Neuropsychologia 146 (2020) 107538

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Feedback processing is enhanced following exploration in continuous environments

Cameron D. Hassall *, Olave E. Krigolson

Centre for Biomedical Research, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada

ARTICLE INFO

Keywords:
Electroencephalography
Computational modelling
Explore-exploit dilemma
P300

ABSTRACT

Decision-making is typically studied by presenting participants with a small set of options. However, real-world behaviour, like foraging, often occurs in continuous environments. The degree to which human decision-making in discrete tasks generalizes to continuous tasks is questionable. For example, successful foraging comprises both exploration (learning about the environment) and exploitation (taking advantage of what is known). Although progress has been made in understanding the neural processes related to this trade-off in discrete tasks, it is currently unclear how, or whether, the same processes are involved in continuous tasks. To address this, we recorded electroencephalographic data while participants "dug for gold" by selecting locations on a map. Participants were cued beforehand that the map contained either a single patch of gold, or many patches of gold. We then used a computational model to classify participant responses as either exploitations, which were driven by previous reward locations and amounts, or explorations. Our participants were able to adjust their strategy based on reward distribution, exploring more in multi-patch environments and less in single-patch environments. We observed an enhancement of the feedback-locked P300, a neural signal previously linked to exploration in discrete tasks, which suggests the presence of a general neural system for managing the explore-exploit trade-off. Furthermore, the P300 was accompanied by an exploration-related enhancement of the late positive potential that was greatest in the multi-patch environment, suggesting a role for motivational processes during exploration.

1. Introduction

There is a growing body of literature on how animals – including humans – manage the trade-off between exploiting prior experience and exploring new options. The explore-exploit trade-off is affected by several factors, including environmental volatility (Behrens et al., 2007), stress (Lenow et al., 2017), the total number of remaining decisions (Wilson et al., 2014), and reward distribution (Constantino and Daw, 2015). The effect of reward distribution on exploration rate is of particular interest because it cuts across multiple species. For instance, snail communities are affected by the patchiness, or spatial clustering, of available food (Chase et al., 2001). Highly patchy environments, which are more heterogenous, are dominated by snail species that tend to explore ("grazers"). Conversely, less patchy environments, which are more homogenous, are dominated by snail species that tend to exploit ("diggers").

Unlike snails, which can only be grazers or diggers, humans can flexibly adjust their exploration rate. For example, Constantino and Daw (2015) observed that individuals will tailor their patch-leaving decisions to the current reward distribution; thus, our decision-making is adaptable to the environment. However, individual factors also play a role. For instance, patch-leaving strategies in a simulated fishing game show considerable inter-subject variability (Hutchinson et al., 2008). There, participants were asked to make a series of decisions – to either fish or switch ponds – and were told that the number of fish in each pond might vary. Due to response variability, and contrary to the authors' predictions, patch-leaving decisions were unaffected by reward distribution (Hutchinson et al., 2008). It is therefore unclear in what way humans are able to use reward distribution knowledge, if at all.

The brain presumably plays a role in our ability to adjust how often we explore. The neural basis of exploration has been studied using a variety of neuroimaging techniques, including electroencephalography (EEG). Early work suggests that a machine learning classifier can use the EEG at frontal and parietal sites to accurately predict whether an individual will explore or exploit (Bourdaud et al., 2008; Tzovara et al., 2012). Similarly, we identified a parietal component of the event-related

* Corresponding author. Centre for Biomedical Research, University of Victoria, P.O. Box 1700 STN CSC, Victoria, British Columbia, V8W 2Y2, Canada.
E-mail address: chassall@uvic.ca (C.D. Hassall).

https://doi.org/10.1016/j.neuropsychologia.2020.107538
Received 16 August 2019; Received in revised form 31 May 2020; Accepted 12 June 2020
Available online 20 June 2020

potential (ERP) called the P300 that is associated with decisions to explore (Hassall et al., 2013, 2019). In Hassall et al. (2013) participants pressed a button to simultaneously inflate a balloon and increase a pot of money (the Balloon Analogue Risk Task, or BART: Lejuez et al., 2002). If the balloon burst, the accumulated money for that round was lost. Participants learned, through trial-and-error, how much a balloon could be safely inflated before a burst became likely. Exploration in the BART has been defined as a decision to continue pumping the balloon following a long pause, while exploitations are fast and automatic (Pleskac and Wershbale, 2014). We observed that the feedback/response-locked P300 was enhanced prior to explorations.

Because feedback and response (i.e., balloon pump) occur simultaneously in the BART, it was unclear in Hassall et al. (2013) which event – feedback or response – drove the exploration-related P300. To address this, in Hassall et al. (2019) we used a very different kind of task in which participants had to learn which of two options yielded a greater average reward (a "two-armed bandit"). Explorations were identified using a reinforcement-learning (RL) model fit to our participants' data. Importantly, response and feedback were separated temporally. Again, the feedback-locked P300 amplitude was greater before exploration than exploitation. This suggested that the exploration-related P300 generalizes across tasks and is driven by feedback, not the motor response.

In both experiments, we interpreted the P300 as indicative of a phasic release of the neuromodulator norepinephrine from locus coeruleus (the LC-NE P300 or LC-P3 hypothesis: Nieuwenhuis et al., 2005). According to the LC-P3 hypothesis, phasic NE activity (and the associated P300 deflection in the ERP) results from processing salient events. This account integrates previous P300 results by considering the motivational significance of the eliciting event (stimulus or feedback). For example, P300 effects are driven by stimulus frequency, novelty, and task relevance (Nieuwenhuis et al., 2005; Polich, 2007). Additionally, the feedback-locked P300 is modulated by reward magnitude (enhanced for large rewards compared to small rewards; Yeung and Sanfey, 2004) and decision type (enhanced for exploratory feedback; Hassall et al., 2013, 2019).

To summarize our previous work: we observed an exploration-related P300 in two different decision-making tasks, the BART and a two-armed bandit. A potential confound in both experiments was that exploitation was the more frequent strategy (exploration rate was 3% in the BART and 20% in the two-armed bandit: Hassall et al., 2013, 2019). Other work has shown that the frequency of a decision type can affect subsequent feedback processing. For example, the risk-related P300 (enhanced feedback processing following risky decisions) is modulated by rate of risk taking. When risk-taking is rare, the risk-related P300 is enhanced (Zheng et al., 2015; Zheng and Liu, 2015). Thus, it was unclear from our previous work whether the enhanced neural feedback processing associated with exploration was due to exploration per se, or to the fact that exploration was rare. To test whether the exploration-related P300 reflects the switch from a frequent to a rare mode of decision-making, as we originally argued, we designed a task in which *exploration* would be the dominant strategy. We had participants search for sparse but spatially-correlated rewards on a continuous two-dimensional map. By "sparse" we mean that the majority of possible responses resulted in little or no reward; our hope in designing the task this way was to encourage more exploration than exploitation. If the exploration-dependent P300 enhancement we observed previously was due to the relative infrequency of exploration, then we ought to observe an *exploitation*-dependent P300 enhancement in the current study. On the other hand, if our previous results replicate, we would conclude that the effect is not due to frequency, but rather to some other property of exploration.

Another issue with our previous studies was that our analyses focused on the effect of *upcoming* trial type on the feedback-locked P300. Our motivation for this choice came from previous machine learning work showing that it was possible to predict whether someone would explore or exploit on the next trial by examining the pre-response EEG (Bourdaud et al., 2008; Tzovara et al., 2012). We had not considered the possibility of an effect of *current* trial type (i.e., whether the participant had just explored or exploited), nor had we considered the possibility of an interaction between current-trial type and next-trial type. In other words, is feedback processing driven more by what we just did, or what we will do? To answer this secondary research question, we decided to examine the effect of current/next trial type on the feedback-locked P300 in the present study.

Additionally, we manipulated the distribution of rewards across blocks. Participants were cued that they would encounter either many reward patches (a multi-patch environment) or one reward patch only (a single-patch environment). The purpose of this manipulation was to further test the hypothesis that the exploration-dependent P300 effect is due to the frequency of exploration relative to exploitation. Stimulus frequency is known to modulate P300 amplitude such that the more infrequent a stimulus is, the larger the P300 that is elicited (Duncan--Johnson and Donchin, 1977). Our hope was that, like snails, our participants would explore more in highly patchy environments, and exploit more in less patchy environments. If a greater exploration-dependent P300 was observed when exploration was less frequent (i.e., in the single-patch environment), this would lend support to a frequency hypothesis. Cues were used to help encourage these behaviours, since previous research on our ability to adapt to different reward distributions is mixed (Constantino and Daw, 2015; Hutchinson et al., 2008).

Finally, our P300 analysis suggested a more sustained difference between exploratory and exploitative feedback compared to previous work (Hassall et al., 2013, 2019). Based on an exploration of our data, we identified a difference in the time range of the late positive potential (LPP), a P300-like ERP component that is also linked to motivational significance (Olofsson et al., 2008; Schupp et al., 2000). This is perhaps unsurprising, given that the P300 is thought to be a major subcomponent of the LPP (Foti et al., 2009; Hajcak and Foti, 2020; MacNamara et al., 2009; Weinberg and Hajcak, 2011). Previously, enhanced LPPs have been seen for emotional compared to neutral images (Schupp et al., 2000), for monetary gains compared to monetary losses (Broyd et al., 2012), for large rewards compared to small rewards (Meadows et al., 2016), and for unambiguous compared to ambiguous decisions (Sun et al., 2017). Later, we will discuss post hoc explanations for our LPP result in light of this previous work.

2. Material and methods

2.1. Participants

Twenty-four university-aged participants (3 male, all right-handed, M_{age} = 21.00, 95% CI [19.30, 22.70] with no known neurological impairments and with normal or corrected-to-normal vision took part in the experiment. All of the participants received credit in an undergraduate course for their participation. Additionally, participants were paid a performance-dependent bonus of up, M_{bonus} = $10.11, 95% CI [8.89, 11.35]. The participants provided informed consent approved by the Human Research Ethics Board at the University of Victoria.

2.2. Apparatus and procedure

Participants were seated 60 cm in front of a 22-inch LCD display (75 Hz, 2 ms response rate, 1680 by 1050 pixels, LG W2242TQ-GF, Seoul, South Korea). Visual stimuli were presented using the Psychophysics Toolbox Extension (Brainard, 1997; Pelli, 1997) for MATLAB (Version 8.3, Mathworks, Natick, USA). Participants were given written and verbal instructions to minimize head and eye movements throughout the experiment.

Participants played 40 rounds of "gold rush", a mining simulator in which the goal was to find as much gold as possible. Each round

consisted of 20 trials. In each trial, participants used a mouse to select a map location at which to dig for gold. Participants were then shown a point total from 1 to 100, representing the amount of gold they had found. The total amount of gold found was tracked for each round, and at the end of the experiment the participant was paid for their best round at a conversion rate of $0.01 per point.

Prior to beginning the experiment, participants were shown on-screen instructions indicating that the distribution of gold was spatially correlated. The distribution of gold was fixed within a round but changed between rounds. Two types of reward distribution were possible: single-patch, and multi-patch. In single-patch maps, rewards were concentrated at one map location. Multi-patch maps contained rewards concentrated at between four and six "peaks". Participants were unaware of the total number of peaks in the multi-patch maps, only that there was more than one. The maximum reward at each peak was randomly chosen from a uniform distribution from 50 to 100 points. Thus, each map for each environment had a mean maximum reward of 75 points. Peak locations were also randomly chosen (uniform distribution). Participants were unaware of the maximum reward available within each map, only that there was always a "best" location to dig. The distribution of rewards around each peak was Gaussian, computed using the MATLAB function mvnpdf (Statistics and Machine Learning Toolbox, Release 2014a, Mathworks, Natick). The Gaussian reward distributions were circular (i.e., identity covariance matrix). See Supplementary Material for participant instructions, and for examples of each map type.

Prior to each round, participants were shown a cue indicating whether the upcoming map was single-patch or multi-patch. The meaning of these cues was explained in the on-screen instructions (Supplementary Material). Participants completed two practice rounds – one for each reward distribution type. On each trial, participants were shown the outline of the map (the dig boundary), a centrally-presented fixation cross, and an 'x' at each previous dig location. After each practice round, participants were shown the underlying reward distribution, with their choices overlaid. During the experiment, participants were never shown the underlying reward distribution. See Fig. 1 for a block/trial overview.

2.3. Data collection

Sixty-three channels of EEG data, referenced to channel AFz, were recorded using Brain Vision Recorder (Version 1.20, Brain Products GmbH, Munich, Germany). Sixty-one electrodes were placed in a fitted cap according to the 10–20 system. Additionally, two electrodes were attached to the left and right mastoids. Conductive gel was used to ensure that electrode impedances were below 20 kΩ prior to recording, and the EEG data were sampled at 500 Hz and amplified (actiCHamp, Brain Products GmbH, Munich, Germany) with a 245 Hz antialiasing low-pass filter.

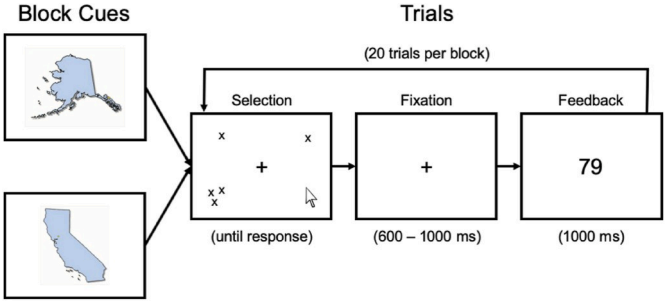

Fig. 1. Task with timing details. Blocks started with a cue indicating the type of reward distribution (single-patch or multi-patch). Participants chose a dig location and were rewarded with an amount of gold from 1 to 100. Previous dig locations were marked on the map and were shown until the participant responded (clicked on the desired dig location).

2.4. Computational models

Several computational models were implemented in MATLAB and evaluated based on how well they accounted for our participants' chosen dig locations. The goal of this modelling was to classify trials as either exploitations or explorations. In general, exploitations were defined as trials for which a participant responded in a value-maximizing way, e.g. choosing the location with the best-known reward. All other responses were explorations. Although several models were tested, only the best-fitting model was used to classify trials for our ERP analysis.

Participant decisions were classified as either exploitations or explorations using computational models. Several models were evaluated for their ability to account for our participants' decisions. First, each model was fit to each participant's data using the MATLAB function fmincon (Optimization Toolbox, Release 2018a, Mathworks, Natick). This function works by searching for parameters that minimize a specified objective function. In our case, the objective function was the negative log-likelihood of a participant's responses, given a particular model. Specifically, each model maintained a probability P_t associated with every possible action a on trial t (i.e., each location on the map). In practice, to reduce the computational complexity of our model-fitting procedure, we further discretized our 800 by 800 pixel maps to an 80 by 80 grid (6400 possible actions).

A good fit meant that the model was assigning high probabilities to a participant's actions (i.e., the chosen map locations). The trial-to-trial probabilities were combined according to the log-likelihood function:

$$-LL = -\sum_{t} \log(P_t(a_s)) \Bigg)$$

where a_s was the selected action. We then computed the mean -*LL* across participants; the best-fitting model, defined as the one the one with the smallest mean –*LL*, was later used to classify trials as exploitations or explorations.

All of our models generated a probability associated with each map location. For all but one of our models (the win-stay, lose-shift model) this was done by first generating a value for each map location. The values v were then converted to action probabilities for each a_i and trial t according to the softmax equation:

$$P_t(a_i) = \frac{e^{v_t(i)_\tau}}{\sum_j e^{v_t(j)_\tau}}$$

where i was the index of the chosen action, j indexed over all possible actions, and τ (temperature) determined the degree of bias towards choosing high-valued locations. Next, we will describe how the values were computed for each function-approximation model.

2.4.1. Nearest-neighbours

This model computed a value for each map location using the nearest-neighbours approach (i.e., the value at a point was equal to the value of the closest previously-chosen point). The values were updated following feedback. In particular, we used MATLAB's griddata function with the "nearest" method. See Fig. 2 for an illustration of how action probabilities were represented after sampling from an example reward distribution.

2.4.2. Inverse euclidean distance

Here, the value associated with each map location was defined as the inverse Euclidean distance from the previously-chosen location (i.e., a bias towards making the same action as before).

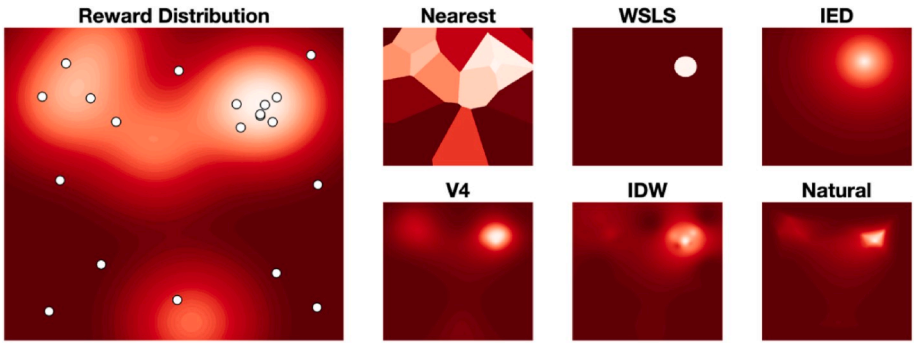

Fig. 2. Sample responses and model representations. The patchiness of the underlying reward distribution was high in this case (left). The participant's responses are shown as white dots. Participant responses were modelled several ways – the final action probabilities are shown on the right (lighter areas were more likely to be chosen, according to the model).

2.4.3. Spline interpolation

This model attempted to estimate the underlying reward distribution using the history of feedback. We again used MATLAB's griddata function, but with the "V4" method.

2.4.4. Inverse distance weighting

This model was similar to the inverse Euclidean distance model; action values were determined based on the inverse distance to each previously-chosen reward, weighted by the values of the previously-chosen rewards.

2.4.5. Natural neighbours

This model estimated the underlying reward distribution using the natural-neighbours method (MATLAB's griddata function with the "nearest" option), which provides a smoother interpolation compared to nearest-neighbours.

2.4.6. Win-stay, lose-shift

As mentioned, we also tested a win-stay, lose-shift model. Generally speaking, these models implement a simple heuristic that tends to repeat an action following a win but switch to a different action following a loss. Our win-stay, lose-shift model assigned a single action probability ε to a radius *r* around the best-chosen option (the win-stay probability), and 1- ε to the rest of the map (Wu et al., 2018). See Fig. 2.

2.5. Data analysis

2.5.1. Modelling data

All of our function-approximation models had a single tunable parameter: the softmax temperature, τ. The win-stay, lose-shift model had two tunable parameters: the win-stay probability, ε, and the affected radius, *r*. The model-fitting procedure described earlier (minimization via MATLAB's fmincon) yielded, for each participant and model, final *–LL* values, and final model parameters.

2.5.1.1. Trial classification. Previously, we classified a trial as an exploitation if the participant's choice matched the model's value-driven choice – i.e., the most likely action, according to the model. All other actions were considered to be explorations. Here, however, the action space was quite large (an 800 by 800 grid), so rather than focus on single actions we expanded our definition of exploitation to include a *range* of likely actions. This was possible because the model-generated action probabilities were continuous. For each participant and block patch type (single/multi) we computed the mean action probability. Trials with greater-than-average action probabilities were defined as exploitations; all other trials were explorations.

2.5.2. Behavioural data

For each participant and environment (single-patch/multi-patch) we computed the mean number of trials of each decision type (exploit/explore). This was also done on a trial-by-trial basis (i.e., for trial 2, 3, … 20). We then computed, for each participant, environment (single-patch/multi-patch), and decision type (exploit/explore) the mean response time, displacement from previous response, and reward.

2.5.3. Electroencephalographic data

EEG data were downsampled to 250 Hz, filtered through a (0.1 Hz–30 Hz pass band) phase shift-free Butterworth filter (60 Hz notch), and re-referenced to the average of the two mastoid channels. Next, ocular artifacts were removed using independent component analysis (ICA). In particular, ICA was used to identify components associated with eye movements. These components were then removed when the data were subsequently reconstructed. Subsequent to this, 1300 ms epochs of EEG data were constructed from 200 ms prior to 1100 ms following feedback onset. All trials were then baseline corrected using a 200 ms pre-feedback window. Finally, trials in which the change in voltage in any channel exceeded 10 μV per sampling point or the change in voltage across the epoch was greater than 100 μV were discarded. On average, we removed 28% of epochs (95% CI [23, 34]).

2.5.3.1. Examination of the grand-grand waveform. To avoid biasing our analysis in favour of a statistically-significant difference between our conditions of interest, we defined the P300 by first examining the "grand-grand" average waveform (Kappenman and Luck, 2016). For each participant, we averaged across all EEG epochs (regardless of condition), then averaged across participants. Next, we identified the time/locations at which the most positive-going deflection occurred (388 ms post-stimulus at electrode P4). To capture the apparent P300 deflection, we then identified the times at which 75% of the maximum voltage was reached (288 to 544 ms post feedback), which formed our analysis window. Although right-lateralized P300s have been observed (Alexander et al., 1996; Amaral et al., 2015; Cacioppo et al., 1996), they are not the norm. We therefore repeated our main P300 analysis at typical P300 locations along the midline (Fz, Cz, Pz: see Supplemental Material).

2.5.3.2. Effect of current/next trial decision. Previously, we analyzed the effect of the upcoming decision type (exploit/explore) on the P300 and reported an exploration-related enhancement (Hassall et al., 2013, 2019). Here, we were interested in the possibility that this neural signal may be affected by both the current and the upcoming trial type. We therefore binned trials based on the current and next trial type: exploit-exploit, exploit-explore, explore-exploit, and explore-explore. Transitions between different trial types were infrequent (e.g., explore-exploit – see trial counts in Fig. 4a) so we chose to combine trials across the different environments (single-patch/multi-patch) to improve the signal-to-noise ratio of the resulting waveforms. We then examined

the effect of current/next trial decision on the P300, defined in the same way as in our main analysis (the mean voltage from 288 to 544 ms post feedback at electrode P4). Four trial groupings (and four waveforms) were created for each participant based the current and next trial type. We then conducted a 2 (current trial type: explore/exploit) by 2 (next trial type: explore/exploit) repeated measures ANOVA. The P300 effect here appeared to be driven more by the current trial type than the next trial type (see Results). For our main analysis, described below, we therefore decided to focus on the current-trial decision instead of the next-trial decision.

2.5.3.3. Feedback-locked P300. Conditional waveforms were created by averaging the feedback-locked EEG for each participant, environment (single-patch/multi-patch), and current-trial decision type (exploit/explore). Finally, a P300 was computed as the mean voltage within our analysis window (288 to 544 ms post feedback) at electrode P4, for each participant, block patch type (single/multi), and decision type (exploit/explore). See Fig. 5 for the resulting waveforms and scalp topographies.

Our behavioural results revealed that exploitation was more rewarding than exploration. The reason why this is relevant to our P300 analysis is that the feedback-locked P300 is known to scale with reward magnitude, e.g. larger for high-magnitude wins compared to low-magnitude wins (Sato et al., 2005; Wu and Zhou, 2009; Yeung and Sanfey, 2004). Thus, the effect of decision type (exploit/explore) is likely confounded here by reward magnitude (low/high) such that our exploration-related P300 is weaker than it otherwise would have been. To investigate the role of reward magnitude in our experiment, we constructed low- and high-reward waveforms for each decision type (exploit/explore) by performing a median split (low/high reward) on all feedback-locked EEG trials. For all four waveforms (low-exploit, high-exploit, low-explore, high-explore) we then computed P300 scores at the same electrode and in the same time window as described above. This was done to confirm the presence of a reward-magnitude effect that diminished (not enhanced) our exploration-related P300.

2.5.3.4. Feedback-locked LPP. Upon examining the feedback-locked waveforms (Fig. 5), we noted that the effect of decision type (exploit/explore) on feedback processing was sustained well beyond the usual P300 time range. The difference appeared to be in the LPP time range (Olofsson et al., 2008; Schupp et al., 2000). To investigate this difference, we averaged our waveforms across environment (single--patch/multi-patch) and constructed a difference wave (explore minus exploit) to define a second analysis window. The grand-grand-average approach was not used here because, unlike our P300 analysis, no peaks were apparent in the conditional waveforms at this later time.[1] As before, we located the time/location of the maximum voltage – of the difference wave, this time – and computed the interval within which 75% of this value was reached. This yielded a later time range, at a more central location: 440–804 ms post feedback at electrode POz.

2.5.4. Inferential statistics

The effect of environment (single-patch/multi-patch) on exploration rate was determined using a paired-samples *t*-test. Cohen's *d* was computed according to:

$$d = \frac{M_{\text{diff}}}{s_{\text{diff}}}$$

where M_{diff} was the difference score mean and s_{diff} was the difference score standard deviation (Cumming, 2014). To determine whether exploration rate changed within a block, a linear model relating exploration rate to trial number (2, 3, … 20) was fit to each participant's data using the MATLAB function polyfit. The effect of trial number on the model slopes was assessed using a single-sample *t*-test (and Cohen's *d* computed by dividing the slope mean by the slope standard deviation). Next, our behavioural scores (mean response time, mean displacement, and mean reward) and ERP scores (P300, late potential) were subjected to a 2 (decision: exploit, explore) by 2 (environment: single-patch, multi-patch) repeated-measures ANOVA. Two different effect-size measures were computed: η_p^2 and η_g^2 (Olejnik and Algina, 2003). To help illustrate how our effects of interest (decision, environment) changed over time, we computed η_p^2 for each on a 200 ms sliding window. Post hoc, we computed observed power using G*Power 3.1 (Faul et al., 2007).

3. Results

3.1. Modelling data

A comparison of the mean *–LL* scores revealed that the natural-neighbours method provided the best fit for our participants' data, regardless of block type (Fig. 3a). This was the model we used to classify trials as exploitations or explorations for our EEG analysis. Although model fit varied across blocks and participants, we chose to focus on a single model for our EEG analysis to have a consistent definition of explore/exploit.

Participants explored slightly more in the multi-patch environment (69.0%, 95% CI [67.1, 70.9]) compared to the single-patch environment (66.2%, 95% CI [64.2, 68.3]), $t(23) = 3.13$, $p = .005$, Cohen's $d = 0.64$, observed power = 0.85. We also noted that participants tended to explore less as they discovered the location of the rewards – the slope of the relationship between exploration rate and trial number was non-zero in both the single-patch environment, $t(23) = -29.27$, $p < .001$, Cohen's $d = -5.97$, observed power = 1.00, and the multi-patch environment, $t(23) = -13.53$, $p < .001$, Cohen's $d = -4.06$, observed power = 1.00. See Fig. 4.

3.2. Behavioural data

3.2.1. Effect of current/next trial decision

After collapsing across task to examine the effect of current/next trial type on trial counts, we observed no effect of current-trial type, $F(1,23) = 0.6$, $p = .5$, $\eta_p^2 = 0.02$, $\eta_g^2 = 0.00$, observed power = 0.10 or next-trial type, $F(1,23) = 3.3$, $p = .08$, $\eta_p^2 = 0.12$, $\eta_g^2 = 0.01$, observed power = 0.41. There was also no current-trial by next-trial interaction, $F(1,23) = 0.8$, $p = .4$, $\eta_p^2 = 0.03$, $\eta_g^2 = 0.00$, observed power = 0.13. We also examined the effect of current/next trial type on reward and found an effect of both current-trial type (larger rewards for exploitations), $F(1,23) = 328$, $p < .001$, $\eta_p^2 = 0.93$, $\eta_g^2 = 0.79$, observed power = 1.00, and next-trial type (larger for exploitations), $F(1,23) = 179$, $p < .001$, $\eta_p^2 = 0.88$, $\eta_g^2 = 0.70$, observed power = 1.00. There was an interaction effect between current-trial type and next-trial type – it appeared to take less of a points difference to switch from exploitation to exploration than it took to switch from exploration to exploitation, $F(1,23) = 109$, $p < .001$, $\eta_p^2 = 0.83$, $\eta_g^2 = 0.34$, observed power = 1.00. See Table 1 for exact values, and Fig. 4.

3.2.2. Response time

Response times were affected by decision type, $F(1,23) = 12.68$, $p = .002$, $\eta_p^2 = 0.36$, $\eta_g^2 = 0.04$, observed power = 0.94. There was no effect of environment, $F(1,23) = 0.66$, $p = .4$, $\eta_p^2 = 0.03$, $\eta_g^2 = 0.00$, observed power = 0.13, and no decision by environment interaction, $F(1,23) = 1.12$, $p = .3$, $\eta_p^2 = 0.05$, $\eta_g^2 = 0.00$, observed power = 0.19. See Table 2

[1] LPP analysis windows are often identified using either previous literature (Stevens et al., 2019) or the difference-wave approach (Brown et al., 2012, 2012; Hajcak et al., 2009).

[2] The interaction between current-trial type and next-trial type was unexpected and a potentially interesting area of future study.

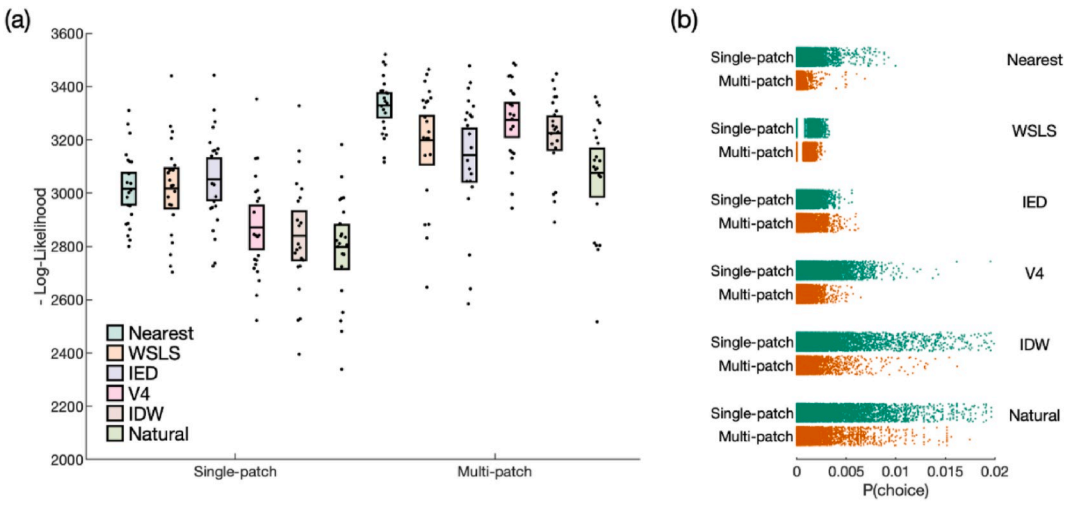

Fig. 3. Model fit results. Most models maintained a value associated with each map location: nearest-neighbours, inverse Euclidean distance: IED, spline interpolation (V4), inverse distance weighting: IDW, and natural neighbours. The win-stay, lose-shift model (WSLS) tended to choose map locations close to the location of greatest previous reward. (a) The models provided comparable fits (lower is better). The natural-neighbours model provided the best mean fit in each environment. (b) Softmax probabilities – these are the model-generated likelihoods for each trial (all participants). Models that yielded better fits tended to generate greater trial-by-trial likelihoods.

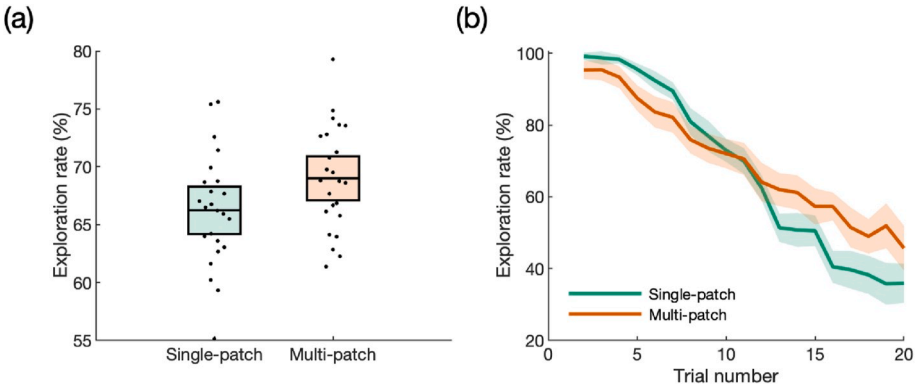

Fig. 4. Trial classification.(a) Overall, participants explored more in the multi-patch environment. (b) The exploration rate decreased throughout a block as participants learned the reward locations. Error bars/shaded regions show 95% confidence intervals.

and Fig. 6 for mean response times.

3.2.3. Displacement

Decision type also affected displacement from previous choice – explorations covered a greater distance, $F(1,23) = 379.13$, $p < .001$, $\eta_p^2 = 0.94$, $\eta_g^2 = 0.82$, observed power = 1.00. There was a smaller effect of environment (greater displacements in the multi-patch environment), $F(1,23) = 6.06$, $p = .02$, $\eta_p^2 = 0.21$, $\eta_g^2 = 0.04$, observed power = 0.68. No interaction was detected, $F(1,23) = 0.16$, $p = .69$, $\eta_p^2 = 0.01$, $\eta_g^2 = 0.00$, observed power = 0.08. See Table 1 and Fig. 6 for mean displacements.

3.2.4. Reward

Exploitations resulted in greater point gains, on average, compared to explorations, $F(1,23) = 1077.00$, $p < .001$, $\eta_p^2 = 0.98$, $\eta_g^2 = 0.93$, observed power = 1.00. The single-patch environment yielded more rewards compared to the multi-patch environment, $F(1,23) = 53.38$, $p < .001$, $\eta_p^2 = 0.70$, $\eta_g^2 = 0.37$, observed power = 1.00. Finally, there was an interaction between decision and environment on reward; the points-advantage of exploiting over exploring appeared to be greatest in the single-patch environment, $F(1,23) = 7.49$, $p = .01$, $\eta_p^2 = 0.25$, $\eta_g^2 = 0.05$, observed power = 0.77. See Table 2 and Fig. 6.

3.3. Electroencephalographic data

3.3.1. Effect of current/next trial decision

There was an effect of current-trial type ($F(1,23) = 7.51$, $p = .01$, $\eta_p^2 = 0.25$, $\eta_g^2 = 0.07$, observed power = 0.76), but not next-trial type ($F(1,23) = 0.00$, $p = .95$, $\eta_p^2 = 0.00$, $\eta_g^2 = 0.00$, observed power = 0.05). There was a current-trial by next-trial interaction, $F(1,23) = 11.58$, $p = .002$, $\eta_p^2 = 0.33$, $\eta_g^2 = 0.13$, observed power = 0.91. Specifically, the effect of exploring on the current trial appeared to be modulated by next-trial decision type (greater when the next trial was an exploitation).[2] See Table 1 and Fig. 4.

3.3.2. P300

There was an effect of decision type on the feedback-locked P300 (enhanced for explorations), $F(1,23) = 25.18$, $p < .001$, $\eta_p^2 = 0.52$, $\eta_g^2 = 0.05$, observed power = 1.00. There was no effect of environment, $F(1,23) = 0.19$, $p = .7$, $\eta_p^2 = 0.01$, $\eta_g^2 = 0.00$, observed power = 0.07, and no interaction, $F(1,23) = 0.80$, $p = .4$, $\eta_p^2 = 0.03$, $\eta_g^2 = 0.00$, observed power = 0.14. See Table 3 for condition means, and Fig. 7.

3.3.3. Effect of reward magnitude

There was an effect of reward magnitude (low/high) on the feedback-locked P300 for exploitations, $t(23) = 3.23$, $p = .004$, Cohen's

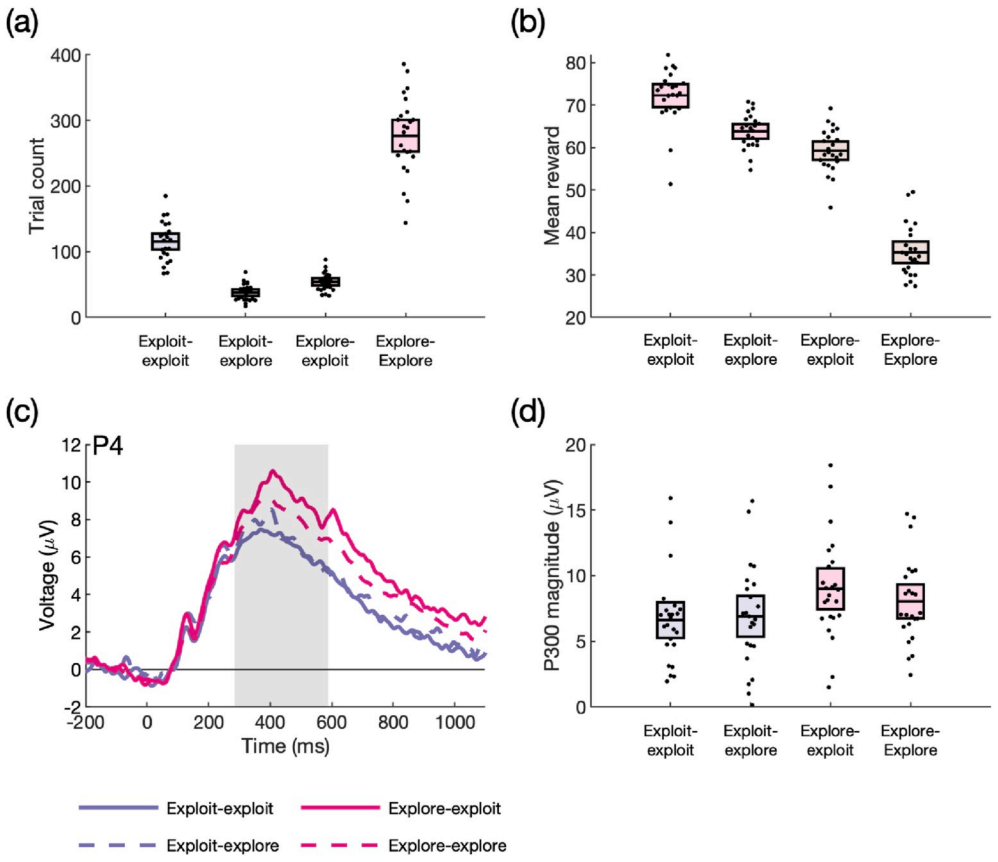

Fig. 5. Behavioural and EEG data by current/next trial type.(a) Participants were more likely to exploit following exploitations, and more likely to explore following explorations. (b) Mean reward by current/next trial. (c) Feedback-locked waveforms for each current/next trial type. The shaded area shows the region of analysis. (d) P300 scores by current/next trial type.

Table 1
Effects of current/next trial decision.

Measure	Exploit (current)				Explore (current)			
	Exploit (next)		Explore (next)		Exploit (next)		Explore (next)	
	M	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
Trial count	115.5	[103.4, 127.6]	37.8	[32.8, 42.9]	54.2	[48.7, 59.7]	276.7	[252.6, 300.8]
Reward (points)	72.2	[69.5, 74.9]	63.8	[62.1, 65.5]	59.3	[57.1, 61.4]	35.3	[32.8, 37.8]
P300 (μV)	6.6	[5.3, 8.0]	6.9	[5.3, 8.5]	9.0	[7.4, 10.6]	8.0	[6.8, 9.3]

Table 2
Behavioural means, with 95% confidence intervals.

	Single-patch				Multi-patch			
	Exploit		Explore		Exploit		Explore	
Measure	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
Response time (ms)	426.4	[377.2, 475.6]	482.4	[429.3, 535.5]	427.0	[378.4, 475.5]	468.4	[408.7, 528.1]
Displacement (mm)	6.9	[5.5, 8.3]	47.8	[43.4, 52.2]	10.3	[8.5, 12.1]	52.2	[45.4, 59.0]
Reward (points)	75.3	[74.2, 76.4]	40.9	[38.4, 43.5]	66.4	[64.4, 68.3]	36.0	[34.0, 38.0]

$d = 0.66$, observed power = 0.87. A similar effect was seen for explorations, $t(23) = 2.45$, $p = .02$, Cohen's $d = 0.50$, observed power = 0.65. Both effects were positive (enhanced for high rewards compared to low rewards). Furthermore, when we compared the high-explore P300 with the low-exploit P300, which had comparable point totals (Table 4), we still observed a large exploration-related P300 enhancement, $t(23) = 7.75$, $p < .001$, Cohen's $d = 1.58$, observed power = 1.00. See Table 4 for mean point amounts in each median split and resulting P300 scores.

3.3.4. LPP

The LPP was affected by decision type, $F(1,23) = 25.28$, $p < .001$, $\eta_p^2 = 0.52$, $\eta_g^2 = 0.08$, observed power = 1.00. There was also a small effect of environment, $F(1,23) = 5.37$, $p = .03$, $\eta_p^2 = 0.19$ $\eta_g^2 = 0.01$, observed power = 0.62. No interaction was detected, $F(1,23) = 0.78$, $p = .4$, $\eta_p^2 = 0.03$, $\eta_g^2 = 0.00$, observed power = 0.14. See Table 3 and Fig. 8.

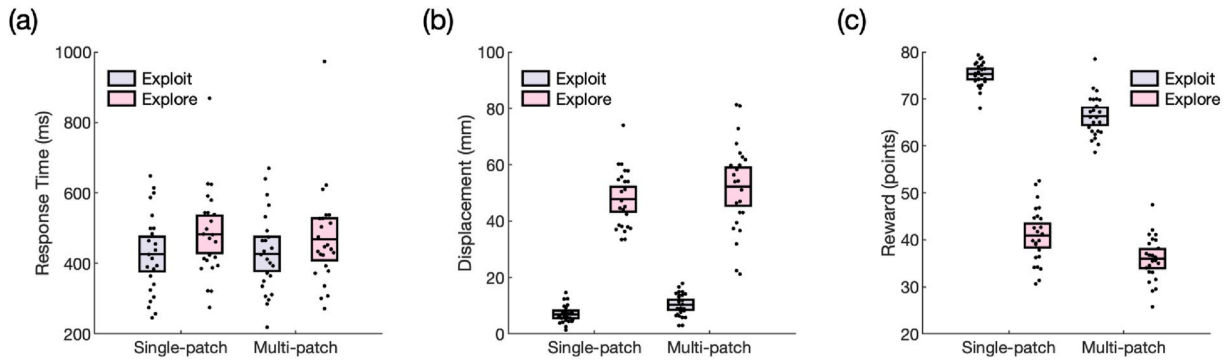

Fig. 6. Behavioural results.Explorations were (a) slower and (b) farther from the previous choice. (c) Exploitation resulted in a greater average point gain.

Table 3
ERP scores, with 95% confidence intervals.

Measure	Single-patch				Multi-patch			
	Exploit		Explore		Exploit		Explore	
	M	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
P300 (μV)	6.8	[5.2, 8.4]	8.2	[6.7, 9.8]	6.8	[5.2, 8.3]	8.5	[7.1, 10.0]
LPP (μV)	3.9	[2.5, 5.3]	5.6	[4.2, 7.1]	4.3	[2.9, 5.7]	6.4	[4.8, 8.0]

not our main goal, the model-fitting procedure itself was important because this was how we classified trials as exploitations or explorations. By choosing the best of several models, we gained confidence in our trial classification. We discovered that a model that approximated the underlying value function provided the best fit for our data (in particular, using the natural-neighbours approach). This discovery is in line with work by Wu et al. (2018) who, after comparing many different models, found evidence that humans rely on function approximation to find spatially correlated rewards in a large decision space (an 11-by-11 grid). We have shown here that this finding holds true in a more continuous space (an 800-by-800 grid).

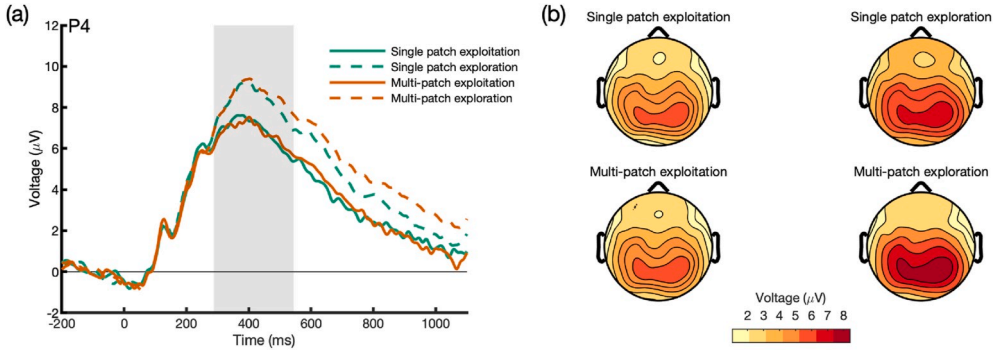

Fig. 7. Feedback-locked P300 waveforms (left) and scalp topographies (right).

Table 4
Reward magnitude and P300 scores, with 95% confidence intervals.

Measure	Explore				Exploit			
	Low Reward		High Reward		Low Reward		High Reward	
	M	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
Reward (points)	17.5	[15.1, 19.9]	58.8	[56.9, 60.1]	57.3	[54.0, 60.1]	82.1	[80.5, 83.8]
P300 (μV)	7.9	[6.4, 9.4]	8.7	[7.3, 10.2]	5.9	[4.4, 7.5]	7.3	[5.7, 8.8]

4. Discussion

In this experiment, we observed an enhanced P300 for feedback following decisions to explore, a result previously observed when exploration was rare compared to exploitation (Hassall et al., 2013, 2019). Here we showed that exploration enhances the feedback-locked P300 even when exploration is frequent. Furthermore, the exploration-related P300 appears to be unaffected by reward distribution knowledge, even though such knowledge affects exploration rate and choice behaviour.

We began by testing how well several models could account for our participants' trial-to-trial decisions. Although model comparison was

After classifying participant decisions as exploitations or explorations, we confirmed two critical features of our experiment. First, and in contrast to earlier work, explorations were more common than exploitations — a feature that allowed us to test whether or not frequent exploration would elicit a P300 enhancement (discussed below). Second, we verified that our between-block manipulation had worked; participants explored more when they were shown a multi-patch cue compared to when they were shown a single-patch cue.[3] In other words,

[3] Though significant, this effect was small, and we noted considerable inter-participant variability in exploration rate (Fig. 4a).

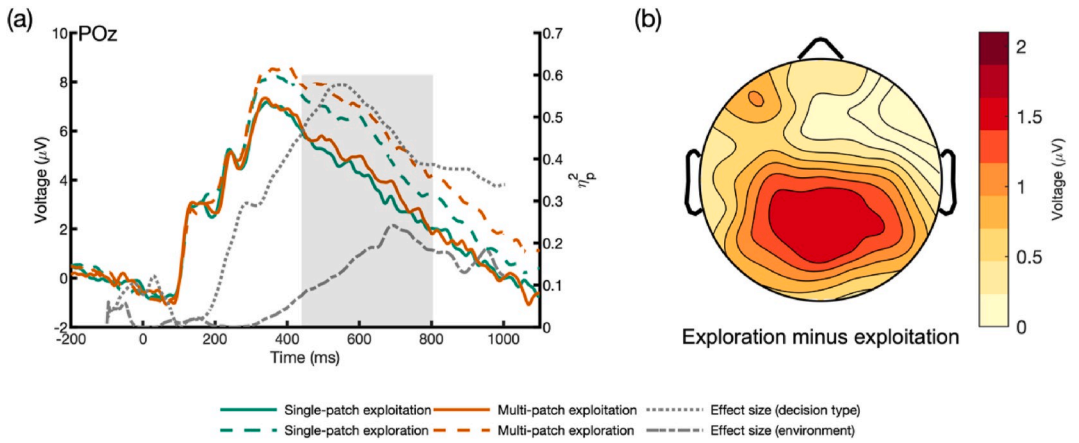

Fig. 8. Feedback-locked LPP (left) and scalp topography of the explore-minus-exploit difference scores (right). The grey lines show the decision/environment effect sizes computed on a 200-ms sliding window.

our participants' decisions to explore were influenced by the reward distribution. This observation is in line with some (Constantino and Daw, 2015), but not all previous work (Hutchinson et al., 2008). Furthermore, explorations were slower compared to exploitations, in line with other studies (Beharelle et al., 2015; Hassall et al., 2013). This is somewhat unsurprising here, given that explorations covered a greater distance than exploitations. Unlike these previous studies, our participants were told the upcoming reward distribution, and adjusted their strategy accordingly. It remains to be seen whether our task would elicit these adaptations if the nature of the reward distribution was initially unknown. Others (Wu et al., 2018) have found that humans assume smooth, spatially-correlated reward distributions (which ours are). But would naïve participants assume the presence of one reward patch or many reward patches?

In line with previous work (Hassall et al., 2013, 2019), our examination of the neural response to feedback revealed an exploration-related P300. In particular, the feedback-locked P300 was greater following exploration than following exploitation. Although a reward-magnitude confound was present – exploitative feedback was more rewarding than exploratory feedback – this confound likely diminished the exploration-related effect since lower-magnitude rewards resulted in a reduced P300 (Sato et al., 2005; Wu and Zhou, 2009; Yeung and Sanfey, 2004). Future studies may be able to control for reward magnitude by making exploration more or less rewarding, e.g. by making the environment non-stationary or stationary.

This presence of an exploration-related P300 here is noteworthy because, unlike previous work examining the exploration-related P300, exploration in our experiment was the more common decision type. Thus, these results rule out the possibility that the exploration-related P300 is driven entirely by the frequency of exploration relative to exploitation. Furthermore, an exploratory analysis identified that the observed exploration-related effect likely peaked around 600 ms post feedback – later than what is usually associated with the P300, but not unheard-of (Polich, 2007). Consistent with previous work, we have labelled this component the LPP (sometimes called the late positive component, or LPC). However, the distinction between the P300 and the LPP may be subtle or event nonexistent (the P300 is thought to be a major contributor to the LPP: Foti et al., 2009; Hajcak and Foti, 2020; MacNamara et al., 2009; Weinberg and Hajcak, 2011). Thus, a simple explanation for our LPP result is that it reflects a delayed P300 effect. One factor affecting P300 latency is cognitive load (Duncan-Johnson, 1981; Krigolson et al., 2012), potentially relevant here because the gold rush task has a much larger action space compared to our previous tasks (Hassall et al., 2013, 2019). The hypothesis that cognitive load shifts the latency of the exploration-related P300 could be tested in future work by, for example, manipulating the size of the action space.

Alternatively, we could consider factors known to affect the amplitude of the LPP. The LPP is thought to reflect the engagement of a general motivational system in the brain – general because it is sensitive not only to emotional images, but also to task-relevant features (Bradley, 2009). For example, the LPP is enhanced if participants are asked to count emotional but not neutral images (Ferrari et al., 2008). Relevant here, the LPP is still present after repeated viewings of the same stimulus (Codispoti et al., 2007). In decision-making contexts, an enhanced LPP is seen for gains versus losses (Broyd et al., 2012) and for larger rewards (Meadows et al., 2016). We can probably rule out a "reward magnitude" explanation since exploratory feedback was less rewarding than exploitative feedback in our experiment. A "motivation" explanation is possible, provided that exploratory feedback in this task had more motivational significance than exploitative feedback (a significance that was further enhanced in the multi-patch environment). Although we did not manipulate or test for level of motivation here, this approach might prove promising in the future as previous research has shown a link between overall task involvement and the feedback-locked P300 (Yeung et al., 2005).

We suggested previously that the exploration-related P300 may be linked to a neural interrupt signal (Hassall et al., 2013, 2019). We speculated that such a mechanism would be useful in suppressing a default strategy (e.g., exploration) in favour of trying something new (e.g., exploitation). In those studies, we interpreted "default" as "more frequent", an interpretation that does not apply to the current results. Here, participants exploited less often than they explored, yet exploration still yielded the greater P300. To maintain a neural-interrupt explanation for our results, we would no longer define as default whichever decision type (exploit/explore) was more frequent. Instead, we would conclude that exploration may be the default strategy generally (i.e., regardless of task). This is a difficult claim to test because it is not clear what exploration and exploitation mean outside of a laboratory or task context. However, it has been suggested that mind-wandering may be a form of exploration, and goal-directed thinking a form of exploitation (Sripada, 2018). Interestingly, mind-wandering is thought to be related to the brain's default state (the default mode network, or DMN: Raichle, 2015). Thus, there are theoretical reasons to suspect that switching from exploration to exploitation always requires neural interruption, regardless of rate of exploration.[4]

Here we have shown that exploration in continuous environments is

[4] Estimates of mind-wandering rates vary from a third (Kane et al., 2007) to half of our daily lives (Killingsworth and Gilbert, 2010). It is therefore unclear whether mind-wandering or goal-directed thought is the more common mental state.

followed by enhanced feedback processing, even when exploration is the dominant strategy. We suggest that this effect is driven mainly by the neural processes required to switch from exploration to exploitation (a neural interrupt signal). These neural processes are general; they operate across different task types (discrete and continuous) and exploration rates (rare and common).

CRediT authorship contribution statement

Cameron D. Hassall: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Olave E. Krigolson:** Conceptualization, Methodology, Writing - review & editing, Supervision.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2020.107538.

References

Alexander, J.E., Bauer, L.O., Kuperman, S., Morzorati, S., O'Connor, S.J., Rohrbaugh, J., Porjesz, B., Begleiter, H., Polich, J., 1996. Hemispheric differences for P300 amplitude from an auditory oddball task. Int. J. Psychophysiol. 21 (2), 189–196. https://doi.org/10.1016/0167-8760(95)00047-X.

Amaral, C.P., Simões, M.A., Castelo-Branco, M.S., 2015. Neural signals evoked by stimuli of increasing social scene complexity are detectable at the single-trial level and right lateralized. PloS One 10 (3), e0121970. https://doi.org/10.1371/journal.pone.0121970.

Beharelle, A.R., Polanía, R., Hare, T.A., Ruff, C.C., 2015. Transcranial stimulation over frontopolar cortex elucidates the choice attributes and neural mechanisms used to resolve exploration–exploitation trade-offs. J. Neurosci. 35 (43), 14544–14556. https://doi.org/10.1523/JNEUROSCI.2322-15.2015.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., Rushworth, M.F.S., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10 (9), 1214–1221 https://doi.org/10/ddsv2g.

Bourdaud, N., Chavarriaga, R., Galan, F., Millan, J. d R., 2008. Characterizing the EEG correlates of exploratory behavior. IEEE Trans. Neural Syst. Rehabil. Eng. 16 (6), 549–556. https://doi.org/10.1109/TNSRE.2008.926712.

Bradley, M.M., 2009. Natural selective attention: orienting and emotion. Psychophysiology 46 (1), 1–11 https://doi.org/10/fpbzmw.

Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vis. 10, 433–436.

Brown, S.B.R.E., van Steenbergen, H., Band, G.P.H., de Rover, M., Nieuwenhuis, S., 2012. Functional significance of the emotion-related late positive potential. Front. Hum. Neurosci. 6 https://doi.org/10/gfwr9f.

Broyd, S.J., Richards, H.J., Helps, S.K., Chronaki, G., Bamford, S., Sonuga-Barke, E.J.S., 2012. An electrophysiological monetary incentive delay (e-MID) task: a way to decompose the different components of neural response to positive and negative monetary reinforcement. J. Neurosci. Methods 209 (1), 40–49 https://doi.org/10/f36dn6.

Cacioppo, J.T., Crites, S.L., Gardner, W.L., 1996. Attitudes to the right: evaluative processing is associated with lateralized late positive event-related brain potentials. Pers. Soc. Psychol. Bull. 22 (12), 1205–1219. https://doi.org/10.1177/01461672962212002.

Chase, J.M., Wilson, W.G., Richards, S.A., 2001. Foraging trade-offs and resource patchiness: theory and experiments with a freshwater snail community. Ecol. Lett. 4 (4), 304–312 https://doi.org/10/c4wz6j.

Codispoti, M., Ferrari, V., Bradley, M.M., 2007. Repetition and event-related potentials: distinguishing early and late processes in affective picture perception. J. Cognit. Neurosci. 19 (4), 577–586 https://doi.org/10/dgnmzn.

Constantino, S.M., Daw, N.D., 2015. Learning the opportunity cost of time in a patch-foraging task. Cognit. Affect Behav. Neurosci. 15 (4), 837–853. https://doi.org/10.3758/S13415-015-0350-Y.

Cumming, G., 2014. The new statistics: why and how. Psychol. Sci. 25 (1), 7–29. https://doi.org/10.1177/0956797613504966.

Duncan-Johnson, C.C., 1981. Young psychophysiologist award address, 1980. Psychophysiology 18 (3), 207–215. https://doi.org/10.1111/j.1469-8986.1981.tb03020.x.

Duncan-Johnson, C.C., Donchin, E., 1977. On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology 14 (5), 456–467 https://doi.org/10/c34pf5.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191. https://doi.org/10.3758/BF03193146.

Ferrari, V., Codispoti, M., Cardinale, R., Bradley, M.M., 2008. Directed and motivated attention during processing of natural scenes. J. Cognit. Neurosci. 20 (10), 1753–1761 https://doi.org/10/bqcqfw.

Foti, D., Hajcak, G., Dien, J., 2009. Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. Psychophysiology 46 (3), 521–530. https://doi.org/10.1111/j.1469-8986.2009.00796.x.

Hajcak, G., Dunning, J.P., Foti, D., 2009. Motivated and controlled attention to emotion: time-course of the late positive potential. Clin. Neurophysiol. 120 (3), 505–510 https://doi.org/10/b6whzw.

Hajcak, G., Foti, D., 2020. Significance?... Significance! Empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: an integrative review. Psychophysiology, e13570. https://doi.org/10.1111/psyp.13570.

Hassall, C.D., Holland, K., Krigolson, O.E., 2013. What do I do now? An electroencephalographic investigation of the explore/exploit dilemma. Neuroscience 228, 361–370 https://doi.org/10/f4j5r4.

Hassall, C.D., McDonald, C.G., Krigolson, O.E., 2019. Ready, set, explore! Event-related potentials reveal the time-course of exploratory decisions. Brain Res. 1719, 183–193 https://doi.org/10/gf3d7b.

Hutchinson, J.M.C., Wilke, A., Todd, P.M., 2008. Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? Anim. Behav. 75 (4), 1331–1349 https://doi.org/10/fmk7px.

Kane, M.J., Brown, L.H., McVay, J.C., Silvia, P.J., Myin-Germeys, I., Kwapil, T.R., 2007. For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. Psychol. Sci. 18 (7), 614–621 https://doi.org/10/cwt295.

Kappenman, E.S., Luck, S.J., 2016. Best practices for event-related potential research in clinical populations. Biol. Psychiatr.: Cognitive Neuroscience and Neuroimaging 1 (2), 110–115 https://doi.org/10/gfz986.

Killingsworth, M.A., Gilbert, D.T., 2010. A wandering mind is an unhappy mind. Science 330 (6006), 932–932.https://doi.org/10/fw6xg5.

Krigolson, O.E., Heinekey, H., Kent, C.M., Handy, T.C., 2012. Cognitive load impacts error evaluation within medial-frontal cortex. Brain Res. 1430, 62–67. https://doi.org/10.1016/j.brainres.2011.10.028.

Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L., Strong, D.R., Brown, R.A., 2002. Evaluation of a behavioral measure of risk taking: the balloon Analogue risk task (BART). J. Exp. Psychol. Appl. 8 (2), 75–84. https://doi.org/10.1037//1076-898X.8.2.75.

Lenow, J.K., Constantino, S.M., Daw, N.D., Phelps, E.A., 2017. Chronic and acute stress promote overexploitation in serial decision making. J. Neurosci. 37 (23), 5681–5689 https://doi.org/10/gbhw5g.

MacNamara, A., Foti, D., Hajcak, G., 2009. Tell me about it: neural activity elicited by emotional pictures and preceding descriptions. Emotion 9 (4), 531–543. https://doi.org/10.1037/a0016251.

Meadows, C.C., Gable, P.A., Lohse, K.R., Miller, M.W., 2016. The effects of reward magnitude on reward processing: an averaged and single trial event-related potential study. Biol. Psychol. 118, 154–160 https://doi.org/10/f83987.

Nieuwenhuis, S., Aston-Jones, G., Cohen, J.D., 2005. Decision making, the P3, and the locus coeruleus—norepinephrine system. Psychol. Bull. 131 (4), 510–532 https://doi.org/10/b3mh34.

Olejnik, S., Algina, J., 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychol. Methods 8 (4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434.

Olofsson, J.K., Nordin, S., Sequeira, H., Polich, J., 2008. Affective picture processing: an integrative review of ERP findings. Biol. Psychol. 77 (3), 247–265 https://doi.org/10/bbdfp6.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vis. 10 (4), 437–442. https://doi.org/10.1163/156856897X00366.

Pleskac, T.J., Wershbale, A., 2014. Making assessments while taking repeated risks: a pattern of multiple response pathways. J. Exp. Psychol. Gen. 143 (1), 142–162 https://doi.org/10/gf3xp4.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Raichle, M.E., 2015. The brain's default mode network. Annu. Rev. Neurosci. 38 (1), 433–447 https://doi.org/10/gdqcqz.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., Kuboki, T., 2005. Effects of value and reward magnitude on feedback negativity and P300. Neuroreport 16 (4), 407–411 https://doi.org/10/b3nrr5.

Schupp, H.T., Cuthbert, B.N., Bradley, M.M., Cacioppo, J.T., Ito, T., Lang, P.J., 2000. Affective picture processing: the late positive potential is modulated by motivational relevance. Psychophysiology 37 (2), 257–261 https://doi.org/10/cf9hfx.

Sripada, C.S., 2018. An exploration/exploitation trade-off between mind wandering and goal-directed thinking. In: Fox, K., Christoff, K. (Eds.), The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming. Oxford University Press, pp. 23–34 https://doi.org/10/gf3sjr.

Stevens, E.M., Frank, D., Codispoti, M., Kypriotakis, G., Cinciripini, P.M., Claiborne, K., Deweese, M.M., Engelmann, J.M., Green, C.E., Karam-Hage, M., Minnix, J.A., Ng, J., Robinson, J.D., Tyndale, R.F., Vidrine, D.J., Versace, F., 2019. The late positive potentials evoked by cigarette-related and emotional images show no gender differences in smokers. Sci. Rep. 9 (1), 3240 https://doi.org/10/gf2zhx.

Sun, S., Zhen, S., Fu, Z., Wu, D.-A., Shimojo, S., Adolphs, R., Yu, R., Wang, S., 2017. Decision ambiguity is mediated by a late positive potential originating from cingulate cortex. Neuroimage 157, 400–414 https://doi.org/10/gbxhnk.

Tzovara, A., Murray, M.M., Bourdaud, N., Chavarriaga, R., Millán, J. del R., De Lucia, M., 2012. The timing of exploratory decision-making revealed by single-trial topographic EEG analyses. Neuroimage 60 (4), 1959–1969 https://doi.org/10/f3w3s6.

Weinberg, A., Hajcak, G., 2011. The late positive potential predicts subsequent interference with target processing. J. Cognit. Neurosci. 23 (10), 2994–3007 https://doi.org/10/dqz4js.

Wilson, R.C., Geana, A., White, J.M., Ludvig, E.A., Cohen, J.D., 2014. Humans use directed and random exploration to solve the explore–exploit dilemma. J. Exp. Psychol. Gen. 143 (6), 2074–2081 https://doi.org/10/f6tr8t.

Wu, C.M., Schulz, E., Speekenbrink, M., Nelson, J.D., Meder, B., 2018. Generalization guides human exploration in vast decision spaces. Nature Human Behaviour 2 (12), 915–924 https://doi.org/10/gfjwtq.

Wu, Y., Zhou, X., 2009. The P300 and reward valence, magnitude, and expectancy in outcome evaluation. Brain Res. 1286, 114–122 https://doi.org/10/fxczvj.

Yeung, N., Holroyd, C.B., Cohen, J.D., 2005. ERP correlates of feedback and reward processing in the presence and absence of response choice. Cerebr. Cortex 15 (5), 535–544. https://doi.org/10.1093/cercor/bhh153.

Yeung, N., Sanfey, A.G., 2004. Independent coding of reward magnitude and valence in the human brain. J. Neurosci. 24 (28), 6258–6264 https://doi.org/10/dbn7qt.

Zheng, Y., Li, Q., Wang, K., Wu, H., Liu, X., 2015. Contextual valence modulates the neural dynamics of risk processing. Psychophysiology 52 (7), 895–904 https://doi.org/10/f7gf3k.

Zheng, Y., Liu, X., 2015. Blunted neural responses to monetary risk in high sensation seekers. Neuropsychologia 71, 173–180 https://doi.org/10/f7dhp9.