NeuroImage 186 (2019) 137–145

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Oscillatory signatures of reward prediction errors in declarative learning

Kate Ergo [*,1], Esther De Loof [1], Clio Janssens, Tom Verguts

Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000, Ghent, Belgium

ARTICLE INFO

Keywords:
Declarative learning
Signed reward prediction error
EEG
Oscillatory signatures of reward processing

ABSTRACT

Reward prediction errors (RPEs) are crucial to learning. Whereas these mismatches between reward expectation and reward outcome are known to drive procedural learning, their role in declarative learning remains underexplored. Earlier work from our lab addressed this, and consistently found that signed reward prediction errors (SRPEs; "better-than-expected" signals) boost declarative learning. In the current EEG study, we sought to explore the neural signatures of SRPEs. Participants studied 60 Dutch-Swahili word pairs while RPE magnitudes were parametrically manipulated. Behaviorally, we replicated our previous findings that SRPEs drive declarative learning, with increased recognition for word pairs accompanied by large, positive RPEs. In the EEG data, at the start of reward feedback processing, we found an oscillatory (theta) signature consistent with unsigned reward prediction errors (URPEs; "different-than-expected" signals). Slightly later during reward feedback processing, we observed oscillatory (high-beta and high-alpha) signatures for SRPEs during reward feedback, similar to SRPE signatures during procedural learning. These findings illuminate the time course of neural oscillations in processing reward during declarative learning, providing important constraints for future theoretical work.

1. Introduction

While interacting with the world, we continuously build up predictions. One important variable we make predictions about, is reward. Whenever a difference occurs between the predicted reward and the obtained reward, a reward prediction error (RPE) is elicited. It is a widely accepted notion that RPEs are fundamental to learning and the formation of new memories. This was already formalized in the seminal Rescorla and Wagner (1972) model, but has generated considerable interest in the recent predictive coding (Mathys et al., 2011; Rao and Ballard, 1999) and Reinforcement Learning (RL) (Silvetti et al., 2014; Sutton and Barto, 1998) frameworks. In these models, learning primarily takes place when a reward or an outcome is unexpected.

One factor in the success of RPEs is the strong support from biological data on the reward circuitry. In particular, dopaminergic (DA) neurons in the midbrain were found to encode RPEs (Schultz et al., 1997), in a manner very similar to the temporal-difference algorithm of RL (Montague et al., 1996; Sutton and Barto, 1981). Furthermore, experimentally manipulating the RPE disrupts learning (Steinberg et al., 2013). Recent work has more specifically verified the subtractive form (observed – predictive signal) of the RPE signal (Eshel et al., 2015; Eshel et al., 2016). Computationally, models that learn by means of RPEs have tackled among the hardest cognitive tasks successfully. For example, in a study by Silver et al. (2016) RPE-based neural networks were trained to play the game of Go and were able to defeat the world champion multiple times.

The role of RPEs has been extensively studied in procedural learning, where learning happens gradually by means of repeated practice (e.g., Ohlsson, 1996; Steinberg et al., 2013). In everyday life, information is often of a declarative nature and must be learned via a single exposure. While many studies provided evidence for the role of reward anticipation and reward delivery in declarative learning (Adcock et al., 2006; Wittmann et al., 2005), there is a substantial void linking RPEs with declarative learning.

Recent studies have started to investigate the effect of RPEs on declarative learning. One valuable approach is to present a set of stimulus-response outcomes repeatedly and fit a learning (e.g., Rescorla-Wagner or Kalman filter) model to the behavioral performance data, which then generates prediction errors (Howard-Jones et al., 2011), possibly for stimuli that are unrelated to the outcome of the learning task (Davidow et al., 2016). For example, in the study by Davidow et al. (2016), stimuli that elicit prediction errors (PEs) are unrelated to the stimuli that have to be remembered. PEs can thus be elicited for stimuli independent of what has to be learned, while still influencing the

* Corresponding author. Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000, Ghent, Belgium.
E-mail address: kate.ergo@ugent.be (K. Ergo).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2018.10.083
Received 15 June 2018; Received in revised form 28 September 2018; Accepted 31 October 2018
Available online 2 November 2018

encoding of these stimuli in memory. However, this approach depends on the fit of the learning model, and generally has yielded mixed results. Although a computational model can make exact predictions about what would happen in different cells of one's design, this requires one to first estimate the parameters of the model before such predictions can be spelled out. This approach can have some downsides; in particular, the predictions may be driven by spurious characteristics of the model, perhaps due to a specific (optimally fitted) parameter setting. Moreover, this approach requires repeated presentation of the stimuli that need to be learned.

An alternative approach is to experimentally manipulate the RPE. We recently applied this approach in a Dutch-Swahili word learning paradigm, and found that signed reward prediction errors (SRPEs; "better-than-expected" signals) drive declarative learning (De Loof et al., 2018). More specifically, recognition improved linearly with larger and positive RPEs. However, given that our previous study only reported behavioral data, it remains unclear whether, as the theory predicts, the SRPE effect can be localized in the reward feedback phase. Given the very high temporal resolution of EEG, this method is ideally suited for this purpose. Moreover, given that earlier studies on procedural learning observed RPE effects in specific frequency bands during reward feedback (HajiHosseini and Holroyd, 2015; HajiHosseini et al., 2012), it was of interest whether we could observe similar oscillatory signatures in our declarative learning paradigm. We thus performed an EEG study to test whether SRPEs modulate EEG oscillations during reward feedback in the acquisition task, reflecting the computation of a "better-than-expected" signal. Additionally, our approach allowed us to detect a neural signature for unsigned reward prediction errors (URPEs; "different-than-expected" signals). We anticipated to see these neural signatures (SRPE and URPE) reflected in an oscillatory power modulation in the theta (4–8 Hz; Cohen et al., 2007), alpha (8–12 Hz; Oya et al., 2005) or high-beta frequency band (20–30 Hz; HajiHosseini and Holroyd, 2015).

2. Materials & methods

2.1. Sample size

Based on previous research, we opted to use a comparable number of participants (40) and a comparable number of word pairs to memorize (60 word pairs). This sample size resulted in a stable pattern of results in both an immediate and delayed test group in the previous and in this study (see below), attesting to the reliability of our results with the current sample size.

2.2. Participants

Forty-one healthy, Dutch speaking, right-handed participants participated in the study after signing an informed consent form. No participant had prior knowledge of Swahili. Twenty participants (15 females, 25.6 years on average) were randomly assigned to the immediate recognition test group, while the remaining twenty-one participants (18 females, 25.5 years on average) were assigned to the delayed test group and performed the recognition test after a one-day delay. Participants were told they would earn at least €20, but possibly up to €25 depending on their performance. To assure that participants were highly motivated, we additionally rewarded two gift vouchers of €20 to participants with the best performance in the immediate and delayed recognition test, respectively.

2.3. Material

A total of 300 words (60 Dutch and 240 Swahili words) (see Table A in Appendix) were used. The experiment was run on a Dell Optiplex 9010 mini-tower running E-Prime software (Schneider et al., 2012). Responses were registered using the Cedrus RB-730 response box enhanced with four time-accurate push buttons (Cedrus Corporation, San Pedro, California).

2.4. Procedure

2.4.1. Familiarization task

To familiarize participants with the stimuli, a familiarization task was included at the start of the experiment. Familiarization was done to control for the novelty of the foreign Swahili words and to level out the effect that some Swahili words might look more familiar than others. Dutch ($n = 60$) and Swahili ($n = 240$) words were randomly presented for a duration of 2 s. Participants were instructed to read the words in silence and press a response button only when a Dutch word was presented.

2.4.2. Acquisition task

In the acquisition task, the aim for the participants was to learn the Swahili translation of 60 Dutch words. On each trial, one Dutch word appeared on top of the screen together with four Swahili words below, of which only one was the correct translation (Fig. 1a). After 4 s, frames surrounded either one, two or four Swahili options. These frames indicated out of which Swahili words participants were allowed to choose to select their guess for the translation of the Dutch word. In the one-option condition, only one Swahili translation was framed and participants could choose with certainty. In the two-option condition, two Swahili translations were framed and participants had a 50% probability of choosing the correct option (and thus of obtaining reward). This probability reduced to 25% when they were presented with the four-option condition in which all four Swahili translations were framed. Each Swahili option was assigned one key and participants had to respond with the middle and index finger of the left and right hand, respectively. There was no time limit to respond. After choosing their answer, a reward anticipation phase of 3 s was inserted to exclude motor activity contamination in the EEG signal. This was followed by reward and performance feedback for 3 s. Next, the to-be-learned Dutch-Swahili word pair was presented for a duration of 5 s. The Dutch word, an equation sign, and its Swahili translation appeared on the screen. Participants were encouraged to use this time to encode the word pair as they knew their memory for the word pairs would be tested in a recognition test, either on the same day or after a one-day delay. If the chosen Swahili translation was rewarded, a green frame was presented around the Dutch word and the chosen Swahili word. Alternatively, if the chosen Swahili translation was unrewarded, a red frame appeared around the Dutch word and one of the other possible Swahili word options. Each trial ended with a reward update for 2.5 s, indicating how much money they earned up until the last completed trial. Participants won €0.70 on rewarded trials, while no money was added on unrewarded trials. In Fig. 1a, the two-option condition with unrewarded choice is illustrated.

2.4.3. Filler task

To avoid recency effects, participants performed a magnitude comparison task immediately after the acquisition task. 400 digits ranging from 1 to 9, with the exclusion of 5, were sequentially presented onto the screen. Participants pressed the left response button for digits smaller than 5 and the right response button for digits larger than 5. All participants executed this filler task, independent of whether they performed the recognition test immediately or after a one-day delay.

2.4.4. Recognition task

In the recognition task, the 60 Dutch-Swahili word pairs were presented. On each trial, the Dutch word appeared on top of the screen together with the same four Swahili translations from the acquisition task (Fig. 1a). The position of the Swahili options was randomly chosen to avoid that participants would learn the translation based on spatial position. No frames surrounded the Swahili translations during the recognition task. No time constraint was imposed. Participants made their choice by pressing one of the four designated response buttons. After indicating their choice, they were asked how certain they were of their

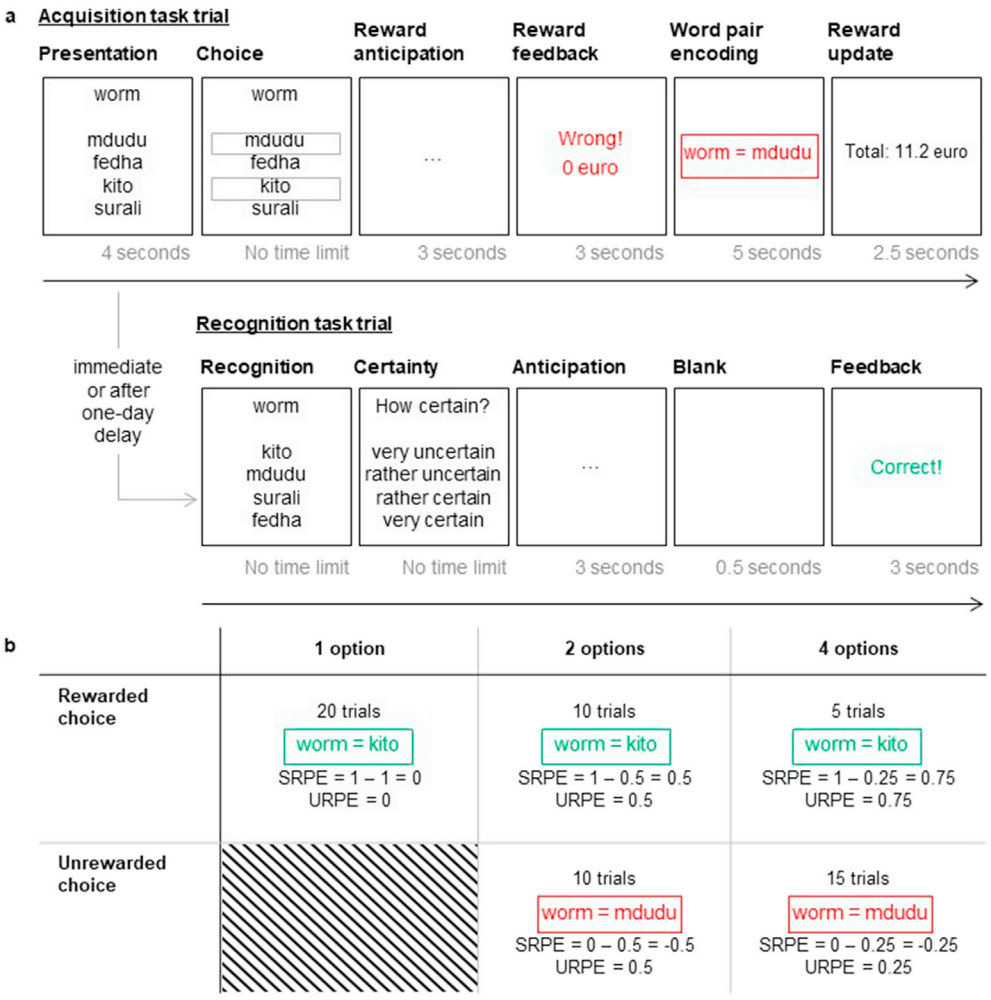

Fig. 1. Overview of the experiment (a) and experimental design (b). (a) In the acquisition task, participants chose between one, two or four Swahili translations. An acquisition trial is illustrated assuming that this is a trial from the "2 options, unrewarded choice" cell in the experimental design, and the participant chose 'kito' as the translation for 'worm'. Hence the word pair to encode is worm = mdudu. If the participant would alternatively have chosen 'mdudu', the feedback would again (by design) have been negative, but in this case the word pair to encode would have been worm = kito. After the 60 acquisition trials, the recognition test was performed either immediately or after a one-day delay. (b) The 2 (obtained reward) × 3 (number of options) experimental design, including number of trials and associated signed and unsigned RPE (SRPE and URPE). SRPEs were calculated by subtracting probability of reward from obtained reward; URPE is the absolute value of SRPE.

answer: 'very uncertain', 'rather uncertain', 'rather certain' or 'very certain' (measured on a scale from 1 'very uncertain' to 4 'very certain'). At the end of each trial, they were given feedback about whether they chose the correct Swahili translation.

2.5. Design

RPE magnitudes were parametrically manipulated by a priori determining the number of options (one, two or four options) as well as the accuracy (correct or incorrect; which here corresponds also to rewarded vs. unrewarded) of each trial. By doing so, we assured that there was a fixed number of trials in each cell of the design (Fig. 1b). Moreover, this allowed us to compute an RPE for each cell of the design as described in more detail below. Note that the predetermined reward feedback implies that participants did not necessarily learn the actual Swahili translations of the Dutch words. For example, if a trial belonged to the rewarded condition, subjects would receive positive feedback irrespective of their choice. Moreover, a random sample of four Swahili words was presented with each Dutch word, often not including the actual Swahili translation of the Dutch word. Participants did not know this during the experiment and thus thought the feedback they received was genuinely about their own choice. They were informed of this manipulation at the end of the experiment.

SRPEs were computed by subtracting the reward probability from the actual reward. In the case of a rewarded trial, the actual reward is one (Fig. 1b, expressed in arbitrary units; see paragraph Acquisition task for the reward scheme). The probability of the reward depends on the number of options. For a one-option condition trial, this probability is equal to one. For a two-option condition trial, the reward probability equals 0.50, whereas the probability reduces to 0.25 for four-option condition trials. The URPE was obtained by taking the absolute value of the SRPE. The URPE and SRPE account predict different outcomes. Recognition should decrease with large negative RPEs according to the SRPE account. In contrast, according to the URPE account, recognition should increase both for large negative and large positive RPEs.

2.6. Data analysis

All behavioral data were analyzed using the linear mixed effects framework in R software (R Core Team, 2014). For continuous dependent variables (e.g., certainty ratings in the recognition test) linear mixed effects models were used, while for categorical dependent variables (e.g., recognition accuracy) generalized linear mixed effects models were applied. A random intercept for participant was included in each model, while all predictors were mean-centered. Note that SRPEs were treated as a continuous predictor allowing the inclusion of all 60 trials per participant to estimate its regression coefficient, with the exception of invalid trials (i.e., trials on which a non-framed Swahili translation was chosen during the acquisition task).

2.7. EEG recordings

Prior to the onset of the acquisition task, participants were informed that EEG would be recorded and were asked to assume a comfortable position and to avoid unnecessary movements. Although EEG recordings were also made during the recognition test in the group who performed

the test immediately, the current analysis is restricted to the EEG recordings made during the acquisition task (recorded in all participants).

Electrophysiological data were recorded using a BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands) with 64 Ag—AgCl electrodes arranged in the standard international 10–20 electrode mapping (Jasper, 1958), with a posterior CMS-DRL electrode pair. Two reference electrodes were positioned at the left and right mastoids. Eye movements were registered with a pair of electrodes above and below the left eye and two additional electrodes at the outer canti of both eyes. EEG signals were recorded at a 1024 Hz sampling rate.

The data were preprocessed in MATLAB (MATLAB R2013a, The MathWorks, Inc., Natick, Massachusetts, United States) using an EEGLAB preprocessing pipeline (Delorme and Makeig, 2004). The data were re-referenced offline to the average of the mastoid electrodes and data stretches with excessive noise were removed after visual inspection. Next, eyeblink artifacts were removed through EEGLAB independent component analysis (ICA), applied on the 0.5–30 Hz band-pass Butterworth filtered data. Eleven participants required the interpolation of one electrode and one participant required the interpolation of two electrodes. Electrode P2 was particularly noisy and made up for ten of the thirteen electrode interpolations. The cleaned data were subsequently filtered with a 60 Hz low-pass filter. Finally, the data were epoched time-locked to the onset of the reward feedback in the acquisition task.

The procedure for the time-frequency analysis was based on the code provided in chapter 16 of Cohen (2014). In order to extract the oscillatory power, a Hanning taper was first applied to the epoched EEG data and these tapered data were subsequently subjected to a short-time Fourier transformation. Because of the intended single-trial analysis on the power estimates a single Hanning taper was used (Cohen, 2014; Haegens et al., 2011). The tapering and a fast Fourier transform were performed in a sliding time window of 600 ms which was advanced in steps of 100 ms between −300 and 3000 ms relative to the onset of the reward feedback. A 600 ms time window was used to satisfy both estimation efficiency and temporal resolution. The oscillatory power was extracted in 18 frequency bands, spaced linearly between 1.67 and 30 Hz in steps of 1.67 Hz. This time-frequency analysis procedure was performed separately for each participant, electrode, and trial.

A baseline correction was applied to the power estimates for each participant, electrode and frequency separately, based on the average baseline activity across all 60 trials. Because we were mainly interested in the brain response to the reward feedback, the power estimates from 400 ms to 300 ms before the onset of the reward feedback were used as the baseline. Note that, because of the 600 ms time window, these power estimates are informed by activity in a window ranging from 700 ms to 0 ms prior to feedback onset. It is worth noting that performing this algorithm with a trial-specific baseline rather than an average baseline resulted in nearly identical results.

Finally, the baseline-corrected data underwent a decibel conversion. In analogy to the behavioral analysis, trials were removed when an invalid (i.e., non-framed) Swahili translation was chosen in the acquisition task. Combined with the removal of EEG data during preprocessing this resulted in a removal of on average 1.63% of the power estimates per participant (ranging between 0% and 5%).

So far, a time-frequency analysis was described where oscillatory power estimates were mapped out in epochs of 3000 ms starting at reward feedback onset. The goal of the subsequent clustering analysis on the power estimates was to test whether oscillatory power in several frequency bands reflected the occurrence of SRPEs. As a multiple comparisons correction, a non-parametric clustering procedure was applied (for a more detailed description of the method, see Maris and Oostenveld (2007)). We first computed correlations between SRPEs and the standardized power estimates at each point in channel, frequency and time. The 0.5% most extreme correlations were entered into the clustering analysis. From these, we clustered adjacent neighbors in the channel, frequency and time domains. To calculate our cluster-level statistic, we multiplied the number of items (i.e., (channel, frequency, time) points) in the cluster with the largest correlation statistic of that cluster. A significance threshold of 5% was imposed on the subsequent non-parametric permutation test with 1000 iterations.

To further clarify the correlation between the SRPEs and the (baselined, decibel-transformed and standardized) power estimates in the significant clusters, the average activity in each significant cluster was calculated on a trial-by-trial basis. These average activities were then regressed onto the RPEs by applying a (trial-by-trial) linear mixed effects model with a random intercept across participants and SRPEs (or URPEs) as a mean-centered predictor.

The significant clusters were visualized in a time-frequency plot and a topographical plot (Fig. 3). For the time-frequency plot, the clustering statistics pertaining to each significant cluster were summarized across channels for each time-frequency combination. The resulting statistics were log transformed before plotting. For the topographical plot, a separate plot was made per significant cluster. For each channel, the clustering statistics were summarized across all significant time-frequency combinations. Again, the resulting statistics were log transformed before plotting. Thus, a topographical plot was constructed for each significant cluster, with small white dots indicating the channels represented in the cluster and large white dots showing the ten most significant channels within each cluster.

3. Results

3.1. Behavioral results

Recognition accuracy in the immediate test group ($M = 73.3\%$, $SD = 16.7\%$, range: 37%–100%) was significantly higher than in the delayed test group ($M = 51.5\%$, $SD = 14.2\%$, range: 32%–81%), $\chi^2(1, N = 41) = 18.6$, $p < 0.001$ (Fig. 2).

To distinguish the URPE from the SRPE account we investigated the effect of number of options and reward. Recognition accuracy was significantly increased with an increasing number of options (one-option: $M = 58.8\%$, $SD = 22.8\%$, range = 20%–100%; two-option: $M = 63.4\%$, $SD = 18.2\%$, range = 30%–100%; four-option: $M = 65.7\%$, $SD = 20.4\%$, range = 20%–100%), $\chi^2(1, N = 41) = 24.0$, $p < 0.001$. Furthermore, rewarded choices were remembered more accurately ($M = 64.1\%$, $SD = 19.6\%$, range = 31%–100%) than unrewarded choices ($M = 60.7\%$, $SD = 20.2\%$, range: 28%–100%), $\chi^2(1, N = 41) = 18.9$, $p < 0.001$. There were no interactions with the test delay (all $p > 0.30$). The data did show a trend towards an interaction between the number of options and the obtained reward, $\chi^2(1, N = 41) = 3.44$, $p = 0.064$. Fig. 2 shows that the trend goes against the direction predicted by the URPE account. In particular, recognition increases with number of options for both rewarded and unrewarded trials. This is the pattern predicted by the SRPE account; the URPE account instead predicts an increase for rewarded trials but a decrease for unrewarded trials (see De Loof et al. (2018) for further explanation). Follow-up tests were performed to verify whether the effect of number of options was statistically increasing for both the rewarded and unrewarded trials. These tests revealed that there was a significant positive effect of the number of options in the rewarded trials, $\chi^2(1, N = 41) = 24.3$, $p < 0.001$. More importantly, in the unrewarded trials there was a positive effect of the number of options that trended towards significance, $\chi^2(1, N = 41) = 3.21$, $p = 0.073$. This data pattern provides evidence for the SRPE account and runs against the URPE account. Furthermore, recognition improved linearly with increasing SRPEs, $\chi^2(1, N = 41) = 24.5$, $p < 0.001$. These recognition results again support the SRPE account for declarative learning (De Loof et al., 2018). For this reason, the remainder of the results are analyzed from an SRPE perspective.

For the certainty ratings, there was no significant main effect of SRPEs, $\chi^2(1, N = 41) = 0.023$, $p = 0.88$, but there was a significant interaction between the effect of SRPEs and delay, $\chi^2(1, N = 41) = 7.01$, $p = 0.0081$. Follow-up tests revealed that SRPEs had no significant effect on certainty ratings for the correctly recognized word pairs in the

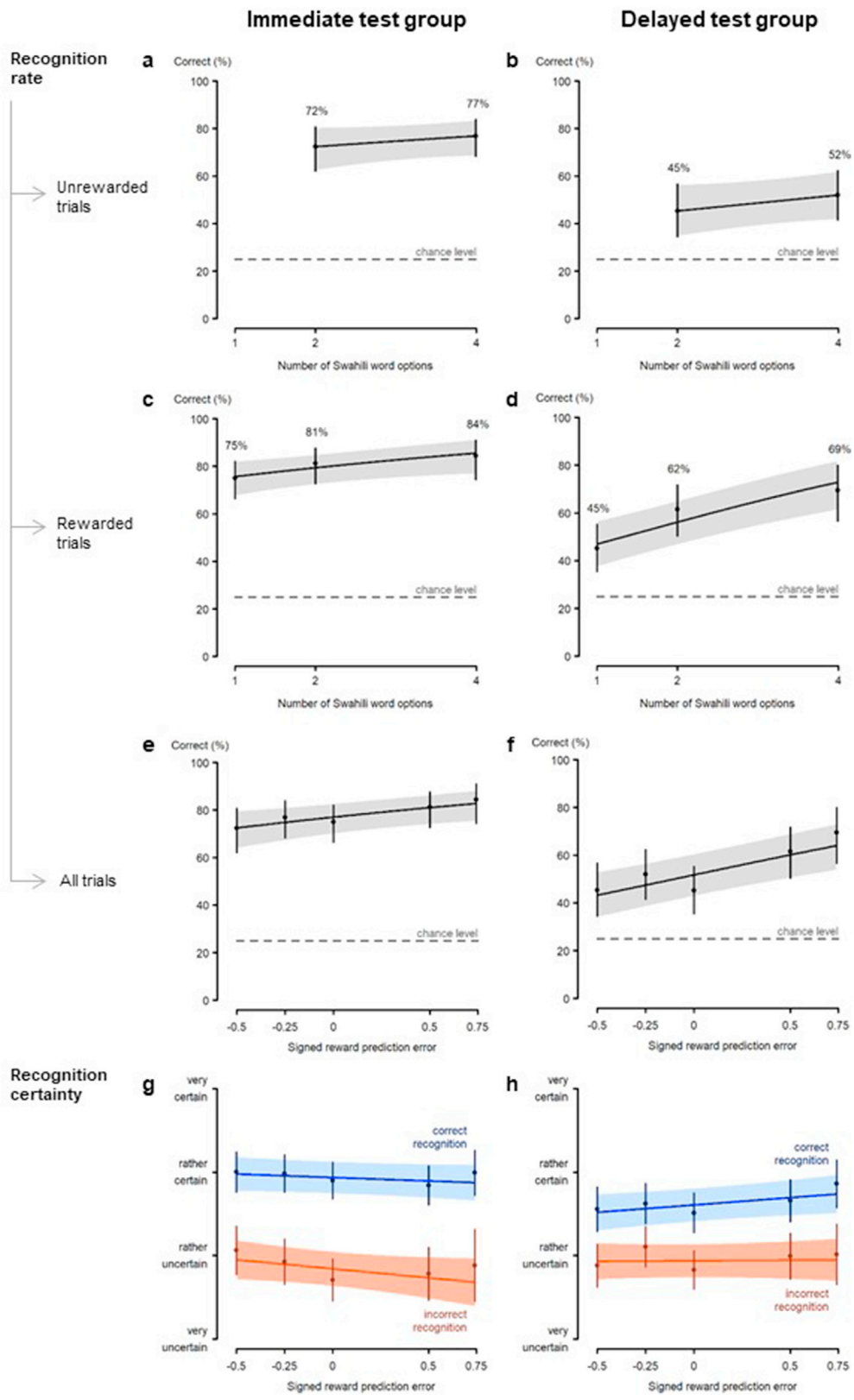

Fig. 2. Recognition accuracy (panel a through f; y-axis) and certainty ratings (panel g and h; y-axis) are plotted as a function of the number of options (panel a through d; x-axis) or as a function of SRPE (panel e through h; x-axis) in the immediate test group (left column) and in the delayed test group (right column). Note that in the one-option condition the chosen translation was always rewarded (panel a through d). For each number of options, reward, and delay condition (as well as for the SRPEs), the average recognition accuracy/certainty and its 95% confidence interval were estimated and superimposed for illustrative purposes only. (a–f) Recognition increased significantly with an increasing number of options and recognition was enhanced for rewarded word pairs; thus recognition increased significantly with higher SRPEs. Performance at chance level is indicated by the gray dashed line at 25% accuracy. (g and h) SRPEs significantly predicted certainty ratings for correctly recognized word pairs (depicted in blue), though only on the delayed test; SRPEs were not predictive for incorrectly recognized word pairs (depicted in orange).

immediate test, $\chi^2(1, N = 20) = 1.15$, $p = 0.28$, but resulted in significantly higher certainty ratings for correct recognitions in the delayed test, $\chi^2(1, N = 21) = 3.90$, $p = 0.048$. Thus, the linear effect of SRPEs on certainty ratings also largely replicates the pattern of findings from our previous study, mainly in the delayed test (De Loof et al., 2018).

3.2. EEG results

For the reward feedback phase, the average power estimates across all midline electrodes are displayed in Fig. 3a. We next investigated the neural signature of SRPEs by examining the oscillatory power

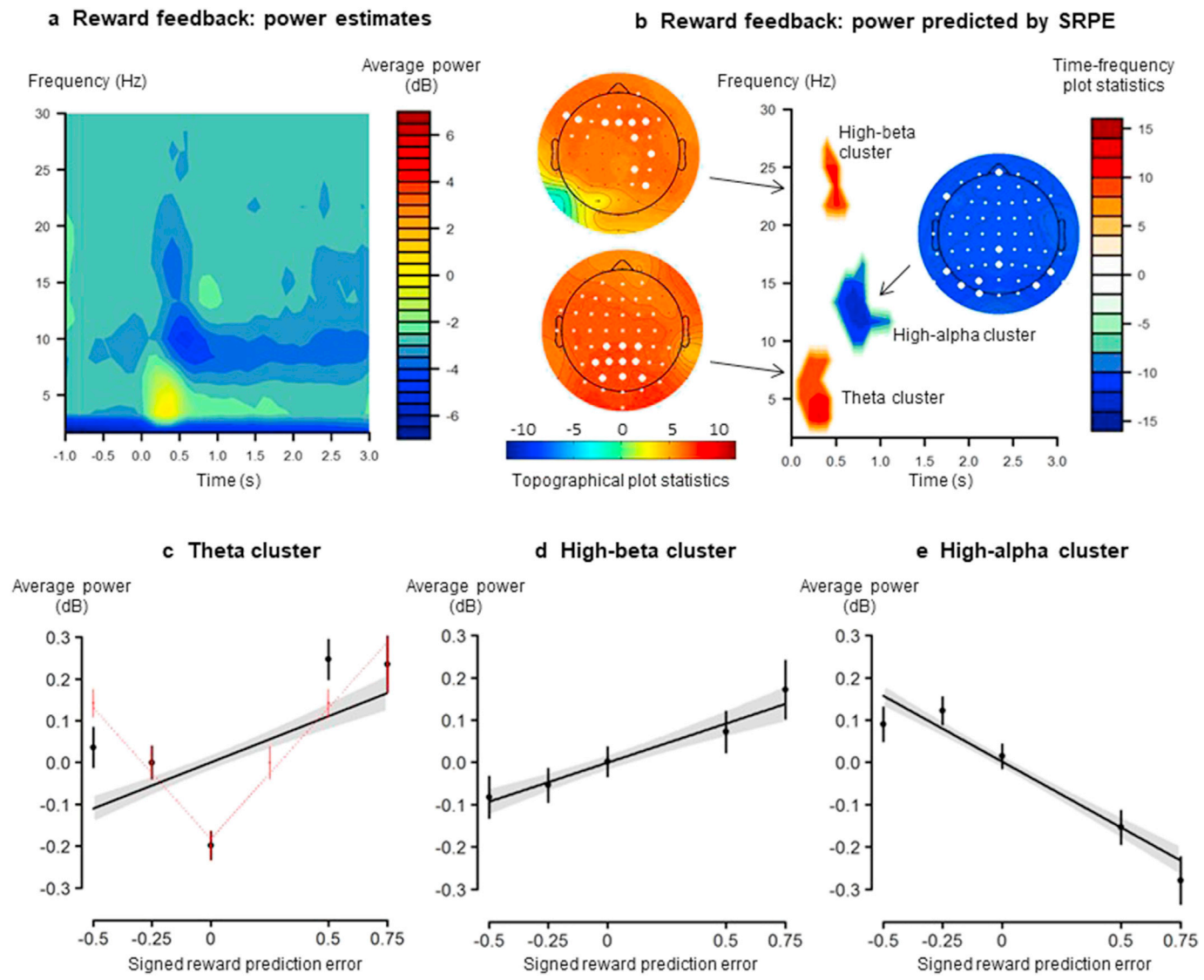

Fig. 3. (a) Overall EEG data: the average power (in decibels; dB) across all midline electrodes is depicted for the reward feedback phase. A clear burst in the theta frequency band is followed by suppression in the alpha frequency band. (b) Permutation-based clustering of the EEG data: oscillatory power during reward feedback is predicted by SRPEs. The significant positive (red) and negative (blue) clusters are plotted in the time-frequency domain, accompanied by their topographic representations (for plotting details see Methods, last paragraph). Small white dots indicate the channels represented in the cluster, while large white dots indicate the ten most significant electrodes within the cluster. SRPEs significantly predict the average power (in decibels, dB) in the theta cluster (panel c), although these results are better explained by the URPE account (red dotted line). Furthermore, SRPEs accounted for activity in the high-beta cluster (panel d) and high-alpha cluster (panel e).

modulations during reward feedback (family-wise corrected clustering procedure; see Methods). Three clusters correlated with SRPEs during reward feedback (Fig. 3b). First, a positive relation was found between SRPEs and power estimates in the theta band (4–8 Hz), peaking at approximately 300 ms ($p < 0.001$). A second cluster correlating with SRPEs was located in the high-beta range (20–30 Hz), peaking around 500 ms post feedback onset ($p = 0.036$). Third, there was a negative relation between SRPEs and high-alpha power (10–17 Hz), peaking between 500 ms and 1000 ms after feedback onset ($p < 0.001$).

For each of the three clusters, we further investigated the trial-by-trial relation between the average cluster activation and SRPEs in a linear mixed model. As Fig. 3c reveals, although the theta cluster correlated positively with SRPEs, the average power in the theta cluster did not exhibit an SRPE profile. On the contrary, because the theta power increased for large positive as well as large negative RPEs, theta power exhibits an URPE profile; this was confirmed by a significant relation between URPEs and activity in the theta cluster, $\chi^2(1, N = 41) = 248$, $p < 0.001$. The average activity in the high-beta cluster closely matched the predictions from the SRPE account as Fig. 3d demonstrates, $\chi^2(1, N = 41) = 48.9$, $p < 0.001$. The high-alpha power in this third cluster (Fig. 3e) also exhibited the linear SRPE profile, $\chi^2(1, N = 41) = 209$, $p < 0.001$. In sum, these results demonstrate an early occurrence of URPEs in the theta band, and two slightly later SRPE signatures in the high-beta and high-alpha bands during reward feedback.

4. Discussion

In this paper, we manipulated RPEs and replicated an effect of RPEs on one-shot declarative learning (De Loof et al., 2018). Large and positive RPEs were associated with increased recognition of word pairs, with stronger effects in the delayed recognition test, despite only a single exposure during declarative learning. We further demonstrated oscillatory signatures of RPEs during reward feedback in the theta (4–8 Hz), high-beta (20–30 Hz), and high-alpha (10–17 Hz) bands. There is an initial URPE (theta) signature in the EEG feedback signal, which later evolves towards an SRPE (beta, alpha) signature.

We interpret the effect of RPEs on declarative learning in the neo-Hebbian framework (Lisman et al., 2011), where declarative learning depends on pre- and postsynaptic activity. Here, the experience of an RPE is accompanied by the release of dopamine (DA) in the hippocampus, where DA modulates hippocampal encoding (Shohamy and Adcock, 2010). This process enables long-term potentiation (LTP), a biological mechanism underlying synaptic strength and memory formation (Munakata and Pfaffly, 2004). For this reason, experiencing an RPE is an ideal time to learn. Because DA seems to particularly influence late LTP, the release of DA should especially enhance memory for longer retention intervals. The observation of a stronger SRPE-effect in the delayed recognition test in our study, may be in line with this prediction; however, this interpretation requires caution and follow-up, given that it was not significant in the current study.

Importantly, we also tested whether this SRPE effect could be localized in the reward feedback phase, as our theory predicts, and in which frequency band specifically. We indeed observed several clusters showing a significant SRPE effect. Our first cluster after feedback onset followed an URPE signature and was located in the theta band (4–8 Hz). Earlier studies also found theta power to scale with unsigned (absolute) prediction error (Cavanagh et al., 2011; Cavanagh et al., 2010; Mas-Herrero and Marco-Pallarés, 2014). Interestingly and in line with our findings, activity in the theta frequency band has been shown to signal the evaluation of (reward) feedback based on prior knowledge and current predictions (Cavanagh et al., 2009). Increased theta power during encoding has also been associated with enhanced memory performance. For example, intracranial EEG measures revealed that during the encoding of episodic memories, increased theta power predicted successful subsequent memory recall (Sederberg et al., 2003). These findings lend further support to the claim that theta power is crucial in the formation of new declarative memories.

A second cluster was found in the beta band (20–30 Hz). Beta power increased with increasing SRPEs. Several studies observed a relation between beta power and learning. In a study by Siegel et al. (2009), monkeys performed a short-term memory task where two visual objects had to be remembered. During memory encoding, Siegel et al. (2009) observed increased beta power and spike-LFP (local field potential) coupling in the beta band (32 Hz). Moreover, in an RL task, van de Vijver et al. (2011) found increased beta power over the medial frontal cortex. Beta power was especially enhanced for trials on which participants received positive feedback. Interestingly, in a procedural learning experiment, HajiHosseini et al. (2012) found an effect on beta power of both reward valence and probability, consistent with our study. HajiHosseini et al. (2012) source-localized beta power to the dorsolateral prefrontal cortex, which is in line with the frontal beta topography we found. Recent evidence suggests increased activity in the beta band (16–25 Hz) to be implicated in gluing together individual events into a coherent memory representation (Morton and Polyn, 2017). Taken together, these findings provide confirmatory evidence that during declarative learning, reward valence and probability are taken into account and are reflected in the beta band, just like in procedural learning. As such, these findings lend support to our claim that SRPEs are imperative in declarative learning, as is the case in procedural learning.

Our third cluster was located in the alpha band (10–17 Hz). Alpha power has also been associated with enhanced semantic encoding (Klimesch, 1999) and ongoing fluctuations in alpha activity have consistently been shown to boost encoding of declarative information in long-term memory (Fell and Axmacher, 2011; Khader, 2010; Kleberg et al., 2014). Specifically, alpha power is thought to index inhibition (Janssens et al., 2017; Jensen and Mazaheri, 2010) and has been associated with the gating of relevant information and suppression of irrelevant information (Ketz et al., 2015; Park et al., 2014). Consistently, we observed a negative relation between SRPE and alpha. Alpha power was increased for negative RPEs and decreased with larger and positive RPEs. This might indicate participants are suppressing the irrelevant word pair for incorrect (i.e., unrewarded) trials, while anticipating the correct to-be-learned word pair to appear on the screen.

One advantage of our approach is that we constructed the experimental design in such a way that we are able to derive predictions from the model independent of parameter settings. In particular, we presented each word only a single time during the acquisition task and chose the stimuli such that there would likely be no bias toward any of the words. We can therefore make predictions about the relative sizes of a participant's prediction error without requiring participant-specific parameters (e.g. learning rate).

Recently, there has been an increasing interest in examining the intricate relationship between reward, RPEs and declarative memory; we propose that further exploration of such relations may shed new light on extant findings from the declarative memory literature. For example, our finding that SRPEs boost declarative learning, might offer an alternative explanation to the testing effect: the finding that testing, rather than mere studying, is more beneficial to memory (Karpicke and Roediger, 2008; Roediger et al., 2011). Although testing is known to drive declarative learning, its underlying mechanism remains ambiguous. We propose that the testing effect emanates from RPEs. During testing, predictions are generated that are then verified by (external or internal) feedback. This feedback might generate RPEs, initiating a time window for enhanced learning, and therefore making testing more advantageous to memory formation.

Our study also has some limitations. First of all, the number of trials ($n = 60$) is rather small. We mitigated this concern by treating SRPEs as a continuous predictor so that only a single regression coefficient had to be estimated across all 60 trials. This analytical approach is more feasible than increasing the number of trials and letting participants study many more word pairs. Especially considering the fact that participants are exposed to each word pair only once, increasing the number of trials would lead to floor effects in the memory performance. Second, we tested many more females ($n = 33$) than males ($n = 8$). Indeed, there might be a difference in reward sensitivity between genders. In particular, one study showed a difference in information processing styles between males and females during decision making (Byrne and Worthy, 2015). In another study by Ding et al. (2017), a significant difference was observed between males and females in how reward and punishment feedback are processed. Although including gender differences might be interesting, our current sample size, with an unbalanced number of participants in the two gender groups, does not allow for such follow-up, between-subject analyses. Moreover, as the main effects of interest (i.e., RPE and test delay) are uncorrelated with gender, any effects of gender can only make our design less sensitive (more conservative), but not bias our results. Third, although our EEG methodology was well-suited for identifying the frequency bands and time course of reward feedback processing, spatial localization is more difficult. Earlier fMRI work identified RPEs in the striatum and hippocampus (O'Doherty et al., 2003), also in declarative learning tasks (Davidow et al., 2016; Scimeca et al., 2016). The deep (subcortical) localization of these sources in combination with volume conduction may explain why the effect is spread across the scalp, but this obviously needs further testing.

In summary, we found a beneficial effect of SRPEs on declarative learning. The occurrence of reward prediction errors was further validated by oscillatory signatures, first in the theta band reflecting an URPE signal, then in the high-beta and high-alpha band reflecting an SRPE signal. These findings confirm that RPE-triggered oscillatory power variations prior to encoding relates to the successful formation of declarative memories. As such, we provide empirical evidence for SRPEs enhancing declarative learning on both behavioral and neural levels, similar to typical procedural learning paradigms. The interplay and mechanistic roles of URPE and SRPE behavioral and neural signatures, remains to be investigated.

Author contributions

KE, EDL and TV designed research; KE collected data; KE, EDL and CJ

analyzed data; KE, EDL and TV wrote the paper. All authors approved the final version of the manuscript for submission.

Acknowledgements

The authors declare no conflict of interest. This work was supported by the Research Foundation Flanders grant numbers 1153418 (to KE) and G012816 (to TV and KE). TV was supported by grant BOF17-GOA-004 from Ghent University.

Appendix

Table A
Swahili words (240)

adhabu	chupi	jeraha	kioo	maisha	msitu	nyundo	surali
adui	daima	jibini	kisiwa	maji	msumari	nyundu	takatak
afya	dakika	jikoni	kisu	mali	mtawa	nzuri	tamasha
aibu	daraja	jiwe	kitanda	mamba	mtirka	ofisi	tanuri
akili	dari	jokofu	kitande	mapafu	mundamo	osha	tembo
alizeti	dizeli	jua	kiti	mashua	mungu	panya	trekta
amani	duka	jumatu	kito	matumai	mvringo	petye	tumbili
asili	elfu	juuya	kitovu	matumbo	mvua	picha	tumbo
baadaye	farasi	kaburi	kofia	maua	mvuke	pombe	twai
bafuni	fedha	kahawa	kovuli	mazishi	mwanake	punda	uadui
bahari	filimbi	kalamu	kuacha	mbolea	mwanga	punguza	uchorai
baharia	funzi	kamba	kuandika	mbuzi	mwezi	pwani	ufagio
baiski	furaha	kamwe	kubale	mbwa	mzungu	rafiki	ugomvi
bandari	garisi	kartasi	kubwa	mchanga	nanga	rangi	uhuru
barua	geza	katika	kudhibi	mchawi	nchi	rombus	ukame
basi	godoro	kawaida	kuhesa	mchuzi	ndaniya	sabuni	ukweli
bega	goti	kazi	kujenga	mdudu	ndege	sahani	umasijo
bendi	gundi	kelele	kukimba	mechezo	ndevu	samaki	uongo
bilaska	guruwe	kemia	kumba	mekno	ndizi	sayari	usiku
bloke	haki	kengele	kumbuka	mfuko	ndogo	seesaw	uyoga
buli	hamsi	kesho	kununa	mgonjwa	ndoora	sehemu	viatu
bunifu	hasira	kiatu	kunywa	miaka	ndugu	seri	wakala
bustani	hatua	kichwa	kupanda	mkasi	neyemba	shimoni	washia
chaki	hazini	kidole	kusanya	mkate	ngazi	shule	welder
chombo	hofu	kifua	kushoto	mkoba	ngono	simu	wengine
choori	ijayo	kihozi	kusikiza	mkuu	ngozi	singizi	wimbo
chubani	imani	kijiko	kuzama	mlango	nopya	soko	wingi
chuki	ishara	kikapu	kweli	moyo	nyange	starehe	wingu
chuma	ishiri	kimysa	leso	mpishi	nyeusi	stork	yatima
chupa	jansa	kinywa	mageho	mraba	nyota	sufuria	zeituni

Dutch words (60)

agent	bord	ezel	kaas	mest	rijst	stoel	wolk
anker	brief	fiets	kassa	nacht	schat	stoom	wonde
appel	bril	goud	knie	neus	sjaal	stuur	worm
bezem	broek	graf	laken	olijf	slaap	touw	zomer
bier	brood	hamer	lamp	oven	slang	trein	
bloem	doos	haven	lepel	paard	slot	tuin	
boer	eend	hond	lijm	poort	stier	verf	
boot	emmer	hoofd	melk	regen	stift	water	

References

Adcock, R.A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., Gabrieli, J.D., 2006. Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron 50, 507–517. https://doi.org/10.1016/j.neuron.2006.03.036.

Byrne, K.A., Worthy, D.A., 2015. Gender differences in reward sensitivity and information processing during decision-making. J. Risk Uncertain. 50 (1), 55–71. https://doi.org/10.1007/s11166-015-9206-7.

Cavanagh, J.F., Cohen, M.X., Allen, J.J.B., 2009. Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring, pp. 98–105, 29(1). https://doi.org/10.1523/JNEUROSCI.4137-08.2009.

Cavanagh, J.F., Frank, M.J., Klein, T.J., Allen, J.J.B., 2010. Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. Neuroimage 49 (4), 3198–3209. https://doi.org/10.1016/j.neuroimage.2009.11.080.

Cavanagh, J.F., Figueroa, C.M., Cohen, M.X., Frank, M.J., 2011. Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. Cerebr. Cortex 12, 2575–2586. https://doi.org/10.1093/cercor/bhr332.

Cohen, M.X., 2014. Analyzing Neural Time Series Data: Theory and Practice. MIT Press.

Cohen, M.X., Elger, C.E., Ranganath, C., 2007. Reward expectation modulates feedback-related negativity and EEG spectra. Neuroimage 35 (2), 968–978. https://doi.org/10.1016/j.neuroimage.2006.11.056.

Davidow, J.Y., Foerde, K., Galván, A., Shohamy, D., 2016. An upside to reward sensitivity: the hippocampus supports enhanced reinforcement learning in adolescence. Neuron 92 (1), 93–99. https://doi.org/10.1016/j.neuron.2016.08.031.

De Loof, E., Ergo, K., Naert, L., Janssens, C., Talsma, D., Van Opstal, F., Verguts, T., 2018. Signed reward prediction errors drive declarative learning. PloS One 13 (1) e0189212. https://doi.org/10.1371/journal.pone.0189212.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Ding, Y., Wang, E., Zou, Y., Song, Y., Xiao, X., Huang, W., Li, Y., 2017. Gender differences in reward and punishment for monetary and social feedback in children: an ERP study. PloS One 12 (3) e0174100. https://doi.org/10.1371/journal.pone.0174100.

Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., Uchida, N., 2015. Arithmetic and local circuitry underlying dopamine prediction errors. Nature 525 (7568), 243–246. https://doi.org/10.1038/nature14855.

Eshel, N., Tian, J., Bukwich, M., Uchida, N., 2016. Dopamine neurons share common response function for reward prediction error. Nat. Neurosci. 19 (3), 479–486. https://doi.org/10.1038/nn.4239.

Fell, J., Axmacher, N., 2011. The role of phase synchronization in memory processes. Nat. Rev. Neurosci. 12 (2), 105–118. https://doi.org/10.1038/nrn2979.

Haegens, S., Nácher, V., Hernández, A., Luna, R., Jensen, O., Romo, R., 2011. Beta oscillations in the monkey sensorimotor network reflect somatosensory decision

making. Proc. Natl. Acad. Sci. U. S. A. 108 (26), 10708–10713. https://doi.org/10.1073/pnas.1107297108.
HajiHosseini, A., Holroyd, C.B., 2015. Sensitivity of frontal beta oscillations to reward valence but not probability. Psychophysiology 602, 99–103. https://doi.org/10.1016/j.neulet.2015.06.054.
HajiHosseini, A., Rodríguez-Fornells, A., Marco-Pallarés, J., 2012. The role of beta-gamma oscillations in unexpected rewards processing. Neuroimage 60 (3), 1678–1685. https://doi.org/10.1016/j.neuroimage.2012.01.125.
Howard-Jones, P., Demetriou, S., Bogacz, R., Yoo, J.H., Leonards, U., 2011. Toward a science of learning games. Mind, Brain, and Education 5 (1), 33–41. https://doi.org/10.1111/j.1751-228X.2011.01108.x.
Janssens, C., De Loof, E., Boehler, C.N., Pourtois, G., Verguts, T., 2017. Occipital alpha power reveals fast attentional inhibition of incongruent distractors. Psychophysiology. https://doi.org/10.1111/psyp.13011.
Jasper, H., 1958. The ten twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 371–375.
Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front. Hum. Neurosci. 4, 186. https://doi.org/10.3389/fnhum.2010.00186.
Karpicke, J.D., Roediger, H.L., 2008. The critical importance of retrieval for learning. Science 319 (5865), 966–968. https://doi.org/10.1126/science.1152408.
Ketz, N.A., Jensen, O., O'Reilly, R.C., 2015. Thalamic pathways underlying prefrontal cortex–medial temporal lobe oscillatory interactions. Trends Neurosci. 38 (1), 3–12. https://doi.org/10.1016/J.TINS.2014.09.007.
Khader, P.H., 2010. Theta and alpha oscillations during working-memory maintenance predict successful long-term memory encoding. Neurosci. Lett. 468, 339–343. https://doi.org/10.1016/j.neulet.2009.11.028.
Kleberg, F.I., Kitajo, K., Kawasaki, M., Yamaguchi, Y., 2014. Ongoing theta oscillations predict encoding of subjective memory type. Neurosci. Res. 83, 69–80. https://doi.org/10.1016/j.neures.2014.02.010.
Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29 (2), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3.
Lisman, J., Grace, A.A., Duzel, E., 2011. A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. Trends Neurosci. 34 (10), 536–547. https://doi.org/10.1016/j.tins.2011.07.006.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.
Mas-Herrero, E., Marco-Pallarés, J., 2014. Frontal theta oscillatory activity is a common mechanism for the computation of unexpected outcomes and learning rate. J. Cognit. Neurosci. 26 (3), 447–458. https://doi.org/10.1162/jocn.
Mathys, C., Daunizeau, J., Friston, K.J., Stephan, K.E., 2011. A Bayesian foundation for individual learning under uncertainty. Front. Hum. Neurosci. 5, 39. https://doi.org/10.3389/fnhum.2011.00039.
Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive hebbian learning. J. Neurosci. 76 (5), 1936–1947.
Morton, N.W., Polyn, S.M., 2017. Beta-band activity represents the recent past during episodic encoding. Neuroimage 147, 692–702. https://doi.org/10.1016/J.NEUROIMAGE.2016.12.049.
Munakata, Y., Pfaffly, J., 2004. Hebbian learning and development. Dev. Sci. 72, 141–148. https://doi.org/10.1111/j.1467-7687.2004.00331.x.
Ohlsson, S., 1996. Learning from performance errors. Psychol. Rev. 103 (2), 241–262. https://doi.org/10.1037/0033-295X.103.2.241.
Oya, H., Adolphs, R., Kawasaki, H., Bechara, A., Damasio, A., Howard, M.A., 2005. Electrophysiological correlates of reward prediction error recorded in the human prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 102 (23), 8351–8356. https://doi.org/10.1073/pnas.0500899102.
O'Doherty, J.P., Dayan, P., Friston, K.J., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38 (2), 329–337.
Park, H., Lee, D.S., Kang, E., Kang, H., Hahm, J., Kim, J.S., Jensen, O., 2014. Blocking of irrelevant memories by posterior alpha activity boosts memory encoding. Hum. Brain Mapp. 35 (8), 3972–3987. https://doi.org/10.1002/hbm.22452.
Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2 (1), 79–87.
Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Blake, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton-Century-Croft, New York, pp. 64–99.
Roediger, H.L., Putnam, A.L., Smith, M.A., 2011. Ten benefits of testing and their applications to educational practice. Psychol. Learn. Motiv. 55 (1). https://doi.org/10.1016/B978-0-12-387691-1.00001-6.
Schneider, W., Eschman, A., Zuccolotto, A., 2012. E-prime User's Guide. Psychology Software Tools, Inc, Pittsburgh.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275 (5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593.
Scimeca, J.M., Katzman, P.L., Badre, D., 2016. Striatal prediction errors support dynamic control of declarative memory decisions. Nat. Commun. 7, 13061. https://doi.org/10.1038/ncomms13061.
Sederberg, P.B., Kahana, M.J., Howard, M.W., Donner, E.J., Madsen, J.R., 2003. Theta and gamma oscillations during encoding predict subsequent recall. J. Neurosci. 23 (34), 10809–10814.
Shohamy, D., Adcock, R.A., 2010. Dopamine and adaptive memory. Trends Cognit. Sci. 14 (10), 464–472. https://doi.org/10.1016/j.tics.2010.08.002.
Siegel, M., Warden, M.R., Miller, E.K., 2009. Phase-dependent neuronal coding of objects in short-term memory. Proc. Natl. Acad. Sci. Unit. States Am. 106 (50), 21341–21346.
Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. Nature 529 (7587), 484–489. https://doi.org/10.1038/nature16961.
Silvetti, M., Alexander, W., Verguts, T., Brown, J.W., 2014. From conflict management to reward-based decision making: actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. 46, 44–57. https://doi.org/10.1016/j.neubiorev.2013.11.003.
Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., Janak, P.H., 2013. A causal link between prediction errors, dopamine neurons and learning. Nat. Neurosci. 16 (7), 966–973. https://doi.org/10.1038/nn.3413.
Sutton, R.S., Barto, A.G., 1981. Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev. 88 (2), 135–170. https://doi.org/10.1037/0033-295X.88.2.135.
Sutton, R.S., Barto, A.G., 1998. Reinforcemient Learning: an Introduction. MIT Press, Cambridge, MA.
R Core Team, 2014. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
van de Vijver, I., Ridderinkhof, K.R., Cohen, M.X., 2011. Frontal oscillatory dynamics predict reedback learning and action adjustment. J. Cognit. Neurosci. 23 (12), 4106–4121. https://doi.org/10.1162/jocn_a_00110.
Wittmann, B.C., Schott, B.H., Guderian, S., Frey, J.U., Heinze, H.-J., Düzel, E., 2005. Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron 45 (3), 459–467. https://doi.org/10.1016/j.neuron.2005.01.010.