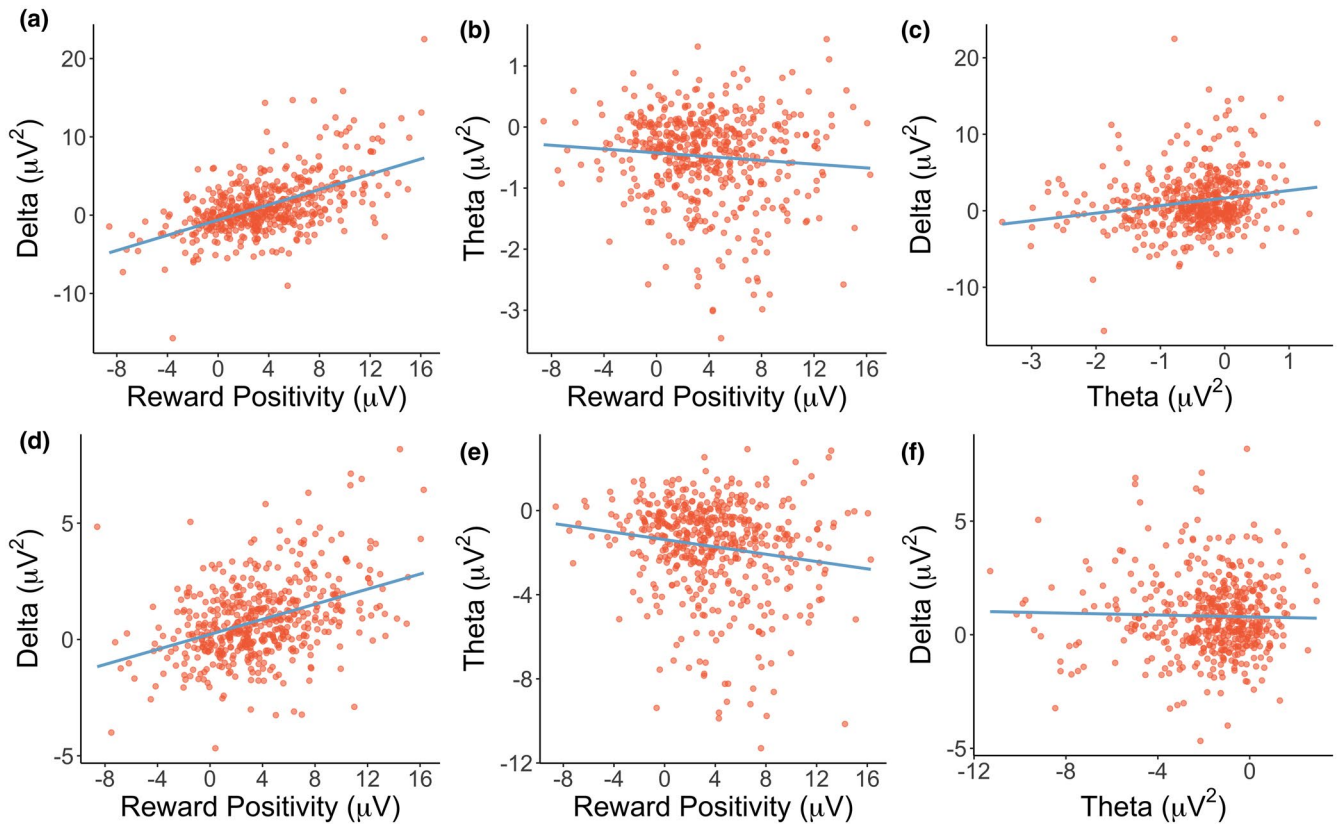


FIGURE 14 Scatterplots between different measures of neural activity. (a–c) frequency data correspond to FFT measures; (d–f) frequency data correspond to time–frequency measures. Blue lines correspond to linear regressions

TABLE 6 Statistical results of gender (female and male) effects on difference neural activity. Difference = female–male

	Difference [95% CI]	<i>t</i> value	<i>p</i> value	Cohen's <i>d</i> [95% CI]
<i>ERP</i>				
Reward positivity	1.29 μ V [0.56 μ V, 2.01 μ V]	3.51	.0005	0.32 [0.12, 0.51]
<i>FFT</i>				
Delta	0.71 μ V ² [0.05 μ V ² , 1.36 μ V ²]	2.12	.0349	0.19 [0.00, 0.38]
Theta	–0.09 μ V ² [–0.22 μ V ² , 0.04 μ V ²]	–1.30	.1945	–0.12 [–0.31, 0.07]
<i>Time–frequency</i>				
Delta	0.43 μ V ² [0.15 μ V ² , 0.72 μ V ²]	2.95	.0033	0.26 [0.07, 0.45]
Theta	–0.19 μ V ² [–0.60 μ V ² , 0.22 μ V ²]	–0.93	.3514	–0.09 [–0.28, 0.10]

Note: 95% CI = 95% confidence intervals.

8.1 | Amplitude and power

For the reward positivity, we provided three different measures of amplitude: the base-to-peak, the maximum peak, and the mean peak. First, we must caution against base-to-peak measures. The base-to-peak method involves contrasting the reward positivity with the amplitude of the P200 ERP component. This method does not isolate the reward positivity but rather contrasts effects between the reward positivity and the P200 (Picton et al., 2000). Indeed, in addition to valence effects on the reward positivity (gains were larger than losses), we found valence effects on the P200 (losses were

larger than gains)⁷. The confounding P200 effect then inflated the measured reward positivity peak in our study and as such we recommend against using this method of peak detection.

We would also like to caution against maximum peak measures as they are biased by noise (e.g., unrelated biological and environmental noise; Luck, 2014; Luck & Gaspelin, 2017; Picton et al., 2000). Specifically, our

⁷We also found an effect of valence for the base (P200) of this measure, $M_d = -2.69 \mu$ V [–2.95 μ V, –2.43 μ V], $SD = 2.95$, Cohen's $d = -0.91$ [–1.04, –0.78].

TABLE 7 Statistical results of gender (female and male) effects on conditional neural activity. Difference = female–male

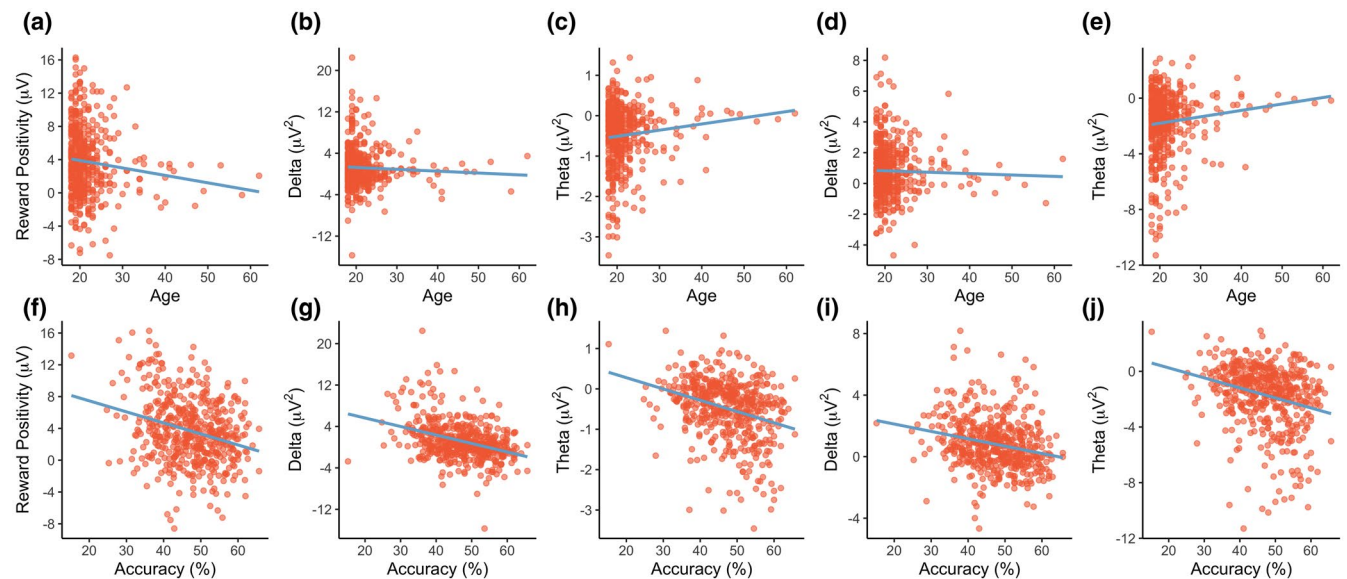
	Difference [95% CI]	<i>t</i> value	<i>p</i> value	Cohen's <i>d</i> [95% CI]
<i>ERP</i>				
Gain	0.77 μ V [−0.22 μ V, 1.76 μ V]	1.53	.1269	0.14 [−0.05, 0.33]
Loss	−0.42 μ V [−1.33 μ V, 0.50 μ V]	−0.90	.3690	−0.09 [−0.28, 0.11]
<i>FFT</i>				
Gain delta	2.68 μ V ² [1.61 μ V ² , 3.75 μ V ²]	4.91	<.0001	0.47 [0.27, 0.66]
Loss theta	0.72 μ V ² [0.45 μ V ² , 0.99 μ V ²]	5.30	<.0001	0.50 [0.31, 0.70]
<i>Time–frequency</i>				
Gain delta	0.02 μ V ² [0.01 μ V ² , 0.03 μ V ²]	4.09	<.0001	0.39 [0.20, 0.59]
Loss theta	0.09 μ V ² [0.05 μ V ² , 0.13 μ V ²]	4.41	<.0001	0.41 [0.22, 0.61]

Note: 95% CI = 95% confidence intervals.

TABLE 8 Correlational results of age and accuracy effects on difference neural activity

	Age		Accuracy	
	Pearson <i>r</i> [95% CI]	<i>p</i> value	Pearson <i>r</i> [95% CI]	<i>p</i> value
<i>ERP</i>				
Reward positivity	−.12 [−.20, −.03]	.0104	−.26 [−.35, −.19]	<.0001
<i>FFT</i>				
Delta	−.05 [−.14, .04]	.2588	−.36 [−.44, −.28]	<.0001
Theta	.11 [.02, .20]	.0140	−.31 [−.39, −.23]	<.0001
<i>Time–frequency</i>				
Delta	−.03 [−.12, .06]	.5249	−.23 [−.31, −.15]	<.0001
Theta	.11 [.02, .19]	.0168	−.26 [−.34, −.18]	<.0001

Note: 95% CI = 95% confidence intervals.

**FIGURE 15** Scatterplots of age (top) and accuracy (bottom) with difference neural activity. (b, c, g, h) frequency data correspond to FFT measures; (d, e, i, j) frequency data correspond to time–frequency measures. Blue lines correspond to linear regressions

analyses demonstrated the confound of noise as the maximum peak amplitudes were positively associated with the window size used for investigation (Figure 4). Indeed, we

also demonstrated that maximum peak analyses produced artificial effects within baseline activity (Figure 4). Our findings indicate then that maximum peaks are biased

TABLE 9 Correlational results of age and accuracy effects on conditional neural activity

	Age		Accuracy	
	Pearson r [95% CI]	p value	Pearson r [95% CI]	p value
<i>ERP</i>				
Gain	-.04 [-.13, .05]	.3637	-.20 [-.28, -.11]	<.0001
Loss	.05 [-.04, .14]	.2595	-.03 [-.12, .06]	.5333
<i>FFT</i>				
Gain delta	-.14 [-.22, -.05]	.0021	-.20 [-.28, -.11]	<.0001
Loss theta	-.20 [-.28, -.11]	<.0001	.14 [.05, .22]	.0022
<i>Time–frequency</i>				
Gain delta	-.15 [-.23, -.06]	.0011	-.07 [-.16, .02]	.1068
Loss theta	-.19 [-.27, -.10]	<.0001	.13 [.04, .21]	.0038

Note: 95% CI = 95% confidence intervals.

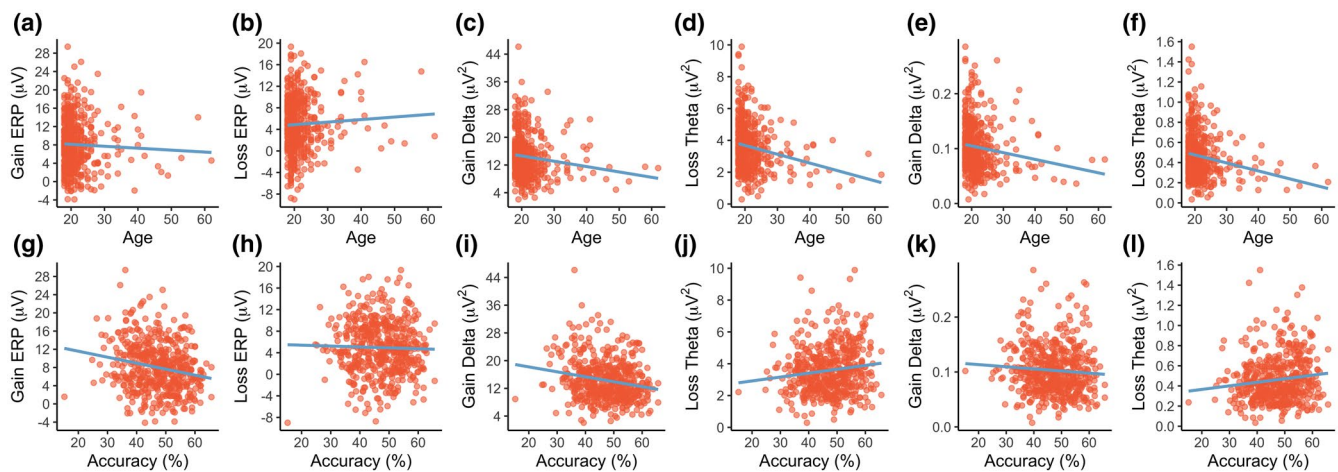


FIGURE 16 Scatterplots of age (top) and accuracy (bottom) with conditional neural activity. (a, b, g, h) ERP data correspond to gain and loss waveforms of the reward positivity. (c, d, i, j) frequency data correspond to FFT measures; (e, f, k, l) frequency data correspond to time–frequency measures. Blue lines correspond to linear regressions

by noise, thus making them unreliable. In line with Luck (2014), we advocate against the use of maximum peak measures in ERP analyses.

As base-to-peak and maximum peak measures are biased methods of extracting ERP component amplitudes, we suggest ERP research use mean peak measures (Luck, 2014; Luck & Gaspelin, 2017; Picton et al., 2000). Our findings demonstrated a large reward positivity effect (Cohen's d of 0.90) with a mean peak of 3.70 μV when contrasting gain and loss feedback processing (Figures 2 and 3)—a value that is in line with the range of amplitudes described in Sambrook and Goslin (2015). With that said, we observed considerable amplitude variability across individuals. Specifically, we found a range of $-8.60 \mu\text{V}$ to $16.28 \mu\text{V}$ for mean peak reward positivity amplitudes across participants (Figure 3). We propose that the descriptives we provide here can be used to understand what characterizes a “typical” reward positivity amplitude and classify atypical populations or potential

outliers when considering neural activity elicited by a two-armed bandit task.

Next, we investigated the EEG response to feedback processing using both FFTs and time–frequency wavelet analyses. We would like to first address that FFT and time–frequency analyses captured delta and theta power effects similarly—delta power was larger for gains and theta power was larger for losses (Figures 6–8 and 10–12). We are unaware of any research using FFT techniques when assessing neural feedback processing and so we here demonstrate FFT results to parallel time–frequency findings and propose this method to be effective.

Delta power was 1.19 and 0.82 μV^2 for FFT and time–frequency analyses, respectively, and theta power was -0.48 and $-1.70 \mu\text{V}^2$ for FFT and time–frequency, respectively. Although the raw values of delta and theta power varied considerably between FFT and time–frequency, the effect sizes were similar. For FFT and time–frequency, we found delta power Cohen's d

effect sizes to be 0.45 and 0.69, respectively, and theta power effect sizes to be -0.93 and -1.07 , respectively. Thus, the differences between methods were minor, yet FFT effects were smaller overall. Our findings are similar to research using principal component analysis (PCA) that found delta power effect sizes to range from 0.80 to 1.31 and theta power effect sizes to range from -0.80 to -1.76^8 (Bernat et al., 2015, 2008; Bowers et al., 2018; Foti et al., 2015; Watts et al., 2017). Although these effect sizes are similar to our findings, they are on average larger—possibly due to major methodological differences (e.g., the use of PCA). It will be important for future research to explicitly compare the different methods of processing frequency amplitudes to understand the relationship between these effect sizes and these methods.

With the rise of using frequency and time–frequency decomposition techniques, it is important to note that the effect sizes for ERP analyses and frequency analyses were generally comparable, even though ERP measures are posited to be confounded by different overlapping components—an issue which may be circumvented when using time–frequency analyses. Specifically, Bernat and colleagues (2015, 2008, 2011; Nelson et al., 2011) have posited that time–frequency decompositions isolate the reward positivity from P300 contamination, resulting in a clearer observation of the underlying mechanisms of feedback processing.

8.2 | Timing

The reward positivity is typically thought to peak between 250 and 350 ms. Using a meta-analysis, Sambrook and Goslin (2015) suggested using a conservative time window of 270–300 ms but also indicated that a more liberal window of 240–340 ms was valid. Here, we found the timing of the reward positivity to center at 310 ms post feedback onset (Figure 3)—a finding in line with the existing body of literature but outside of the conservative range proposed by Sambrook and Goslin (2015). This apparent inconsistency in timing may indicate that it would be more appropriate to find the averaged peak time of participants' difference waves, as we did here, and use a mean amplitude around this time instead of using a predetermined time window. In other words, our results suggest that researchers should allow their data to guide their analyses—especially when considering how task variability across studies may affect component timing. For example, Krigolson and colleagues determined that introducing cognitive load to a time-estimation task increased variability in the timing of feedback-related ERP components (Krigolson et al., 2012).

One topic which is rarely analyzed, however, is how large of a window around this peak to use—something that is often subjectively determined by inspection of the data (see Luck, 2014, Luck & Gaspelin, 2017, and Picton et al., 2000 for alternative approaches). Here, we conducted an analysis where we systematically increased the window size of the mean peak detection window (Figure 4). We then determined the inflection point of the resulting amplitude curve and used this to characterize the optimal window size -92 ms (or ± 46 ms)—for peak detection. Based on our analysis, and as reported by others (Luck, 2014; Luck & Gaspelin, 2017; Picton et al., 2000), we would suggest using a simple mathematical analysis of data, as we did here, to objectively determine the peak time and time window of the reward positivity. Our proposed method of determining window size is especially important due to task variability across research. Here, we used a two-armed bandit task and the specific time window we determined may not generalize to other tasks (Luck & Gaspelin, 2017), just as the timing suggested by Sambrook and Goslin (2015) did not generalize to ours.

Feedback evoked bursts of delta and theta power generally range from 150 to 400 ms and from 200 to 400 ms, respectively (Bernat et al., 2008, 2011, 2015; Foti et al., 2015; Nelson et al., 2011; Watts et al., 2017, 2018). One caveat of time–frequency methods, however, is temporal smearing. Temporal smearing is an artifact of the time–frequency method wherein effects are artificially spread across time. Although temporal precision can be increased via a time–frequency parameter—the Morlet parameter—this is at the cost of distorting across frequencies. Thus, it is important to remember when considering time–frequency analyses that the timing of the underlying cognitive processes may not correspond with the timing of oscillatory clusters. In our data, we found that delta power peaked at 358 ms and theta power peaked at 420 ms post feedback onset (Figure 11). These times diverge from what was found with the reward positivity but are in line with Bernat and colleagues' (2011) claim that the underlying time–frequency correlates extend further in time than their corresponding ERP counterparts.

8.3 | Topography

The reward positivity is generally analyzed at electrodes FCz and Cz (and sometimes Fz or a pool of frontal electrodes). Indeed, our topographic findings determined that when considering the reward positivity, FCz and Cz were both the electrode sites where evoked feedback effects were greatest (Figure 5)⁹. It is interesting, however, that Fz

⁸These Cohen's d effect sizes were computed from partial eta squared measures using the following equation: $2 * \sqrt{\frac{\eta_p^2}{1-\eta_p^2}}$ (Cohen, 1988).

⁹We only present analyses at electrode FCz in this manuscript; however, we have found that analyses at electrode Cz were very similar to FCz as presented here and have included them as Supporting Information.

was considerably less prominent given that it is a popular site for analysis. Based on our findings, which were collected using a two-armed bandit task, we put forth caution against using this latter electrode site when investigating the reward positivity.

The topography of delta and theta power is also consistent across research. In congruence with research on the reward positivity, both delta and theta power are commonly investigated at electrode sites FCz and Cz. Here, we found that the peak topography when using frequency decomposition techniques was consistent with previous accounts (Bernat et al., 2008, 2011, 2015; Bowers et al., 2018; Foti et al., 2015; Nelson et al., 2011; Watts et al., 2017, 2018; Webb et al., 2017). Specifically, for FFTs the most prominent component peaks for delta power were at electrodes FCz and Cz and for theta power were at electrode FCz (Figure 9); for time–frequency delta power the peak was at Cz and for theta power it was at FCz (Figure 13)^{ix}.

8.4 | Uncharacteristic neural activity

For all measures used to examine the neural correlates of feedback processing, we have supplied descriptive statistics and effect sizes to better understand the variability that may be expected. Also, we have supplied graphics of individual data and we note that there was a subset of participants who did not follow the typical pattern of ERP, frequency, and/or time–frequency responses.

Let us first consider the reward positivity where a subset of our participants had a reversed amplitude (i.e., a negative rather than positive amplitude; Figure 3 and S1). When research uses smaller sample sizes than we did here (e.g., 30), it is often easy to disregard these reversed reward positivities by concluding these scores to be due to noise or a lack of motivation. Here, however, we see that there was a significantly sized number of individuals (17.2%) with a negative rather than positive reward positivity. In other words, for these individuals, loss feedback elicited a stronger neural response than gain feedback.

This uncharacteristic reversed pattern of results was not isolated to ERPs and was also found in delta and theta power for FFT and time–frequency analyses (Figures 7, 11, S2, and S4). Frequency decomposition techniques found an increased number of individuals with these patterns (i.e., a reversed effect for delta or theta power) ranging from 19.6% to 39.4% across FFT and time–frequency analyses. Furthermore, the reversal of one frequency band was not linked to the reversal of the other (Figures S2 and S4). Specifically, reversed delta power effects did not correspond with reversed theta power effects (or vice versa)—a result that further supports the notion that delta and theta power reflect distinct neural processes in terms of feedback processing (Bernat et al., 2011, 2015).

Our correlation analyses indicate that these uncharacteristic patterns of neural activity may be due to age and performance. There was a negative relationship between age and the reward positivity and a positive relationship between age and theta activity—indicating a reduced sensitivity to loss feedback with age (Figure 15). Although these findings are in line with other research (Hill et al., 2018), both the current findings and those by Hill and colleagues (2018) demonstrate small effects that rely on samples composed majorly of young adults—thus the true influence of age on neural feedback processing remains unclear.

Participant performance demonstrated moderate negative relations with all neural activity—indicating a reduced sensitivity to gain feedback and enhanced sensitivity to loss feedback with decreased accuracy (Figure 15). These effects cannot simply be explained as reduced participant motivation, however, as this would imply diminished neural ERP effects with reduced performance (e.g., Umemoto et al., 2019), but we demonstrated enhanced ERP effects to reduced performance. Thus, in line with other research (e.g., Williams et al., 2020), our findings indicate a clear relationship between performance and sensitivity to feedback.

Altogether, the pattern of uncharacteristic neural activity more likely reflects a continuum of neural processing rather than qualitative differences within our sample (i.e., several sub-groups). To be specific, although our analyses do categorize individuals as having standard or reversed activity across measures (implying qualitatively different populations), these were simply to demonstrate that different patterns exist. Instead, we posit that these uncharacteristic patterns of data demonstrate the variability that exists within a single population—a notion supported by our age and performance correlational analyses, demonstrating these as potential factors of variability. Future research that concerns a range of individual factors may be able to further describe the differential neural responses across humans.

8.5 | Relationship between ERPs and frequency decomposition techniques

The first note to make is that the reward positivity was associated with both delta and theta activity, albeit more strongly with delta power (Figure 14). Bernat and colleagues (2015) have also found this relationship and posited that delta activity reflects the P300 while theta activity reflects the reward positivity. They conclude this as evidence for a confounding effect on the reward positivity from the P300. Further in line with Bernat and colleagues (2015) is that delta and theta power were not associated with each other (only for time–frequency analyses within

our findings)—indicating that the reward positivity may be measuring the difference between two independent cognitive processes (see also Foti et al., 2011)—but which processes?

Delta activity is posited to reflect the processing of the complexity of rewarding events (Bernat et al., 2015; Walsh & Anderson, 2016) to motivate action selection (Cavanagh, 2015; Walsh and Anderson, 2016). In contrast, Cavanagh and Frank (2014; see also Cavanagh et al., 2012; Lin et al., 2018; Williams et al., 2019) have outlined frontal theta activity to reflect neural signals of surprise and uncertainty. Specifically, they found the presence of theta power with errors, punishment, novelty, and conflict and conclude that this is evidence of its involvement as a signal for the need for cognitive control. If, as Bernat and colleagues (2015) posit, the reward positivity is a reflection of theta activity then it is not a signal specific to learning but of a need for adaptation (Cavanagh & Frank, 2014). Thus, delta activity—and a rewarding event—motivates immediate behaviors while theta activity—and a punishing event—corresponds to behavioral adjustments to improve future performance (Cavanagh, 2015; Cavanagh & Shackman, 2015).

8.6 | Statistical power

Here, we have provided effect sizes for the reward positivity, delta power, and theta power and concomitant sample size power analyses. Before discussing appropriate sample sizes for using EEG to study feedback processing, we would like to discuss desired power in general. Due to recommendations by Cohen (1988), it is common in neuroscience to see researchers strive for a desired power of .80. In other words, the convention is to design research so that we would find a true effect 80% of the time. Over 10 years ago, Dienes (2008) criticized this convention when stating that if we hold chances of a Type I error at 5% (i.e., $\alpha = .05$), we should be holding the same standard for Type II errors (i.e., power = .95). Due to both Cohen's (1988) and Dienes' (2008) notions, here we supply a range of required sample sizes calculated by varying the levels of desired power.

It is reassuring to see that EEG experiments on feedback processing generally use adequate sample sizes. Our findings determined that when examining the reward positivity, achieving a power of .95 requires a sample of 18 participants (or 20 participants for those who prefer round numbers). Although FFT analyses do require more participants than time–frequency analyses, the required sample size is not absurd—65 participants for FFTs compared to 29 participants for time–frequency when seeking .95 power. The latter is closely in line with standard sample sizes (albeit slightly larger than many studies), yet the former is much more than what is traditionally seen in research. Thus, although FFT

analyses are attainable when investigating delta activity, they do require a significant sample size to achieve high power and might then be considered only useful when necessary. Moreover, both FFT and time–frequency theta power sample sizes indicate less than 20 participants for .95 power, comparable to ERP analyses—thus either FFT or time–frequency analyses are appropriate when investigating theta activity. With all that said, here we investigated the needed sample sizes to simply detect these neural components; larger sample sizes are needed to detect within or between-group modulations of the reward positivity, delta power, and theta power.

9 | CONCLUSIONS AND RECOMMENDATIONS

Here, we provide selected methodological comparisons and reference results for researchers investigating the ERP and oscillatory correlates of feedback processing. We have also shared the data analyzed within this article in the spirit of open science so that other researchers can investigate their own research questions about the neural correlates of feedback processing. Based on our findings, we will now provide selected methodological recommendations with the hopes of standardizing future research.

9.1 | Mean peak detection is the most appropriate method for reward positivity quantification

We investigated different methods for detecting the peak of the reward positivity ERP component. Specifically, we analyzed base-to-peak, maximal peak, and mean peak measures. First, we found that the base-to-peak technique was confounded by effects on the P200 thus demonstrating that this method does not strictly measure the reward positivity but rather assesses the difference between these two components. Second, we found noise to inflate the reward positivity amplitude when using the maximal peak detection method. Thus, in line with Luck (2014; Luck & Gaspelin, 2017) and Picton et al. (2000), we advocate against using base-to-peak and maximum peak methods for quantifying ERP components such as the reward positivity in favor of using mean peak detection methods.

9.2 | Consider both ERP and frequency analyses as they provide unique findings

The current study analyzed the neural responses of feedback evaluation with different methodological approaches—specifically, ERPs, FFTs, and time–frequency analyses. We

found both ERP and frequency analyses to determine comparable effects, but also that the reward positivity reflected multiple cognitive mechanisms. Specifically, the associations between delta power and theta power with the reward positivity but not with each other determined that this ERP component is elicited by several independent cognitive processes. Bernat and others (2011, 2015) have posited that frequency analyses better isolate underlying components than do ERP techniques. Thus, we recommend that research on feedback processing consider both methods of analysis as they can each provide unique perspectives.

9.3 | Time–frequency analyses should be preferred over FFT analyses

Regarding frequency analyses, two methods may be used: frequency and time–frequency. We investigated both of these approaches and found that even though theta power effects were comparable across the methods, FFT analyses resulted in a reduced delta effect in comparison to time–frequency analyses. Indeed, achieving adequate power with FFT delta activity would require a sample size beyond what is standard for EEG research (although it is attainable). Thus, our findings indicate that time–frequency analyses are more sensitive to the neural responses associated with feedback processing and may be more practical to use over FFT analyses when possible.

9.4 | Complex designs can be facilitated by measuring individual neural differences

The main goal of this study was to describe different neural components of feedback processing with the intent of presenting reference results for future research. We provided these by collecting data from 500 participants as they performed a two-armed bandit task.

First, we must state that the generalizability of our results is unknown. Although we used a common probabilistic reward learning task, this task is markedly different from other tasks used to elicit the reward positivity such as time estimation tasks (e.g., Williams et al., 2017), complex learning tasks (e.g., Williams et al., 2018), and deterministic learning tasks (e.g., Williams et al., 2020), thus whether our results would apply to other learning tasks is unclear. Furthermore, our participants were all university students at a Canadian university. Although student populations are common within psychology and neuroscience research, our findings may not generalize to other demographic populations and/or other nationalities. Thus, there must be caution when interpreting the results presented here within contexts outside of our research.

We do posit, however, that it may be useful to have participants complete a 5-min two-armed bandit task prior to other learning tasks. The reasoning for this is that researchers can use the results from the two-armed bandit, in conjunction with our findings here, to account for participant individual differences in feedback elicited neural responses. Including individual neural differences from a simple probabilistic learning task can help better explain effects within complex designs and when considering small effects where between-participant variability has great influence.

CONFLICT OF INTEREST

The authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

Chad C Williams: Data curation; Formal analysis; Methodology; Writing-original draft; Writing-review & editing. **Thomas D Ferguson:** Data curation; Formal analysis; Methodology; Validation; Writing-review & editing. **Cameron D Hassall:** Data curation; Methodology; Validation; Writing-review & editing. **Wande Abimbola:** Data curation; Methodology; Validation. **Olav E Krigolson:** Conceptualization; Funding acquisition; Methodology; Writing-review & editing.

ORCID

Chad C. Williams  <https://orcid.org/0000-0003-2016-5614>

Olave E. Krigolson  <https://orcid.org/0000-0002-9053-908X>

REFERENCES

- Bernat, E. M., Nelson, L. D., & Baskin-Sommers, A. R. (2015). Time-frequency theta and delta measures index separable components of feedback processing in a gambling task. *Psychophysiology*, *52*(5), 626–637. <https://doi.org/10.1111/psyp.12390>
- Bernat, E. M., Nelson, L. D., Holroyd, C. B., Gehring, W. J., & Patrick, C. J. (2008). Separating cognitive processes with principal components analysis of EEG time-frequency distributions. *Advanced Signal Processing Algorithms, Architectures, and Implementations XVIII*, 70740S. <https://doi.org/10.1117/12.801362>
- Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., & Patrick, C. J. (2011). Externalizing psychopathology and gain–loss feedback in a simulated gambling task: Dissociable components of brain response revealed by time-frequency analysis. *Journal of Abnormal Psychology*, *120*(2), 352–364. <https://doi.org/10.1037/a0022124>
- Bernat, E. M., Williams, W. J., & Gehring, W. J. (2005). Decomposing ERP time–frequency energy using PCA. *Clinical Neurophysiology*, *116*(6), 1314–1334. <https://doi.org/10.1016/j.clinph.2005.01.019>
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. <https://doi.org/10.1016/j.tics.2004.10.003>

- Bowers, M., Buzzell, G., Bernat, E., Fox, N., & Barker, T. (2018). Time-frequency approaches to investigating changes in feedback processing during childhood and adolescence. *Psychophysiology*, *55*(10), e13208. <https://doi.org/10.1111/psyp.13208>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897x00357>
- Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods*, *137*(2), 321–332. <https://doi.org/10.1016/j.jneumeth.2004.03.002>
- Cavanagh, J. F. (2015). Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times. *NeuroImage*, *110*, 205–216. <https://doi.org/10.1016/j.neuroimage.2015.02.007>
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. <https://doi.org/10.1016/j.tics.2014.04.012>
- Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology-Paris*, *109*(1–3), 3–15. <https://doi.org/10.1016/j.jphysparis.2014.04.003>
- Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*(2), 220–238. <https://doi.org/10.1111/j.1469-8986.2011.01293.x>
- Champely, S. (2018). *Pwr: Basic functions for power analysis* v1.2-2. <https://CRAN.R-project.org/package=pwr>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT press.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.
- 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*, 447–455. [https://doi.org/10.1016/0013-4694\(91\)90062-9](https://doi.org/10.1016/0013-4694(91)90062-9)
- Foti, D., & Weinberg, A. (2018). Reward and feedback processing: State of the field, best practices, and future directions. *International Journal of Psychophysiology*, *132*(Pt B), 171–174. <https://doi.org/10.1016/j.ijpsycho.2018.08.006>
- Foti, D., Weinberg, A., Bernat, E. M., & Proudfit, G. H. (2015). Anterior cingulate activity to monetary loss and basal ganglia activity to monetary gain uniquely contribute to the feedback negativity. *Clinical Neurophysiology*, *126*(7), 1338–1347. <https://doi.org/10.1016/j.clinph.2014.08.025>
- Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from non-rewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, *32*(12), 2207–2216. <https://doi.org/10.1002/hbm.21182>
- Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1990). The error-related negativity: An event-related brain potential accompanying errors. *Psychophysiology*, *27*, S34.
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Physiological Science*, *4*, 385–390. <https://doi.org/10.1111/j.1467-9280.1993.tb00586.x>
- Hill, K. E., Oumeziane, B. A., Novak, K. D., Rollock, D., & Foti, D. (2018). Variation in reward-and error-related neural measures attributable to age, gender, race, and ethnicity. *International Journal of Psychophysiology*, *132*, 353–364. <https://doi.org/10.1016/j.ijpsycho.2017.12.009>
- Holroyd, C. B., & Coles, M. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679. <https://doi.org/10.1037/0033-295x.109.4.679>
- Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, *44*(6), 913–917. <https://doi.org/10.1111/j.1469-8986.2007.00561.x>
- Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. <https://doi.org/10.1037/a0038339>
- Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, *45*(5), 688–697. <https://doi.org/10.1111/j.1469-8986.2008.00668.x>
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. <https://doi.org/10.1016/j.tics.2011.12.008>
- Krigolson, O. E. (2018). Event-related brain potentials and the study of reward processing: Methodological considerations. *International Journal of Psychophysiology*, *132*, 175–183. <https://doi.org/10.1016/j.ijpsycho.2017.11.007>
- Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). How we learn to make decisions: Rapid propagation of reinforcement learning prediction errors in humans. *Journal of Cognitive Neuroscience*, *26*(3), 1–10. https://doi.org/10.1162/jocn_a_00509
- Krigolson, O. E., Heinekey, H., Kent, C. M., & Handy, T. C. (2012). Cognitive load impacts error evaluation within medial-frontal cortex. *Brain Research*, *1430*, 62–67. <https://doi.org/10.1016/j.brainres.2011.10.028>
- Lin, H., Saunders, B., Hutcherson, C. A., & Inzlicht, M. (2018). Midfrontal theta and pupil dilation parametrically track subjective conflict (but also surprise) during intertemporal choice. *NeuroImage*, *172*, 838–852. <https://doi.org/10.1016/j.neuroimage.2017.10.055>
- Luck, S. J. (2014). *An Introduction to the event-related potential technique*. MIT press.
- Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. <https://doi.org/10.1111/psyp.12639>
- Luft, C. (2014). Learning from feedback: The neural mechanisms of feedback processing facilitating better performance. *Behavioural Brain Research*, *261*, 356–368. <https://doi.org/10.1016/j.bbr.2013.12.043>
- Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a “generic” neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798. <https://doi.org/10.1162/jocn.1997.9.6.788>
- Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. Oxford University Press.

- Nelson, L. D., Patrick, C. J., Collins, P., Lang, A. R., & Bernat, E. M. (2011). Alcohol impairs brain reactivity to explicit loss feedback. *Psychopharmacology (Berl)*, *218*(2), 419–428. <https://doi.org/10.1007/s00213-011-2323-3>
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. <https://doi.org/10.1111/1469-8986.3720127>
- Proudfit, G. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, *52*(4), 449–459. <https://doi.org/10.1111/psyp.12370>
- Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, *141*(1), 213–235. <https://doi.org/10.1037/bul0000006>
- Sambrook, T. D., & Goslin, J. (2016). Principal components analysis of reward prediction errors in a reinforcement learning task. *NeuroImage*, *124*, 276–286. <https://doi.org/10.1016/j.neuroimage.2015.07.032>
- Umamoto, A., Inzlicht, M., & Holroyd, C. B. (2019). Electrophysiological indices of anterior cingulate cortex function reveal changing levels of cognitive effort and reward valuation that sustain task performance. *Neuropsychologia*, *123*, 67–76. <https://doi.org/10.1016/j.neuropsychologia.2018.06.010>
- Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, *36*(8), 1870–1884. <https://doi.org/10.1016/j.neubiorev.2012.05.008>
- Watts, A., Bachman, M. D., & Bernat, E. M. (2017). Expectancy effects in feedback processing are explained primarily by time-frequency delta not theta. *Biological Psychology*, *129*, 242–252. <https://doi.org/10.1016/j.biopsycho.2017.08.054>
- Watts, A., Tootell, A. V., Fix, S. T., Aviyente, S., & Bernat, E. M. (2018). Utilizing time-frequency amplitude and phase synchrony measure to assess feedback processing in a gambling task. *International Journal of Psychophysiology*, *132*, 203–212. <https://doi.org/10.1016/j.ijpsycho.2018.04.013>
- Webb, C. A., Auerbach, R. P., Bondy, E., Stanton, C. H., Foti, D., & Pizzagalli, D. A. (2017). Abnormal neural responses to feedback in depressed adolescents. *Journal of Abnormal Psychology*, *126*(1), 19. <https://doi.org/10.1037/abn0000228>
- Williams, C. C., Hassall, C. D., Lindenbach, T., & Krigolson, O. E. (2020). Reward prediction errors reflect an underlying learning process that parallels behavioural adaptations: A trial-to-trial analysis. *Computational Brain & Behavior*, *3*, 189–199. <https://doi.org/10.1007/s42113-019-00069-4>
- Williams, C. C., Hassall, C. D., Trska, R., Holroyd, C. B., & Krigolson, O. E. (2017). When theory and biology differ: The relationship between reward prediction errors and expectancy. *Biological Psychology*, *129*, 265–272. <https://doi.org/10.1016/j.biopsycho.2017.09.007>
- Williams, C. C., Hecker, K. G., Paget, M. K., Coderre, S. P., Burak, K. W., Wright, B., & Krigolson, O. E. (2018). The applications of reward learning in the real world: Changes in the reward positivity amplitude reflect learning in a medical education context. *International Journal of Psychophysiology*, *132*, 236–242. <https://doi.org/10.1016/j.ijpsycho.2017.10.010>
- Williams, C. C., Kappen, M., Hassall, C. D., Wright, B., & Krigolson, O. E. (2019). Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *NeuroImage*, *189*, 574–580. <https://doi.org/10.1016/j.neuroimage.2019.01.048>
- Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience*, *24*(28), 6258–6264. <https://doi.org/10.1523/jneurosci.4537-03.2004>
- Zimmerman, M. E. (2011). Normative data. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology*. Springer.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

FIGURE S1 ERP waveforms at electrode FCz illustrating the reward positivity. A: grand averaged conditional waveforms with 95% confidence intervals of participants with a traditional reward positivity, B: grand averaged conditional waveforms with 95% confidence intervals of participants with a reversed reward positivity

FIGURE S2 FFT line plots at electrode FCz illustrating delta and theta activity. A: grand averaged conditional waveforms with 95% confidence intervals of participants with reversed delta activity, B: grand averaged conditional waveforms with 95% confidence intervals of participants with reversed theta activity

FIGURE S3 Illustration of individual time peak data and error terms for time–frequency derived analyses for frequencies with effects. Each black dot represents a participant’s data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, 3 standard deviations

FIGURE S4 Time–frequency wavelets at electrode FCz illustrating delta and theta activity. A: grand averaged conditional wavelets of participants with reversed delta activity, B: grand averaged conditional wavelets of participants with reversed theta activity. Contour lines demonstrate significant clusters derived from the full dataset

TABLE A1 Descriptive statistics and effect sizes for the reward positivity, delta, and theta across ERP, FFT, and time–frequency analysis methods. All values calculated as the difference between gain and loss conditions (gain-loss). 95% CI = 95% confidence intervals

TABLE A2 Descriptive statistics and effect sizes for conditional (gain, loss) amplitudes of the reward positivity, delta, and theta across ERP, FFT, and time–frequency analysis methods. 95% CI = 95% confidence intervals. All measures are a comparison of the conditional activity versus zero

TABLE A3 Required within-subject sample size calculations for different levels of desired power for reward positivity, delta, and theta measures. Values calculated from Cohen's *d* effect sizes as determined using difference data (gain-loss)

TABLE A4 Required within-subject sample size calculations for different levels of desired power for each frequency. Values calculated from Cohen's *d* effect sizes as determined using difference data (gain-loss)

TABLE A5 Correlational results between the reward positivity, delta activity, and theta activity. 95% CI = 95% confidence intervals

TABLE A6 Statistical results of gender (female, male) effects on difference neural activity. Difference = female – male. 95% CI = 95% confidence intervals

TABLE A7 Statistical results of gender (female, male) effects on conditional neural activity. Difference = female – male. 95% CI = 95% confidence intervals

TABLE A8 Correlational results of age and accuracy effects on difference neural activity. 95% CI = 95% confidence intervals

TABLE A9 Correlational results of age and accuracy effects on conditional neural activity. 95% CI = 95% confidence intervals

FIGURE A1 Topographic map of electrode placements and labels

FIGURE A2 ERP waveforms at electrode FCz illustrating the reward positivity. A: grand averaged conditional waveforms with 95% confidence intervals, B: grand averaged difference waveform with 95% confidence intervals, C: heatmap (two dimensional representation where colour corresponds to magnitude) of amplitude for each participant across time

FIGURE A3 Illustration of individual data and error terms for the reward positivity. A: Difference amplitudes for each amplitude extraction technique, B: conditional amplitudes for the mean peak measure, C: peak timing of the reward positivity. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, 3 standard deviations

FIGURE A4 EEG amplitudes using mean (blue) and maximal (red) measures with different window sizes around the reward positivity peak (A) and within baseline activity (B). Crossbar represents mean and 95% confidence intervals

FIGURE A5 Topographic maps of grand averaged data for the reward positivity. Left: amplitudes from the difference waveform, middle: amplitudes from the gain waveform, right: amplitudes from the loss waveform. Top: traditional topographic maps, Bottom: percentage (%) of participants

where the reward positivity peaked at the respective electrode

FIGURE A6 FFT line plots at electrode FCz illustrating delta and theta effects. A: grand averaged conditional line plots with 95% confidence intervals, B: grand averaged difference line plot with 95% confidence intervals, C: heatmap (two dimensional representation where colour corresponds to magnitude) of power for each participant across frequencies

FIGURE A7 Illustration of individual data and error terms for delta and theta EEG power. A: Difference amplitudes for delta and theta, B: conditional amplitudes for delta and theta. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, 3 standard deviations

FIGURE A8 Illustration of individual power data and error terms for FFT derived analyses. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, 3 standard deviations

FIGURE A9 Topographic maps of grand averaged data for delta and theta EEG power from FFT analyses. From left to right: delta power of the difference between gain and loss, theta power of the difference between gain and loss, delta power for the gain condition, and theta power for the loss condition. Top: traditional topographic maps, Bottom: percentage (%) of participants where delta or theta power peaked at the respective electrode

FIGURE A10 Time–frequency wavelets at electrode FCz illustrating delta and theta power. A: grand averaged conditional wavelet for the gain condition, B: grand averaged conditional wavelet for the loss condition, C: wavelets grand averaged across both the gain and loss conditions, and D: grand averaged difference wavelet. Contour lines were determined using the collapsed isolator method

FIGURE A11 Illustration of individual data and error terms for delta and theta power. A: Difference amplitudes for delta and theta power, B: conditional amplitudes for delta and theta power, C: peak amplitude times for delta and theta power. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, 3 standard deviations

FIGURE A12 Illustration of individual power data and error terms for time–frequency derived analyses for frequencies with effects. Each black dot represents a participant's data and the middle black lines represent the mean. Rectangles

surrounding the mean correspond to error bars. From widest (nearest to the mean) to thinnest (farthest from the mean): 95% confidence intervals, 1 standard deviation, 2 standard deviations, 3 standard deviations

FIGURE A13 Topographic maps of grand averaged data for delta and theta power from time–frequency analyses. From left to right: delta power of the difference between gain and loss, theta power of the difference between gain and loss, delta power for the gain condition, and theta power for the loss condition. Top: traditional topographic maps, Bottom: percentage (%) of participants where delta or theta power peaked at the respective electrode

FIGURE A14 Scatterplots between different measures of neural activity. A–C: frequency data correspond to FFT measures; D–F: frequency data correspond to time–frequency measures

FIGURE A15 Scatterplots of age (top) and accuracy (bottom) with difference neural activity. B, C, G, H: frequency

data correspond to FFT measures; D, E, I, J: frequency data correspond to time–frequency measures. Blue lines correspond to linear regressions

FIGURE A16 Scatterplots of age (top) and accuracy (bottom) with conditional neural activity. A, B, G, H: ERP data correspond to gain and loss waveforms of the reward positivity, C, D, I, J: frequency data correspond to FFT measures; E, F, K, L: frequency data correspond to time–frequency measures. Blue lines correspond to linear regressions

How to cite this article: Williams CC, Ferguson TD, Hassall CD, Abimbola W, Krigolson OE. The ERP, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study. *Psychophysiology*. 2021;58:e13722. <https://doi.org/10.1111/psyp.13722>