Neural prediction errors depend on how an expectation was formed

Blake W. Saurels[1], Tonya Frommelt[1], Kielan Yarrow[2], Ottmar V. Lipp[1, 3], & Derek H. Arnold[1]

1. *School of Psychology, University of Queensland, Australia.*
2. *Department of Psychology, City, University of London, United Kingdom*
3. *School of Psychology and Counselling, Queensland University of Technology, Australia.*

Keywords: Prediction, Prediction Error, Expectation, Confirmation Bias, Oddball Effect, Declared Predictions

Running Title: Neural prediction errors

Final Submission Date: October 2021

Corresponding Author: Blake W. Saurels

e-mail: blake.saurels@uq.net.au

Abstract

When a visual event is unexpected, because it violates a train of repeated events, it excites a greater positive electrical potential at sensors positioned above occipital-parietal human brain regions (the P300). Such events can also seem to have an increased duration relative to repeated (implicitly expected) events. However, recent behavioural evidence suggests that when events are unexpected because they violate a declared prediction – a guess – there is an opposite impact on duration perception. The neural consequences of incorrect declared predictions have not been examined. We replicated the finding whereby repetition violating events elicit a larger P300 response. However, we found that events that violated a declared prediction entrained an opposite pattern of response – a smaller P300. These data suggest that the neural consequences of a violated prediction are not uniform but depend on how the prediction was formed.

Significance statement

The authors demonstrate that how a prediction is formed dictates how the brain responds to unexpected visual information. They replicate the classic finding whereby the P300 wave, detected via EEG at occipital-parietal sensors, is larger following violations of trains of repetitive (implicitly expected) events. Using the same paradigm, they find that events that violate a declared, explicit prediction produce an opposite effect – a smaller P300.

INTRODUCTION

There is great contemporary interest in the neural processes that govern our capacity to predict the near future. Neuroscientists have proposed that this capacity is perhaps the primary function of the human brain, and findings regarding the outcome of perceptual predictions have inspired prominent theories of brain function [1-4].

Recent behavioural evidence has suggested that different types of predictions can have opposite effects on time perception [5-8]. Events that are presumed to be expected due to repetition, although participants tend not to be quizzed about their expectations, can seem to have a shorter duration relative to unexpected 'oddball' events [6-8]. However, if people are asked to guess (to declare) which of two equally likely events might eventuate, then events that confirm a declared prediction can seem to last longer relative to events that are not predicted [5].

In terms of brain activity, neural responses to events that are unexpected, because they break a chain of repeated events (which we will refer to as a *repetition*-based prediction), tend to be exaggerated relative to responses to repeated events [9-12]. For example, a neural signature of violations of repetition-based predictions is a greater positive electrical potential recorded by sensors positioned above occipital-parietal regions of the human brain, peaking ~300ms from event onset – the P300 [13-15].

It has been suggested that neural response magnitudes may typically scale with perceived duration [6, 16, 17]. So repeated visual events may evoke a reduced amount of visual and temporal information processing in the visual system [6]. The exact mechanism for these processes are unknown, but a neural correlate of exaggerated perceived durations has been

identified - the P300, which is enhanced for events that seem longer due to violations of a repetition-based prediction [13].

When we considered the reversed impact of expectation on time perception for declared predictions [5] and repetition-based predictions [6-8], we were motivated to ask if there might be a similar reversed impact on brain activity – will responses elicited by events that confirm a declared-prediction be opposite to responses elicited by events that confirm a repetition-based prediction? To the best of our knowledge, this question has not been examined previously.

We conducted two experiments. In Experiment 1, participants were shown sequences of brief repetitive events (flashed red or green circles; see Figure 1A for a graphic depicting this experimental protocol). Before the first presentation of each sequence, participants guessed what colour the first event would be – a declared prediction. The final event of the sequence was either a repeat of all preceding events, or an oddball (the opposite colour) – encouraging a repetition-based prediction. The results of this experiment are suggestive of a reversed neural outcome for violations of repetition-based predictions and declared predictions.

The design of Experiment 1 had the benefit of closely matching many similar past investigations, at the cost of confounding the effects of repetition with the effects of the probability of encountering a specific type of event (repeated events were both more common and were presumed to be implicitly predicted). In Experiment 2 we adopted a design that broke this confound. Participants viewed sequential pairs of flashed coloured circles (see Figure 2A for a graphic depicting this experimental protocol). Participants guessed the colour of both sequential events before they were presented – a declared prediction. Repetitions of the same

colour were used to explore the effects of repetition, independent of the impact of the probability of encountering a given type of event.

Across both experiments, we found that while P300s were exaggerated for events that violated repetition-based predictions, they were diminished for events that violated declared predictions – a reversed pattern of neural response which we were encouraged to predict based on findings relating to time perception [5-8].

EXPERIMENT 1

METHODS AND DATA ANALYSIS

Experimental procedures, participant numbers, exclusion criteria and analyses for this experiment were pre-registered (https://aspredicted.org/8s3re.pdf), and the research was approved by the University of Queensland ethics committee and conducted in accordance with the Declaration of Helsinki. We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. We have also distinguished the unregistered procedures and analyses from the pre-registered ones.

Participants

A target *N* of 40 participants was pre-registered. A total of 44 participants were recruited for testing via a research participation scheme at the University of Queensland (in exchange for course credit). Four were excluded according to pre-registered exclusion criteria (see data analysis section), leaving a total of 40 (our target *N*; 13 males). All reported having normal or

corrected-to-normal visual acuity. Ages for the final sample ranged from 18 to 33 (*M* ~ 20, *SD* ~ 3).

Stimuli and Apparatus

Stimuli consisted of coloured circles, with a diameter subtending ~17 degrees of visual angle. Green and red circles were matched in luminance (green circles CIE: 0.2858, 0.5939, 8.9311; red circles CIE: 0.6337, 0.3117, 8.9311). Stimuli were presented on an ASUS VG248QE 3D Monitor, using a Cambridge Research Systems ViSaGe stimulus generator and custom Matlab R2015b software [18]. The monitor had a resolution of 1920 x 1080 pixels and was set at a refresh rate of 60Hz. Participants viewed stimuli from 57cm, from directly in front of the monitor with their chin placed on a chin rest. A Biosemi International ActiveTwo system was used to record EEG data (sampling rate: 1024Hz, channel array: 64).

Procedure

Each trial consisted of between five and seven 50ms coloured circle presentations (hereafter referred to as *events*; the precise number was determined at random on a trial-by-trial basis) separated by 500ms inter-stimulus intervals (ISIs).

Each trial began with the participant making a declared prediction regarding the colour of the first event in the forthcoming sequence, by using a mouse to click on the word ‘red’ or ‘green’ (a declared prediction). The colour of the first event in each sequence was contingent on the participant’s declared prediction, set to match it with a probability of ~70%. After the declared prediction, a fixation dot was presented and then, after a variable delay (of between 2 and 3 seconds), the presentation sequence began. The events in the sequence were either all the same

colour (a repeat trial) or all the same colour except the last event, which was oppositely coloured (an oddball trial, encouraging a repetition-based prediction, see Figure 1). Half of the 280 individual trials were repeats, and half were oddballs, both interleaved in a pseudo-random order during a single block of trials. Participants were encouraged to try and detect an unspecified pattern used to determine the colour of the first event in each sequence, to enhance guessing behaviour and attention to whether their guess had been correct or not. After each trial, participants reported if the final circle was the 'same' or 'different' relative to all preceding events, again using a mouse to make their selection. This was done to enhance and monitor attention to the final presentation of each sequence.

Data Analysis

During pre-processing, data were high (1Hz), low (100Hz), and band-stopped (45 – 55Hz) filtered. Data were then subjected to an independent component analysis, implemented by the FieldTrip toolbox for Matlab, [18, 19] to remove blink artefacts (positive patterns of activity at frontal electrodes, often the strongest component). Electrode activity was then average referenced, to correct for baseline skin conductance levels. Data were then sorted into trials, and events within trials, via custom Matlab code.

The response period for each event (450ms) within each trial was baseline-corrected relative to the average of activity recorded by each sensor during a period 100ms prior to the onset of physical events. Individual participant datasets were then averaged according to trial type (i.e. Oddball vs. Repeat – using the last event in each trial sequence, and Incorrect vs. Correct declared predictions – using the first event of each sequence) – producing average wave forms for a 550ms window (-100ms to +450ms from event onset), for each trial type, and for each participant.

Individual differences were then calculated between responses to Oddballs and Repeats (i.e. for violations of repetition-based predictions), and between Incorrect and Correct guesses (i.e. for violations of declared predictions). These two sets of individual difference wave forms were then compared via a cluster-based permutation analysis conducted using the FieldTrip toolbox for Matlab (settings: Monte Carlo method; test statistic set as the maximum of the cluster-level statistics; alpha = .05; cluster alpha = .05; randomisations: 1000; minimum number of neighbourhood channels required sample inclusion in clustering algorithm: 3). This test is designed to detect epochs, after the physical onset of different types of events, where the neural consequences of violations of declared predictions differ from the neural consequences of violations of repetition-based predictions.

RESULTS

Our cluster test identified a positive occipito-parietal cluster (see Figure 1C) extending from 346 until 394ms after event onsets (cluster electrodes: C1, CP1, P1, PO3, O1, Iz, Oz, POz, Pz, CPz, Cz, CP2, P2, P4, PO8, PO4, & O2) where the neural consequences of violations of declared predictions were not interchangeable with the neural consequences of violations of repetition-based predictions. We used the FieldTrip toolbox to create a topographic heat map, depicting the results of this cluster test, in Figure 1C.

Figure 1B plots differences between oddballs and repeats (red line – representing violations of repetition-based predictions) and between incorrect and correct guesses (purple line – representing violations of declared predictions) averaged across all electrodes that contributed to the significant cluster test. This plot highlights a reversed pattern of response (see Supplemental Figure 1 for the event-related potentials (ERPs) used to calculate these two

neural difference waves) for violations of declared predictions and for violations of repetition-based predictions. The period corresponding with the significant cluster test corresponds with a positive for the violation of repetition-based predictions, consistent with a P300 [14, 15], which has been associated with apparent duration dilations for oddballs [13]. Another commonly examined ERP for repetition violations is the mismatch negativity (MMN), which occurs between ~150-300ms [11, 12, 20, 21]. We did find a negative difference between oddballs and repeats that peaked within this time window (225-275ms, $t_{39} = 7.45$, $p < .001$, with a Bayes factor analysis revealing decisive evidence for the alternative hypothesis, that there would be a difference, $BF_{10} > 1000$).

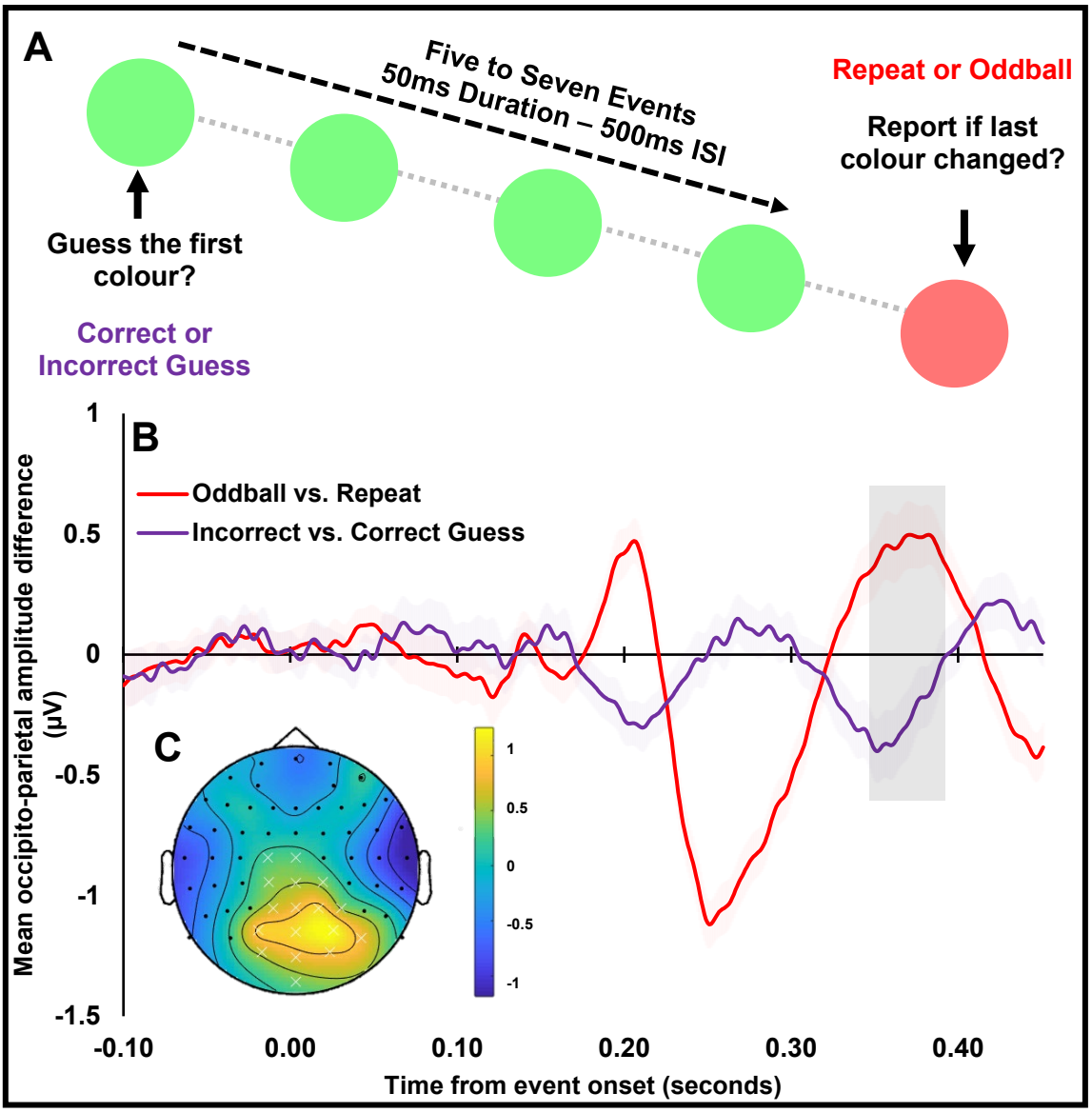

Figure 1. **A)** Graphic depicting the protocol for Experiment 1. **B)** Mean amplitude difference scores taken from the occipito-parietal electrodes identified by the cluster analysis, from 100ms prior to event onset to 450ms after event onset. Difference scores calculated as unexpected outcome (oddball or incorrect guess) minus the expected outcome (repeat or correct guess). Error bars depict ±1 SE across participants for each time point. **C)** Heat map depicting the occipito-parietal cluster identified by the cluster analysis, at 346 to 394ms (shaded bar). White Xs mark the cluster electrodes. A positive cluster (yellow on heat map, taken as red line minus purple line) indicates that the repetition violation difference was greater than the intuition violation difference.

DISCUSSION

In Experiment 1 we found that violations of declared predictions produced an opposite pattern of neural responses (recorded from occipito-parietal sensors) relative to violations of repetition-based predictions. A cluster analysis identified a period between ~350-400ms post-event onsets that differentiated these two types of prediction violations. This period was associated with a

greater positive deflection for oddballs relative to repeat presentations (i.e., for violations of repetition-based predictions), but diminished deflection for events that violated a declared prediction relative to events that confirmed these predictions. This reversed pattern is consistent with our hypotheses, which were motivated by the reversed impact of violations of these different types of predictions on measures of time perception [5-8].

In Experiment 1 we adopted a traditional visual oddball paradigm, which allowed us to examine the effects of violations of repetition-based predictions using a protocol that was commensurate with past investigations [6-8, 10-12], and then contrast this pattern of response to the consequences of violations of declared predictions. However, the design of Experiment 1 had a number of features that we thought could be improved.

First, in Experiment 1 there was a difference in sequence position between our measures of declared predictions (the first event of each sequence) and our measures of repetition-based predictions (the final event of each sequence). In Experiment 2 we therefore adopted a simplified two sequential event protocol, and all analyses related to the *second* of these two events – eliminating this temporal confound. Second, event durations in Experiment 1 were very brief (50ms), which introduced offset transients within analysis periods. While this cannot account for the conditional reversal of neural response patterns, it does undermine the equivalence of our data with typical measures of visual evoked potentials [11-15], so in Experiment 2 we used longer event presentations (600ms) and constrained analyses to this time window.

The third and most important conceptual issue in relation to Experiment 1 was that our measure of repetition-based prediction violations had (at least) two contributing factors – the effect of

repetition per se [which has been linked to adaptation, e.g., 22] vs change, and the effect of the probability of encountering a given type of event. To break this confound, in Experiment 2 we included three blocks of trials in each experimental session, one in which event types (repetition vs change) were equally likely, and two in which one of the two types of events was more probable (80%). Moreover, we informed participants of these contingencies at the outset of each of these blocks of trials. In sum, this allowed us to independently assess the impact of repetition per se, the impact of the probability of events being encountered (regardless of the event type), and the impact of guessing (i.e., of declared predictions) – all by having participants guess the entire sequence before each trial began.

EXPERIMENT 2

METHODS AND DATA ANALYSIS

Experimental procedures, participant numbers, exclusion criteria and analyses for this experiment were pre-registered (https://aspredicted.org/9r6fn.pdf), and the research was approved by the University of Queensland ethics committee and conducted in accordance with the Declaration of Helsinki. We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. We have also distinguished the unregistered procedures and analyses from the pre-registered ones.

Participants

A target *N* of 42 participants was pre-registered. These participants were recruited for testing via a research participation scheme at the University of Queensland (in exchange for course

credit or monetary reimbursement of $40). No participants were excluded according to pre-registered exclusion criteria (see data analysis section). However, 3 participants were excluded due to corruption of the EEG data files – leaving a sample of 39 participants, 16 males. All reported having normal or corrected-to-normal visual acuity. Ages for the final sample ranged from 18 to 47 ($M \sim 23$, $SD \sim 5$).

Stimuli and Apparatus

All features of the stimuli and apparatus were the same as Experiment 1, except the diameter of the circles subtended ~23 degrees of visual angle (35% larger) – to increase the ERP signal strength.

Procedure

Each trial consisted of two sequential 600ms circle presentations. Presentations were separated by a 500-1500ms ISI, with the precise ISI duration determined at random on a trial-by-trial basis.

To manipulate repetition, circles on each trial were either the same colour (Green then Green, or Red then Red – the Repeat condition), or the second colour was different (Green then Red, or Red then Green – the Change condition).

To manipulate statistical likelihood, the experiment was split into three blocks of 100 individual trials, with each block containing a different ratio of Repeat to Change trials: either 80:20, 20:80, or 50:50. Trial blocks were presented in an order that was counter-balanced across participants. More probable trials of unbalanced blocks constituted the ‘More Likely’ condition, and less probable trials the ‘Less Likely’ condition. Trials within balanced 50:50

blocks constituted the 'Equally Likely' condition. Participants were informed of the probability that would prevail in each block before the first block trial.

Guessing (declarations) was manipulated by having participants predict what sequence of colours they were about to see (e.g., red – red, or red – green). To encourage participants to engage in this task, they were told there would be an underlying pattern to the order of sequence presentations that they might discern if they paid close attention. To reduce discouragement due to poor guessing, the first event colour was set to the participants' first 'guess' with a probability of 80%. So, participants were seldom wrong in their guess regarding the first colour. There were 300 trials total, with 100 per block.

Data analysis

Pre-processing and data handling was the same as Experiment 1. The final ten trials from the experimental session where we tested the first participant were corrupted and have been excluded from the analysis.

The response period for each event (600ms) within each trial was baseline-corrected to the average activity taken from the period 100ms prior, for each electrode, for each participant. The data were then averaged across trials for each participant.

Initially, mean difference wave forms were calculated in the same manner as Experiment 1, and activity for the electrodes determined by the original cluster analysis were plotted in Figure 2A. The mean difference scores within a restricted time window, 300-500ms (determined based on previous research examples [13-15]), were taken for each trial type, for each participant. Each of these three sets of difference scores was subject to frequentist and Bayesian paired

sample *t*-tests. All analyses pertained to the second event, and when comparing correct to incorrectly guessed events, we took only trials where participants had correctly guessed the first event colour. We had originally planned to analyse these data independently, via a cluster analysis. However, as Experiment 2 was a conceptual replication of Experiment 1, with the added benefit that the results were temporally comparable to previous literature, we decided that using the same spatial cluster as Experiment 1 was more appropriate.

The first follow-up analysis involved all the same steps, except data were first averaged as sub-conditions – 4 (guess outcome: both correct, first correct only, second correct only, or both incorrect) x 2 (sequence option: Repeat, or Change) x 3 (likelihood: Likely option, Unlikely option, or Balanced). There was a different number of trials within each of these, thus some trials were given a greater weighting than others. However, this method balanced the influence of each sub-condition – meaning the difference scores were not unfairly influenced by sub-conditions with more trials (e.g., correctly guessed sequences that were the statistically more likely sequence). The second follow-up analysis involved the same steps as the original method, except trials from the balanced block only were included.

RESULTS

Using the same occipito-parietal electrodes identified by the cluster analysis of Experiment 1, we again plot separate difference scores, but this time for each of 3 factors, the effect of change (vs repetition, red line Figure 2B), the effect of incorrect guessing (for the second event only, vs correctly guessing both, purple line), and the effect of encountering an improbable (vs a probable, blue line) type of event. All data relate to neural responses to the *second* of the two sequential events on each trial (see Supplementary Figure 2 for original ERPs that were used

for these difference scores). These difference plots illustrate a reversed pattern for the effects of change (vs repetition; red line) vs incorrect guessing (vs correct; purple).

The longer presentation times of Experiment 2 avoided offset transients during the analysis window, which allowed us to test for ERPs described in past similar investigations [e.g. 13-15]. Of central interest is the P300, which has been associated with the recognition of odd events [14, 15], and the strength of apparent time dilations for oddballs [13]. The electrode cluster we identified in Experiment 1, and used for these analyses, is consistent with previous similar investigations [e.g., 13-15]. For the critical time window of similar investigations [300 – 500ms; e.g. 11-15], we found positive differences for change sequences (relative to repeats, see Figure 2C; $t_{38} = 4.42$, $p < .001$), with a Bayes factor analysis revealing decisive evidence for the alternative hypothesis, that there would be a difference ($BF_{10} = 304.29$). We also found an oppositely signed negative difference for incorrect (vs correct) guesses ($t_{38} = 2.93$, $p = .006$), with a Bayes factor analysis revealing substantial evidence for the alternative hypothesis, that there would be a difference ($BF_{10} = 6.72$). Finally, we found a negative difference for improbable (vs probable) types of event sequences ($t_{38} = 2.43$, $p = .02$), with a Bayes factor analysis revealing only anecdotal evidence for the alternative hypothesis, that there would be a difference ($BF_{10} = 2.3$). Note that analyses for this last result were restricted to data from our two unbalanced blocks of trials, whereas other analyses incorporated data from all three blocks. Also, all data relate to the status of the second event, as we only analysed trials where participants had correctly guessed the first colour.

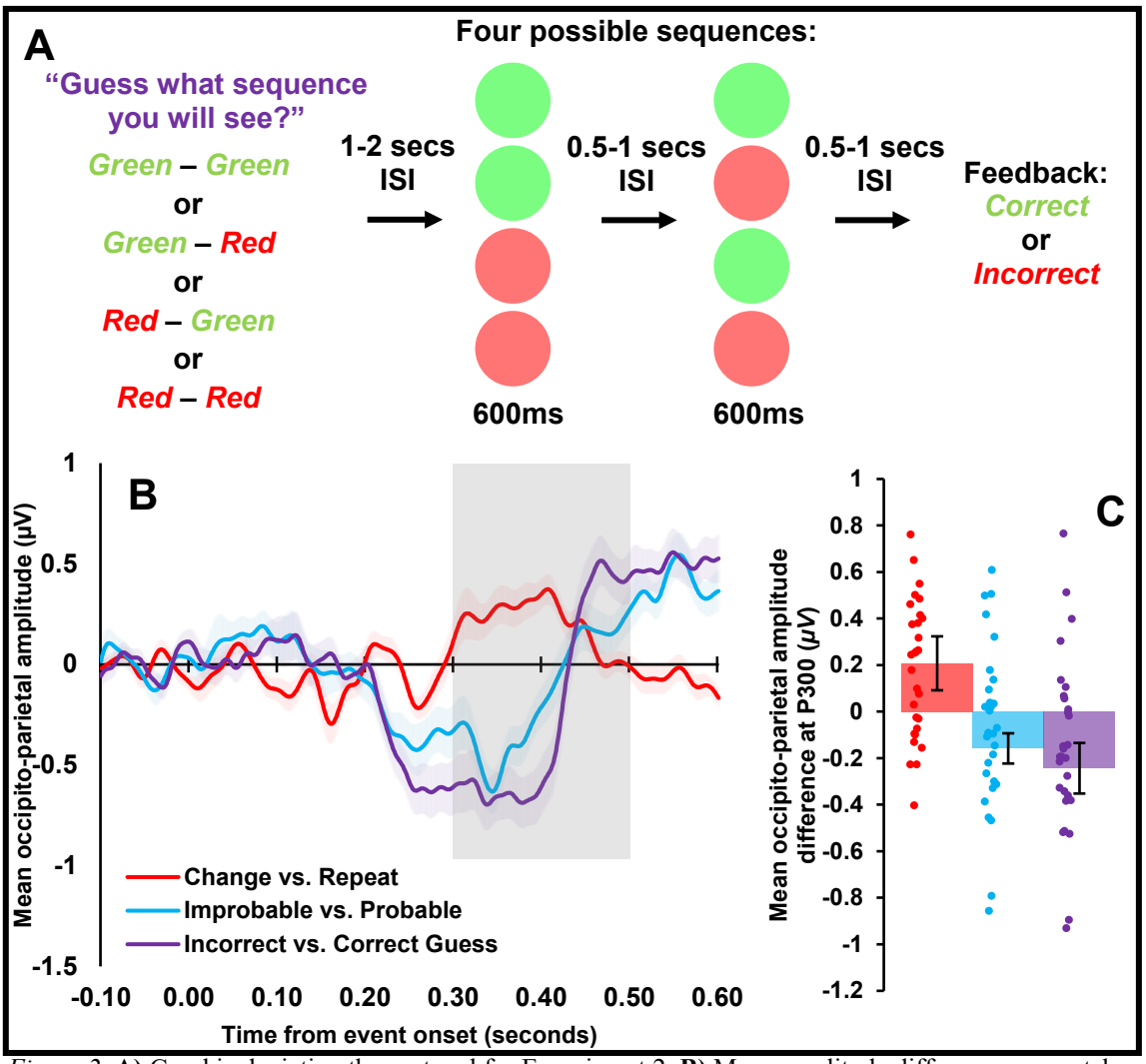

Figure 2. **A)** Graphic depicting the protocol for Experiment 2. **B)** Mean amplitude difference scores taken from the occipito-parietal electrodes identified by the Experiment 1 cluster analysis, from 100ms prior to event onset to 600ms after event onset. Difference scores calculated as unexpected outcome (change, improbable outcome, or incorrect guess) minus the expected outcome (repeat, probable outcome, or correct guess). Error bars depict ±1 SE across participants for each time point. **C)** Mean amplitude difference scores taken from the occipito-parietal electrodes, averaged over the 300-500ms post-event onset time window. Error bars depict ±1 SE across participants. Individual dots depict individual participant scores.

Differences contingent on guess outcome and event probability were similar, with both negatively signed at the outset of our targeted analysis window (see Figure 2B purple and blue lines). This outcome might reflect the non-independence of these two conditions – participants were more likely to guess that the more probable event sequence would eventuate (i.e., a repeat or a change sequence) in the two of three blocks where one type of event was more common. While this would not necessarily result in correct guessing (as the colours in each sequence could equally be red or green), knowledge of the more likely event sequence nonetheless

increased the probability of correct guessing. Percent correct in unequal probability block ($M \sim 55.9\%$, $SD \sim 4\%$) was greater than in equal probability blocks ($M \sim 39.9\%$, $SD \sim 4.2\%$, $t_{38} = 16.89$, $p < .001$), with a Bayes factor analysis revealing decisive evidence for the alternative hypothesis, that there would be a difference ($BF_{10} > 1000$).

To tease apart the impact of guessing from probable (vs improbable) event types, we conducted two additional analyses. First, we replicated our first analysis, but individually averaged data for each sub-condition. For instance, data were independently averaged for correctly guessed repeat trials from blocks where these events were more likely, and from blocks when these events were less likely, before calculating a conditional average from these two estimates for each participant. This applies an equal weighting in analyses to confounding sub-factors (i.e., probable and improbable event sequences contribute equally to estimates of the effect of guessing, and correct and incorrect guesses contribute equally to estimates of the effect probability) at the cost of overweighting performance on less likely types of trials. All data from all trials were included here. Difference scores computed via this method are shown in Figure 3A. For the P300 time window, change sequences produced a positive difference relative to repeats (see Figure 3B; $t_{38} = 2.99$, $p = .005$), with a Bayes factor analysis revealing substantial evidence for the alternative hypothesis, that there would be a difference ($BF_{10} = 7.68$), and incorrect guesses produced negative differences relative to correct ones ($t_{38} = 2.87$, $p = .007$), with a Bayes factor analysis revealing substantial evidence for the alternative hypothesis, that there would be a difference ($BF_{10} = 5.86$). However, there was no discernible difference for less (vs more) probable event type sequences ($t_{38} = 0.87$, $p = .388$), with a Bayes factor analysis revealing substantial evidence for the null hypothesis, that there would be no difference ($BF_{10} = 0.25$).

Our second method for dissociating the impact of declared predictions from event type probability was to restrict analyses to the balanced (50:50) block of trials (100 trials per participant). Analyses of data from these blocks of trials revealed the same difference score reversal (see Figure 3C-D). For the P300 time window, change sequences produced a positive difference relative to repeats (t_{38} = 2.375, p = .023), with a Bayes factor analysis revealing anecdotal evidence for the alternative hypothesis, that there would be a difference (BF_{10} = 2.03), and incorrect guesses produced negative differences relative to correct ones (t_{38} = 3.34, p = .002), with a Bayes factor analysis revealing strong evidence for the alternative hypothesis, that there would be a difference (BF_{10} = 17.7).

One might reasonably ask if participants attempted to predict the second event colour in unbalanced blocks of trials, or if they had simply always chosen the more likely sequence contingent on their first guess (i.e., in blocks of trials where repeats were more likely, a participant might guess at the first colour, and then always predict that colour would repeat). This 'no attempt to guess at the second colour strategy' would deliver a percentage correct on unbalanced blocks of 64%. To explain, by design participants correctly guessed the first colour on 80% of trials and would then incorrectly predict the more likely event would ensue on 20% of these trials (i.e., 0.8 - (0.8 x 0.2) = 0.64). If, however, participants had attempted to guess at the second colour – informed by instructions regarding the more likely sequence (on 80% of trials), we would predict that they should obtain an overall percentage correct on unbalanced blocks of 54%. Again, by design, participants correctly guessed the first colour on 80% of trials (there were 100 trials per unbalance block). Of these, 64 were the more likely sequence, which would be guessed correctly at a probability of 80%. The other 16 trials would be the less likely sequence, which would only be correctly guessed at a probability of 20% (i.e., 0.8*0.8*0.8 +

0.8*0.2*0.2 = 0.544). Actual performance on unbalanced blocks of trials was 55.9% (*SD* ~ 4%).

We compared actual levels of performance on unbalanced blocks of trials against a prediction for a 'no attempt to guess the second colour' strategy (i.e., 64% correct) via a Bayes factor analysis and found decisive evidence for the alternative hypotheses that there would be a difference ($BF_{10} > 1000$). We compared actual levels of performance on unbalanced blocks of trials against the prediction for a 'guessing the second colour' strategy (i.e., 54.4% correct) via a Bayes factor analysis and found only anecdotal evidence for a difference in performance ($BF_{10} = 2.711$). Overall, these data provide conclusive evidence against participants simply choosing the more likely sequence on every trial of unbalanced blocks, although they may have guessed at the more likely outcome slightly more than our task instructions encouraged (although the evidence for that is anecdotal).

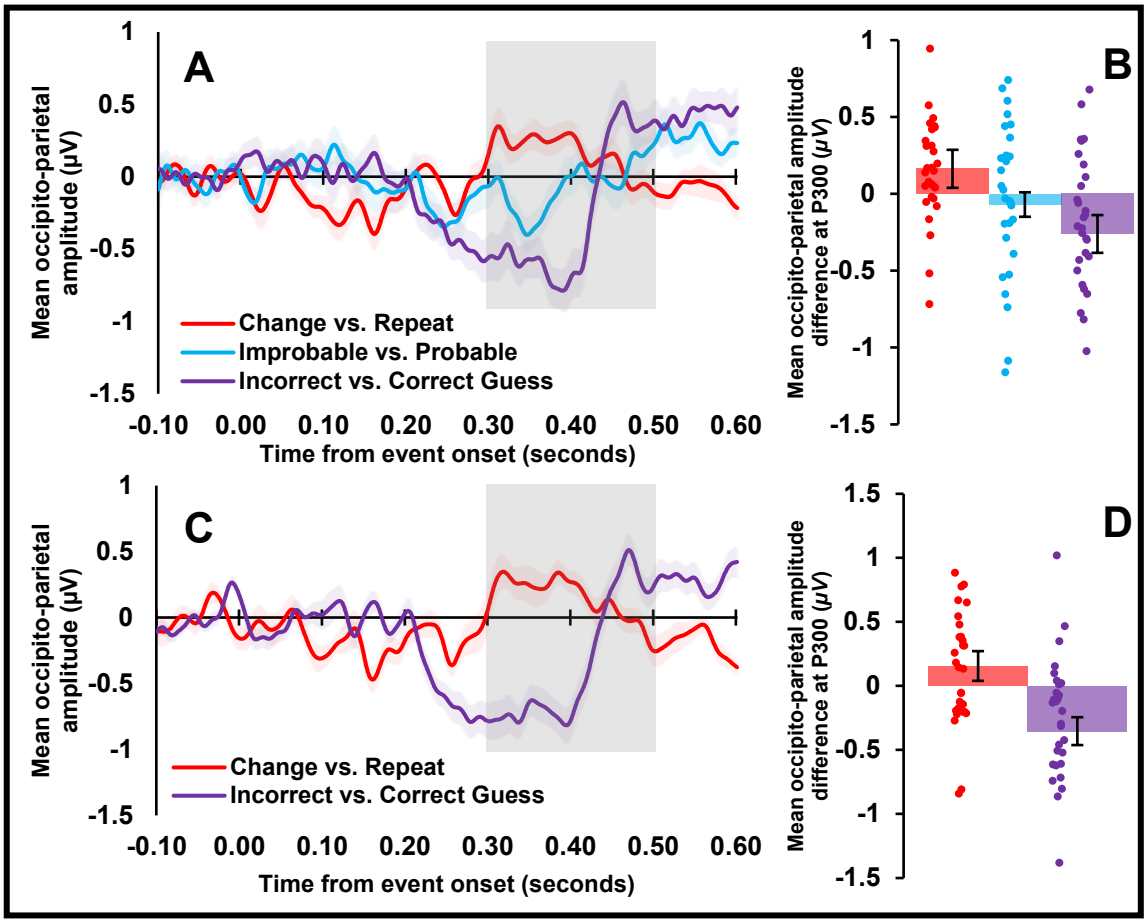

Figure 3. **A)** Mean amplitude difference scores taken from the occipito-parietal electrodes identified by the Experiment 1 cluster analysis, from 100ms prior to event onset to 600ms after event onset. Difference scores calculated as unexpected outcome (change, improbable outcome, or incorrect guess) minus the expected outcome (repeat, probable outcome, or correct guess). Error bars depict ±1 SE across participants for each time point. Data from all blocks were included here. **B)** Mean amplitude difference scores taken from the occipito-parietal electrodes, averaged over the 300-500ms post-event onset time window. Error bars depict ±1 SE across participants. Individual dots depict individual participant scores. Data from all blocks were included here. **C)** Depicts the same type of data as Figure 3A, except only from trials in the block where there was no frequency imbalance for repeats and oddballs. **D)** Depicts the same type of data as Figure 3B, but again, using only trials from the equal block.

GENERAL DISCUSSION

We found that violations of declared predictions produced an opposite pattern of occipito-parietal neural response relative to the combined influence of change (vs repetition) and event type probability (Experiment 1), and to the isolated influence of change (vs event repetition, Experiment 2). We predicted these outcomes based on recent behavioural evidence, suggesting an analogous reversal in the influence of these factors on time perception [5-8].

It has been suggested that prediction is a primary function of the human brain [3, 4]. There are, however, different factors that can encourage a prediction [1, 2]. We have examined the consequences of declared predictions and have distinguished these from the implicit effects of repetition-based predictions [6-8]. A key strength of our work is that we have carefully distinguished the effect of repetition per se from the probability of encountering a given type of event, and have further distinguished both of these effects of declared predictions (see Experiment 2). Intriguingly, with this handling of data, only declared predictions and the effects of repetition (independent of the probability of encountering a given type of event) had a robust (opposite) impact on measures of neural activity. However, the procedure needed to tease apart these different effects (switching from a standard oddball paradigm with many repetitions, to a protocol with just 1 potential repetition) may have considerably weakened the impact of event probability – so we would not question conclusions drawn regarding event probability and expectation from past protocols.

While we have not examined time perception in this study, our data are consistent with findings regarding the neural correlates of time dilations for oddballs vs repeated events – an exaggerated P300 [13]. We have similarly found exaggerated P300s for correctly guessed events (vs incorrectly guessed events, see Figures 2B, 3A and 3C), and correctly guessed events

are also associated with time dilations [5]. In our data these relatively exaggerated P300s for correctly guessed events are expressed as negative deflections, as we gave precedence in our analyses to data from (unexpected) incorrectly guessed trials, in order to match oddball (vs repeat) analyses typical of other studies [e.g., 13-15]. Future research with a targeted protocol could examine a possible direct link between P300 strength and a perceived duration exaggeration for events that confirm declared predictions.

Prediction violations have also been associated with a MMN – that is, with a more downward deflection of voltage recorded by EEG sensors from ~100 to 300ms post stimulus onset when an input unexpectedly breaks a train of identical repeated events [11, 12, 20, 21]. The MMN was originally detected for auditory events but has also been found with trains of repetitive visual events [20, 21]. Like our first experiment, these studies have tended to confound the effects of adaptation and event probability – as both of these effects can be expected to result from having identical repeated visual presentations. When the impact of sensory adaptation is controlled for, it is unclear if the visual MMN persists [23]. Our data seem to echo these observations. In Experiment 1, oddball presentations were associated with an exaggerated negative deflection ~250ms post event onset, relative to repeated presentations, whereas in Experiment 2 we did not find any evidesnce for a robust impact of event probability, after having controlled for the impact of event repetition. Instead of a visual MMN, our data are more consistent with a later modulation – an exaggerated P300.

The impact of declaring a prediction in our data is reminiscent of findings regarding the consequences of confirmation bias [24, 25]. Bias-confirming events elicit greater selective attention, as people monitor input for evidence that confirms their expectations [24-27]. This might be reflected in greater visual and temporal processing, consistent with evidence that bias-

confirming events seem longer [5]. This mirrors a previous explanation for why humans extract less visual information from repeats events [7] and they seem shorter [6-8], which it has been suggested corresponds to a reduction in visual and temporal processing [6].

Patterns of neural response elicited by predictions have been shown to be reversible – contingent on attention [28, 29]. When an event is attended, because it is behaviourally relevant, the typical attenuation of sensory responses elicited by predicted events can be reversed. In our protocol, no condition was given preferential attentional weighting by instruction, but the requirement that people predict future events might have encouraged biased attention to confirmatory evidence. Hence the reversed neural response patterns we observe for events confirming declared predictions might be a conceptual echo of findings regarding the neural consequences of confirmation bias [24-27].

Studies examining the neural correlates of violations of repetition-based predictions have been enormously influential in the formation of contemporary theories of brain function [1-4]. Here, we have examined these in relation to what might seem a conceptually similar situation – violations of declared predictions. We have found that violations of these different types of predictions seem to encourage opposite neural consequences, and that to some degree, the neural consequences previously attributed to violations of repetition-based predictions [6-8, 11-15] might be due in large part to the effects of repetition per se. These findings discourage the idea that the neural consequences of predictions might be uniform [see 30] and encourage scrutiny of contemporary theories of brain function.

Overall, our data show that different factors that can promote an expectation have markedly different neural consequences. Consistent with previous research [e.g., 13-15], we have found

that when repetition and probability effects are combined, there is a relatively enhanced response recorded by occipital-parietal sensors ~300ms post stimulation. By contrast, our novel finding is that explicit declarations encourage a reversed response. Overall, these data suggest that the neural consequences of expectation depend on how an expectation is formed – dispelling the notion of a uniform mapping between prediction outcomes and neural consequences.

REFERENCES

1. Bubic, A., von Cramon, D. Y., & Schubotz, R. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4*(25), doi:10.3389/fnhum.2010.00025
2. Downing, K. L. (2007). Predictive models in the brain. *Connection Science, 21*, 39-74. doi:10.1080/09540090802610666
3. Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815-836. doi:10.1098/rstb.2005.1622
4. Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159*(3), 417-458. doi:10.1007/s11229-007-9237-y
5. Birngruber, T., Schröter, H., Schütt, E., & Ulrich, R. (2018). Stimulus expectation prolongs rather than shortens perceived duration: Evidence from self-generated expectations. *Journal of Experimental Psychology: Human Perception and Performance, 44*(1), 117-127. doi:10.1037/xhp0000433
6. Pariyadath, V., & Eagleman, D. (2007). The effect of predictability of subjective duration. *PLoS ONE, 2*(11), e1264. doi:10.1371/journal.pone.0001264
7. Saurels, B. W., Lipp, O. V., Yarrow, K., & Arnold, D. H. (2019). Predictable events elicit less visual and temporal information uptake in an oddball paradigm. *Attention, Perception, & Psychophysics, 82*(23), 1074-1087. doi:10.3758/s13414-019-01899-x
8. Tse, P., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. Perception & Psychophysics, 66(7), 1171-1189. doi:10.3758/BF03196844
9. Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *Journal of Neurophysiology, 99*(6), 2877-2886. doi:10.1152/jn.90376.2008

10. Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*(1), 14-23. doi:10.1016/j.tics.2005.11.006

11. Kimura, M., Schröger, E., & Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *Neuroreport, 22*(14), 669-673. doi:10.1097/WNR.0b013e32834973ba

12. Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: A review. *Biological Psychology, 63*(3), 199–236. doi:10.1016/s0301-0511(03)00049-8.

13. Ernst, B., Reichard, S. M., Riepl, R. F., Steinhauser, R., Zimmermann, S. F., & Steinhauser, M. (2017). The P3 and the subjective experience of time. *Neuropsychologia, 103*, 12-19. doi:10.1016/j.neuropsychologia.2017.06.033

14. Levi-Aharoni, H., Shriki, O., Tishby, N. (2020). Surprise response as a probe for compressed memory states. *PLoS Computational Biology, 16*(2), e1007065. doi:10.1371/journal.pcbi.1007065

15. Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148. doi:10.1016/j.clinph.2007.04.019

16. Matthews, W. J. (2015). Time perception the surprising effects of surprising stimuli. *Journal of Experimental Psychology: General, 144*(1), 172-197. doi:10.1037/xge0000041

17. Matthews, W. J., & Gheorghiu, A. I. (2016). Repetition, expectation, and the perception of time. *Current Opinion in Behavioral Sciences, 8*, 110–116. doi:10.1016/j.cobeha.2016.02.019

18. The Math Works, Inc. (2015). *MATLAB* (Version 2015b) [Computer software]. https://www.mathworks.com/

19. Oostenveld, R., Fries, P., Maris, E., & Schoffelen, JM. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*(156869). doi:10.1155/2011/156869

20. Cammann, R. (1990). Is there a mismatch negativity (MMN) in visual modality? *Behavioral and Brain Sciences, 13*(2), 234–235. doi:10.1017/S0140525X00078420

21. Czigler, I., & Csibra, G. (1990). Event-related potentials in a visual discrimination task: Negative waves related to detection and attention. *Psychophysiology, 27*(6), 669– 676. doi:10.1111/j.1469-8986.1990.tb03191.x

22. Webster, M. A. (2015). Visual adaptation. *Annual Review of Visual Science, 1*(1), 547-567. doi:10.1146/annurev-vision-082114-035509

23. Male, A. G., O'Shea, R. P., Schröger, E., Müller, D., Roeber, U., & Widmann, A. (2020). The quest for the genuine visual mismatch negativity (vMMN): Event-related potential indications of deviance detection for low-level visual features. P*sychophysiology, 57*(6), e13576. doi: 10.1111/psyp.13576

24. Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. doi:10.1037/1089-2680.2.2.175

25. Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313–329. doi:10.1016/0001-6918(78)90006-9

26. Rajsic, J., Wilson, D. E., & Pratt, J. (2015). Confirmation bias in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 41*(5), 1353-1364. doi:10.1167/14.10.927

27. Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., & Gardner, J. L. (2016). Adaptable history biases in human perceptual decisions. *Proceedings of the National*

Academy of Sciences of the United States of America, 113(25), E3548-E3557. doi:10.1073/pnas.1518786113

28. de Lange, F. P., Kok, P., Rahnev, D., & Jehee, J. F. M. (2011). Attention reverses the effect of prediction in silencing sensory signals. *i-Perception, 2*(4), 341-341. doi:10.1068/ic341

29. Summerfield, C., Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Science, 13*(9), 403-409. doi:10.1016/j.tics.2009.06.003

30. Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences, 17*, 521-534. doi:10.1007/s11097-017-9525-z

Author contributions:

Conceptualization: BWS, DHA
Data curation: BWS
Formal analysis: BWS
Funding acquisition: DHA
Investigation: BWS, TF
Methodology: BWS, DHA
Project administration: BWS
Resources: DHA
Software: BWS, DHA
Supervision: DHA
Validation: DHA
Visualization: BWS
Writing – original draft: BWS
Writing – review & editing: BWS, TF, KY, OVL, DHA

Competing interests: The authors declare that they have no competing interests.

Funding: The research was support by ARC Discovery Project Grant GA68476 awarded to DHA.

Data availability statement: All datasets and analysis scripts can be found at: https://doi.org/10.48610/ad94a19.

SUPPLEMENTAL MATERIAL

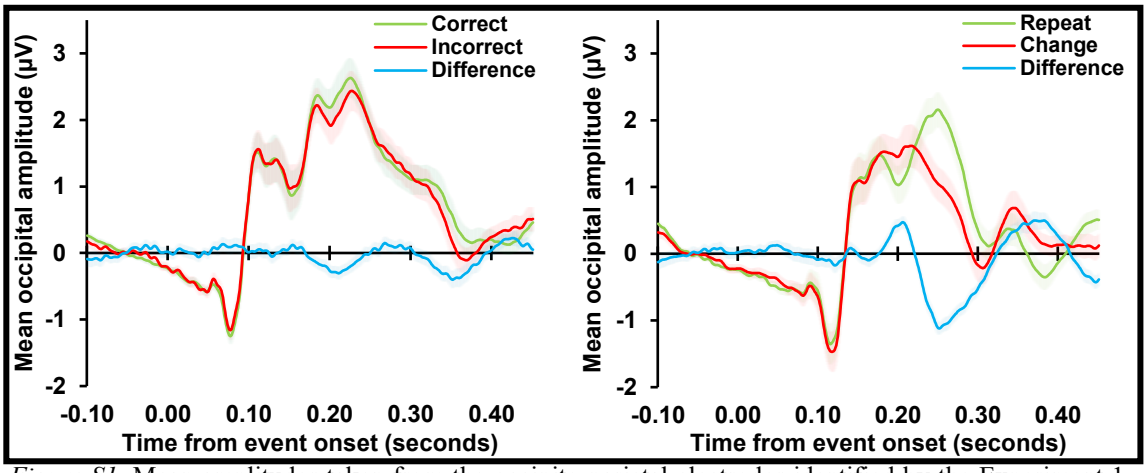

Figure S1. Mean amplitudes taken from the occipito-parietal electrodes identified by the Experiment 1 cluster analysis, from 100ms prior to event onset to 450ms after event onset. Green (red) lines depict the expected (unexpected) outcome in each case, and blue lines depict the difference between these, taken as unexpected minus expected. Error bars depict ±1 SE across participants for each time point.

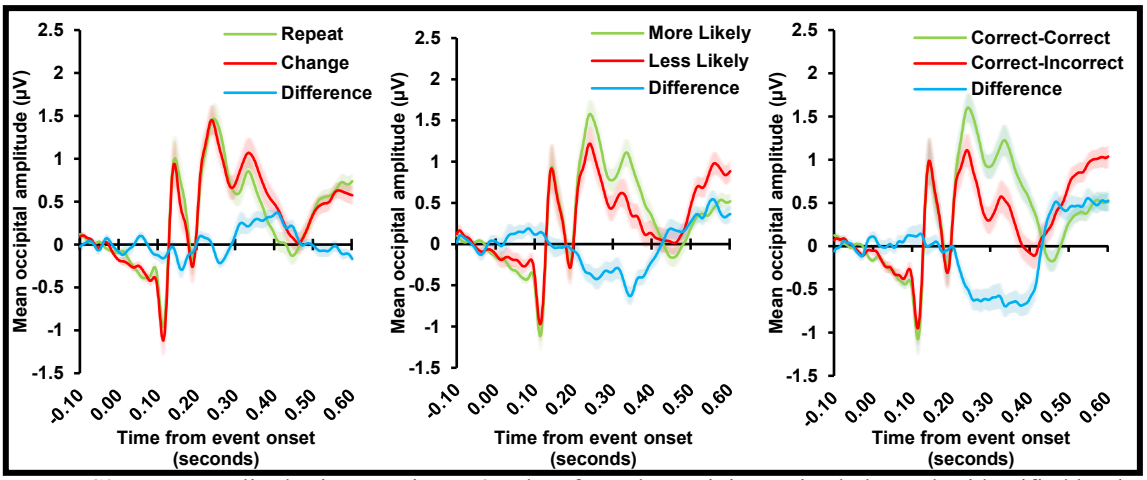

Figure S2. Mean amplitudes in Experiment 2, taken from the occipito-parietal electrodes identified by the Experiment 1 cluster analysis, from 100ms prior to event onset to 650ms after event onset. Green (red) lines depict the expected (unexpected) outcome in each case, and blue lines depict the difference between these, taken as unexpected minus expected. Error bars depict ±1 SE across participants for each time point.