International Journal of Psychophysiology 132 (2018) 236–242

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

The application of reward learning in the real world: Changes in the reward positivity amplitude reflect learning in a medical education context☆

Chad C. Williams[a,*], Kent G. Hecker[b,c], Michael K. Paget[c], Sylvain P. Coderre[c], Kelly W. Burak[c], Bruce Wright[d], Olave E. Krigolson[a]

[a] Centre for Biomedical Research, University of Victoria, Canada
[b] Faculty of Veterinary Medicine, University of Calgary, Canada
[c] Cumming School of Medicine, University of Calgary, Canada
[d] Division of Medical Science, University of Victoria, Canada

ABSTRACT

Evidence ranging from behavioural adaptations to neurocognitive theories has made significant advances into our understanding of feedback-based learning. For instance, over the past twenty years research using electroencephalography has demonstrated that the amplitude of a component of the human event-related brain potential – the reward positivity – appears to change with learning in a manner predicted by reinforcement learning theory (Holroyd and Coles, 2002; Sutton and Barto, 1998). However, while the reward positivity (also known as the feedback related negativity) is well studied, whether the component reflects an underlying learning process or whether it is simply sensitive to feedback evaluation is still unclear. Here, we sought to provide support that the reward positivity is reflective of an underlying learning process and further we hoped to demonstrate this in a real-world medical education context. In the present study, students with no medical training viewed a series of patient cards that contained ten physiological readings relevant for diagnosing liver and biliary disease types, selected the most appropriate diagnostic classification, and received feedback as to whether their decisions were correct or incorrect. Our behavioural results revealed that our participants were able to learn to diagnose liver and biliary disease types. Importantly, we found that the amplitude of the reward positivity diminished in a concomitant manner with the aforementioned behavioural improvements. In sum, our data support theoretical predictions (e.g., Holroyd and Coles, 2002), suggest that the reward positivity is an index of a neural learning system, and further validate that this same system is involved in learning across a wide range of contexts.

1. Introduction

Converging evidence has made significant advances into understanding how humans learn from feedback. Whereas pioneer research has described how behaviours change in response to rewards and punishments (Skinner, 1958), more recent studies have theorized the neural mechanisms that underlie reward learning systems within the brain (Holroyd and McClure, 2015). In particular, neuroimaging studies have discovered that there are at least two neurocognitive mechanisms to learning from feedback. First, it has become evident that there is an early, unconscious system that is sensitive to violations of expectancy (Holroyd and Coles, 2002; Holroyd and Krigolson, 2007; Krigolson et al., 2014; Sutton and Barto, 1998). Second, there also appears to be a later conscious system responsible for updating mental representations of the environment in order to adapt behaviours and predictions (Sato et al., 2005; Yeung and Sanfey, 2004). The former of these processes has been theorized to be driven by the midbrain dopamine system which delivers signals that reflect reward prediction errors – the degree to which the predictions of outcomes do not match the actual outcomes – to the anterior cingulate cortex (ACC; Holroyd and McClure, 2015; Schultz et al., 1997). More precisely, within this specific framework (i.e., Holroyd and Coles, 2002), prediction errors are computed within the basal ganglia, and projected to the ACC via the midbrain dopamine system. Computational theories describe the ACC to be a 'controller' of cognitive resources in that it integrates these dopamine signals and directs how to best use resources across the brain in order to learn from the environment (Holroyd and McClure, 2015).

Over the past twenty years there has been a large body of work

☆ Sources of support: Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN 2016-0943), and the Neuroeducation Network (007).
* Corresponding author at: School of Exercise Science, Physical & Health Education, University of Victoria, P.O. Box 17000, STN CSC, Victoria, British Columbia V8W 2Y2, Canada.
E-mail address: ccwillia@uvic.ca (C.C. Williams).

http://dx.doi.org/10.1016/j.ijpsycho.2017.10.010
Received 31 July 2017; Received in revised form 22 September 2017; Accepted 26 October 2017
Available online 27 October 2017
0167-8760/

examining the electroencephalographic (EEG) responses of these reward signals. In 1997, Miltner, Braun, and Coles first reported the feedback related negativity (FRN), a component of the human event-related brain potential (ERP) evoked by the processing of outcome feedback that is now theorized by some to reflect the arrival of dopamine signals at the ACC (Holroyd and Coles, 2002; Holroyd and McClure, 2015; Holroyd and Yeung, 2012; Schultz et al., 1997). More recently, it has been suggested that the FRN should be framed as reward positivity reflecting the sensitivity of this component to positive as opposed to negative outcomes (Foti et al., 2011; Holroyd et al., 2008; Proudfit, 2015). The reward positivity component arises in frontal-central regions of the scalp 250 to 350 ms following performance feedback (Proudfit, 2015). Specifically, it is theorized to be the ERP analog of reward prediction error dopamine signals arriving at the ACC (Holroyd and Coles, 2002; Holroyd and McClure, 2015; Holroyd and Yeung, 2012).

If the Holroyd and Coles hypothesis is true, it seems logical that the amplitude of the reward positivity should reflect underlying learning processes – yet, to date, findings are mixed. For instance, Krigolson et al. (2014) demonstrated that the amplitude of the reward positivity diminished with learning, a result also reported by the same group in 2009 (Krigolson, Pierce, Tanaka, & Holroyd) and by others (Bellebaum and Colosio, 2014; Bellebaum and Daum, 2008; Eppinger et al., 2008; Luque et al., 2012; Sailer et al., 2010). Other studies, however, have found that behavioural and neural changes linked to learning did not always coincide (Bellebaum et al., 2010; Eppinger et al., 2009; Groen et al., 2007; Hämmerer et al., 2010; Holroyd and Coles, 2002; Nieuwenhuis et al., 2002; see Walsh and Anderson, 2012 for a review). As such, it is unclear if the reward positivity reflects an underlying learning process or whether it is simply sensitive to feedback evaluation.

One potential explanation for the conflicting findings may relate to whether the information in experimental paradigms is relevant and/or learnable. For example, in some of the gambling paradigms typically used to study the reward positivity no learning can actually occur. This was explored by Bellebaum and Colosio (2014) who had participants make decisions about alphabetic characters in which feedback for some stimuli was contingent on participant responses (learning could occur), while for other stimuli it was not (learning could not occur). They found that the reward positivity amplitude decreased across the task only for the contingent stimuli. As such, it appears to be important that we study the reward positivity in tasks where learning can occur. Related to that, is the relationship between information and outcomes. Specifically, in the aforementioned studies participants had to learn about shapes (Bellebaum and Daum, 2008; Bellebaum et al., 2010; Krigolson et al., 2009; Krigolson et al., 2014; Sailer et al., 2010), simple objects (Eppinger et al., 2008; Eppinger et al., 2009; Groen et al., 2007; Holroyd and Coles, 2002; Luque et al., 2012), and alphabetic characters (Bellebaum and Colosio, 2014; Hämmerer et al., 2010; Nieuwenhuis et al., 2002). However, in none of these experiments did the stimuli naturally lead to a correct answer. In other words, the stimulus–response mappings were in a sense both arbitrary and meaningless. Put another way, what was learned by participants in these studies could never be used in, nor ever arise from, any natural environments.

In contrast, behavioural research has explored learning in real-world contexts. For example, two recent studies have demonstrated the efficacy of reinforcement learning in medical education (Anderson et al., 2016; Xu et al., 2016). Anderson et al. (2016) used a reinforcement learning paradigm to enhance the teaching of neuroanatomy to medical students. Specifically, they had participants learn to identify neuroanatomical structures via a computer based trial and error shaping process – participants saw an image with a label, determined whether the structure and label were correctly matched, and were provided with feedback about the accuracy of their response. Importantly, participants learned to identify multiple neuroanatomical structures as was indicated by increasing accuracy rates and decreasing response times (Anderson et al., 2016). Further evidence supporting this in a medical education context comes from Xu et al. (2016) who used a similar approach to teach students to correctly categorize melanoma. These paradigms are progressing in the correct direction, yet still rely on simple stimuli (e.g., an image). We propose that to truly understand how learning occurs organically it is important to extend these findings to learning more complex real-world material while at the same time investigating the neural processes involved.

Here, we seek to demonstrate that changes in the reward positivity are related to an actual learning process and moreover that the system underlying this component plays a role when learning complex real-world material. In the current study, participants were to learn to diagnose liver and biliary diseases by making judgments on patient case-studies and utilizing simple performance feedback while electroencephalographic data were recorded. We hypothesized that participants would be able to learn complex data structures in order to categorize clinical cases through the use of a reinforcement learning paradigm. Specifically, we predicted that accuracy rates would be higher and reaction times (i.e., viewing the patient card and viewing the diagnostic options) would be quicker late in each phase, when learning has occurred, as opposed to early in each phase. Further, we predicted that participants would score higher than chance on a retention test. In regards to neural data, we hypothesized that performance feedback would elicit a reward positivity – indicating the processing of said feedback. Importantly, we also predicted that the amplitude of the reward positivity would diminish with learning – a result in line with previous work and theoretical predictions (i.e., Sutton and Barto, 1998).

2. Methods

2.1. Participants

Thirty undergraduate students with no medical training (23 female, mean age 20 years old [CI: ± 1 year]) from the University of Victoria participated in the experiment. All participants had normal or corrected-to-normal vision, no neurological impairments, and volunteered for extra course credit in a psychology course. Four participants were removed as they did not progress past the first phase (see below) resulting in twenty-six participants (19 female, mean age 20 years old [CI: ± 1 year]). All participants provided informed consent approved by the Human Research Ethics Board at the University of Victoria, and the study followed ethical standards as prescribed in the 1964 Declaration of Helsinki.

2.2. Apparatus and procedure

Participants were seated in a sound dampened room in front of a 19″ LCD computer monitor and used a handheld 5-button RESPONSEPixx (VPixx, Vision Science Solutions, Quebec, Canada) controller to complete an adaptation of the Cards reinforcement learning paradigm (Bannister et al., 2016; Burak et al., 2015; Horrey et al., 2016; Kazoleas, 2016; Tang et al., 2016) written in MATLAB (Version 8.6, Mathworks, Natick, U.S.A.) using the Psychophysics Toolbox extension (Brainard, 1997).

Cards teaches participants through the application of reinforcement learning principles. In our experiment, participants were presented with physiological data (e.g., liver enzyme values) which they then used to make clinical decisions. Specifically, participants learned to classify five types of liver and biliary diseases: cholestatic intrahepatic, cholestatic extrahepatic, mild hepatocellular, moderate hepatocellular, and severe hepatocellular. This classification mimics the first step of cognitive organization structures called “schemes”, a process particularly ascribed to expertise (Coderre et al., 2003). During each clinical case (i.e., trial) of the experiment, participants were shown a patient case-study card followed by a multiple-choice presentation of the diagnostic

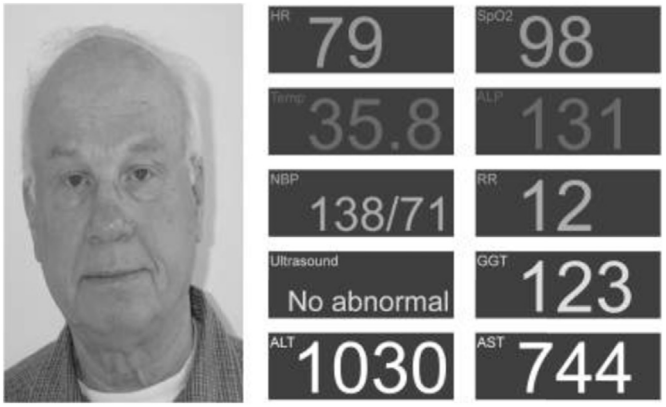

Fig. 1. Example of a patient card that is presented to participants. Card includes a patient photo and ten physiological readings. The card was presented in colour.

classification options. Following a participant's diagnosis, a feedback screen indicated whether or not the diagnostic classification was correct or incorrect.

The patient card presented included a photo of the 'patient' and 10 physiological readings (see Fig. 1). The patient's photo was randomly determined, without replacement, by a pool of 357 profiles (69% female; Minear and Park, 2004). All 'patients' were 50 to 93 years old with no outward manifestations of any liver or biliary diseases. To extend the length of the task, the photos were repeated once in a newly randomized order proceeding the first presentation of the entire set. This resulted in a total of 714 possible trials. The physiological data were displayed in five rows and two columns where the text was displayed in a green, purple, blue, yellow and white Arial font from top to bottom, respectively. To ensure participants were learning to discern which variables were necessary to classify clinical cases (rather than spatial locations), the physiological data were randomly placed across the card on each trial. Five of the ten physiological readings (heart rate, blood oxygen level, blood pressure, respiratory rate, temperature) were distractor variables and were not useful for diagnosing clinical cases. The remainder of the variables (alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, gamma-glutamyl transferase, ultrasound reading) was pertinent to the clinical cases and varied as a function of the patient's disease. For each variable presentation, a random number was generated within a respective and appropriate range thus ensuring that no two cards were the same. All cards were generated and verified by a medical expert in the clinical area as to their accuracy and validity. Importantly, participant did not receive any of the aforementioned information, nor were they trained on any of the variables or diseases in the experiment. Participants were able to view the patient card as long as needed and once ready to make a decision, they pressed a button to continue to the decision stage. For each trial, we measured the card viewing time: the time participants took to indicate that they were ready to make a decision.

The decision stage of each trial involved selecting from one of the five types of liver and biliary diseases. The diagnostic classification options were presented in the same array as the response box in a white Arial font. The options were bordered by a colour (left: green, right: red, top: yellow, bottom: blue) to match the coloured buttons of the response box. Participants selected a specific diagnostic classification option by pushing the appropriate response button, after which the border of the selection raised in brightness. Following this, participants confirmed their response by pressing the center button on the response box, or alternatively, they could unselect their response by pressing any of the outer buttons. To ensure that participants were learning the diagnostic classification (rather than spatial locations), we randomized which diagnostic classifications appeared in a given location. Similar to the patient card stage of the experiment, participants were able to remain in the selection screen as long as they needed which we quantified as diagnostic option viewing time. Once participants submitted their response, a white fixation cross appeared for 400 ms to 600 ms so that the upcoming feedback stimuli could be analyzed independently of motor activity. Feedback of the decision was then presented as either a '✓' (correct) or an 'X' (incorrect) in a white Arial font. This feedback stimulus was presented for 1000 ms. Importantly, this was the only feedback provided to participants – they were never given information on why they had made the correct or incorrect decision, or on how to reach the correct decision. At the offset of feedback, the next trial would begin by presenting the next patient card.

To facilitate learning, the experimental task was broken into four phases. In the first phase, participants only had to learn two of the five possible clinical diagnostic classifications. After participants were able to achieve an accuracy of 90% for two consecutive blocks, each containing 20 trials, the participants moved to the second phase of the experiment wherein another diagnostic classification was added to the potential cases the participant saw. Again, movement to the next phase of the experiment occurred when participants completed two successive blocks of trials within which they achieved an accuracy of at least 90%. The third and fourth phases occurred in a similar fashion, with an additional possible diagnostic classification being added during each phase to bring the total number of possible case types to four, and five, respectively. The number of available responses (answers) matched the number of diseases in each phase. Once the participant completed phase four, the experiment ended. After a twenty-minute break within which participants completed a distractor task, a pen and paper retention test was given to participants.

In the distractor task, participants were to choose between a series of paired coloured squares, one of which was more often rewarding than the other. The purpose of this distractor task was to ensure that participants' performance on the post-test relied on information that had been consolidated to long-term memory (Liu and Fu, 2007). The post-test consisted of twenty multiple-choice questions. Each question presented novel clinical data with the same layout as within the experiment (i.e., patient card) and participants were to indicate the classification by selecting one of the five clinical case options. Unbeknownst to the participants, there were four patients for each disease type randomly distributed throughout the test.

2.3. Data acquisition and processing

Accuracy rates, card view times, and response option view times were recorded using the RESPONSEPixx controller (VPixx, Vision Science Solutions, Quebec, Canada). Behavioural measures were binned as the first twenty (early) and last twenty (late) trials of each phase. For each participant, behavioural analyses examined block accuracy rates, card view times, and response option view times. Grand average behavioural data were created by averaging the results of all corresponding participants.

EEG data were recorded from 64 electrodes which were mounted in a fitted cap with a standard 10-10 layout (ActiCAP, Brainproducts GmbH, Munich, Germany). Electrodes on the cap were initially referenced to a common ground. On average, electrode impedances were kept below 20 kΩ. The EEG data were sampled at 500 Hz, amplified (ActiCHamp, Revision 2, Brainproducts GmbH, Munich, Germany), and filtered through an antialiasing low-pass filter of 8 kHz. To ensure temporal accuracy of stimuli and data a DATAPixx processing box (VPixx, Vision Science Solutions, Quebec, Canada) was used.

EEG data were processed as follows using Brain Vision Analyzer software (Version 7.6, Brainproducts, GmbH, Munich, Germany). First, excessively noisy or faulty electrodes were removed. The ongoing EEG data were down sampled to 250 Hz, re-referenced to an average mastoid, and then filtered using a dual pass Butterworth filter with a passband of 0.1 Hz to 30 Hz in addition to a 60 Hz notch filter. Segments spanning 3000 ms - 1000 ms prior to feedback stimuli to

2000 ms following feedback stimuli onset - were created to complement an independent component analysis which was used to correct ocular artifacts (Luck, 2014). Channels that were initially removed were then interpolated using spherical splines. A re-segmentation of the data was then conducted to yield 800 ms epochs ranging from 200 ms prior to feedback stimuli to 600 ms following feedback onset. Following this, each epoch was baseline corrected using the 200 ms of data prior to feedback stimuli onset. Data were then re-segmented into two conditions (correct, incorrect) and two time points (early, late). All waveforms were then processed through an artifact rejection algorithm with a 10 μV/ms gradient and a 150 μV absolute difference criteria. For each participant, condition and time point, ERP waveforms were created by averaging the epoched data for each electrode.

The ERP component of interest was the reward positivity. This frontal-central ERP component is defined as the maximal difference between correct and incorrect feedback waveforms between 250 ms and 350 ms following feedback stimulus onset at electrode FCz (Holroyd and Krigolson, 2007; Miltner et al., 1997). For statistical purposes, we quantified the reward positivity for each condition and participant as the mean voltage ± 25 ms of the peak difference in the grand average waveforms (292 ms) at channel FCz. We chose this time window and channel based on visual inspection of the data and previous literature (Holroyd and Coles, 2002; Holroyd and Krigolson, 2007; Krigolson et al., 2014; Schultz et al., 1997). Note, we only were able to analyze the reward positivity difference (correct – incorrect) in the initial stages of learning given the lack of error trials in later stages. We also compared the correct ERP waveforms for the first twenty trials of all experimental phases (early condition) to the correct ERP waveforms for the last twenty trials of all experimental phases (late condition) to gauge learning related changes in the amplitude of the reward positivity. Although the reward positivity is generally measured as the difference between correct and incorrect waveforms, the change in amplitude of this component is theorized to be driven by a change in the correct waveforms and not in the incorrect waveforms, thus allowing us to measure the amplitude of the reward positivity by solely focusing on the correct waveforms (Holroyd and McClure, 2015; Holroyd and Yeung, 2012). The result of these analyses yielded average correct and average incorrect conditional waveforms (early learning) and two average correct waveforms (early condition, late condition). For each average waveform, a grand average waveform was created by averaging corresponding ERPs across all participants.

2.4. Data analysis

To confirm that reinforcement learning is an effective approach when learning to diagnose liver and biliary disease types, and that the reward positivity can be used to track learning in a medical context, the following statistical approaches were reported on all behavioural and neural measures of interest: 95% confidence intervals, null hypothesis testing (*t*-tests with $\alpha = 0.05$), and effect sizes (Cohen's d). All statistics were conducted in R (version 3.3.0) using R Studio (version 0.99.902).

Behavioural data were analyzed to assess participants' ability to diagnose liver and biliary disease types. The behavioural data of interest were the block accuracy rates, the card view time, the response option view time, and the post-test accuracy rates. For block accuracy rates, card view time, and response option view time, data were binned into early and late conditions. The early condition was defined as the first twenty trials (first block) of all completed phases, while the late condition was defined as the last twenty trials (last block) of all phases completed. A paired samples *t*-test was conducted between the early and late conditions of each behavioural measure. A single sample *t*-test was conducted on the post-test accuracy comparing participants' rates to chance (20%). EEG data were analyzed via two paired samples *t*-tests that were used to compare the amplitude of the reward positivity between correct and incorrect trials and between stage of learning (early, late).

Table 1
Descriptive statistics for all behavioural and neural data comparing early and late conditions. CI = confidence interval. Accuracy rates were determined as the averaged percentage of correct diagnostic decisions out of the 20 trials of the corresponding blocks. Card view time refers to the averaged length of time participants viewed the patient card in the corresponding blocks. Response view time refers to the averaged length of time participants viewed the diagnostic options prior to finalizing their diagnosis for the corresponding blocks. Correct waveform reflects the mean peak amplitude ± 25 ms centered at 292 ms, where the reward positivity peak was maximal, at electrode FCz time-locked to correct feedback onset of the corresponding blocks.

	Early			Late		
	Mean	95% CI		Mean	95% CI	
		Min	Max		Min	Max
Accuracy rates	80.14%	75.56%	84.73%	96.83%	95.74%	97.91%
Card view time	7.57 s	6.75 s	8.39 s	4.58 s	4.10 s	5.07 s
Response view time	4.66 s	4.43 s	4.90 s	3.67 s	3.55 s	3.80 s
Correct waveform	7.53 μV	5.71 μV	9.35 μV	3.68 μV	1.93 μV	5.44 μV

3. Results

Participants completed the task in an average of 48 min [CI: 43 min 54 min] and 274 trials [CI: 229 trials 318 trials]. Descriptive statistics can be found in Table 1 and statistical results can be found on Table 2. Participants averaged 88% [83% 93%] on the post-test, well above chance (20%). Furthermore, participants had higher accuracy rates, shorter viewing times of the patient's cards, and shorter viewing times of the response choices in the late condition compared to early condition (see Fig. 2). These results suggest that participants learned data patterns that led to the correct diagnostic selection. Accompanying these findings was concurrent neural evidence of learning as evidenced by a reward positivity (see Fig. 3A). Further, the correct feedback waveform of the reward positivity decreased in amplitude from the early condition to the late condition (see Figs. 3B and 4). These data indicate that reward positivity amplitude was sensitive to the degree of learning throughout the task.

4. Discussion

Both our behavioural and neural data demonstrated that participants learned the presented material. Specifically, increases in accuracy and decreases in response times from the early condition to the late condition revealed participant improvement at diagnosing liver and biliary disease types. Improvements in accuracy reflected that participants were able to utilize simple feedback as to whether or not they made the correct decision to direct their exploration of the clinical environment and learn the key diagnostic classification features (Gershman and Niv, 2010; Niv et al., 2015; Wilson and Niv, 2012). Ultimately, this led to participants becoming able to identify the necessary patterns within the physiological data that corresponded to each disease type. Additionally, concomitant decreases in response times indicated that participants were learning to more quickly navigate the information presented to them and optimize their assessment strategy to result in the correct decision. Further, we found that, after a twenty-minute distractor task participants retained knowledge on a post-hoc test of the presented clinical material (88% accuracy). Importantly, the retention test is evidence of participant learning – a relatively stable improvement in performance over time (Skinner, 1958).

Here we provided further evidence that the reward positivity reflected a reward learning system in an experimental context that closely mirrors a real world medical education environment. Indeed, an analysis of the difference between correct and incorrect feedback revealed that early in learning a reward positivity was elicited. Perhaps more importantly given our hypotheses, we also demonstrated that learning related changes in the amplitude of the reward positivity appeared to reflect an underlying learning process. Specifically, the correct aspect of

Table 2
Descriptive and inferential statistics for all behavioural and neural data. CI = confidence interval. Post-test accuracy refers to the difference between participants' percentage of correct diagnoses out of 20 on a pen and paper test and chance (20%). Accuracy rates were determined as the difference between the averaged percentage of correct diagnostic decisions out of the 20 trials of the early and late conditions (late – early). Card view time refers to the difference between the averaged length of time participants viewed the patient card of the early and late conditions (late – early). Response view time refers to the difference between the averaged length of time participants viewed the diagnostic options prior to finalizing their diagnosis across the early and late conditions (late – early). Reward positivity reflects the difference in mean amplitude ± 25 ms centered at 292 ms, where the reward positivity peak was maximal, at electrode FCz between correct and incorrect feedback. Correct waveform corresponds to the difference of the mean peak amplitude ± 25 ms centered at 292 ms at electrode FCz time-locked to correct feedback onset of the early and late conditions (late – early).

	Mean difference	95% CI		*t*-Value	p-Value	Cohen's d
		Min	Max			
Post-test accuracy	68%	63%	73%	27.43	< 0.0001	5.38
Accuracy rates	16.68%	12.44%	20.93%	8.09	< 0.0001	1.59
Card View time	− 2.99 s	− 3.59 s	− 2.38 s	− 10.15	< 0.0001	− 1.99
Response view time	− 0.99 s	− 1.16 s	− 0.82 s	− 11.83	< 0.0001	− 2.32
Reward positivity	5.53 μV	2.37 μV	8.68 μV	3.61	0.0013	0.71
Correct waveform	− 3.85 μV	− 5.30 μV	− 2.40 μV	− 5.46	< 0.0001	− 1.07

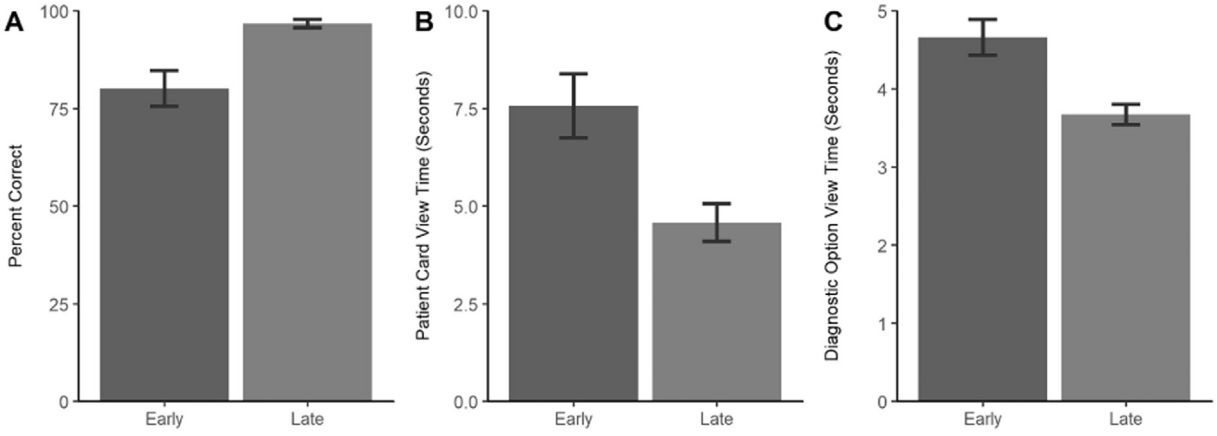

Fig. 2. Behavioural results. A: block accuracy rates, B: card view times, and C: response option view times for early and late conditions in learning. Error bars represent 95% confidence intervals. The early condition refers to the average of the first 20 trials of each phase, while the late condition refers to the average of the last 20 trials of each phase.

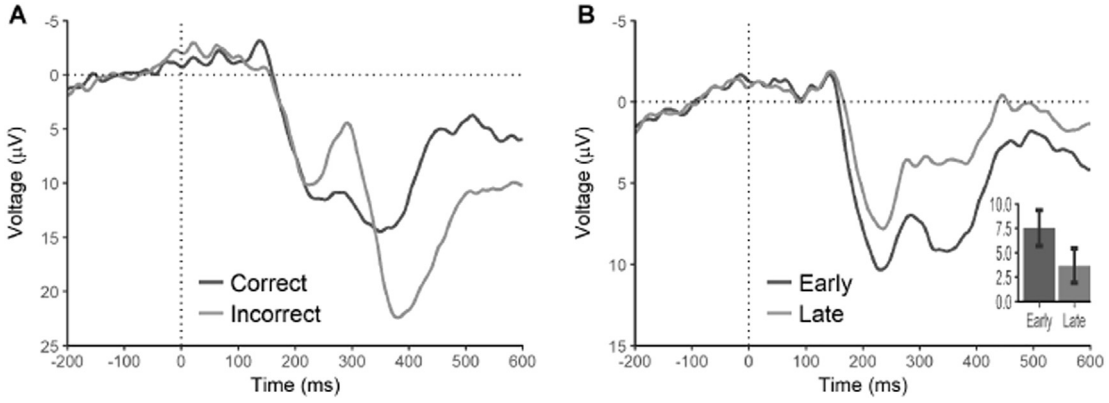

Fig. 3. Neural results at electrode FCz time-locked to feedback stimulus onset. A: correct and incorrect feedback waveforms across the entire experiment. This refers to correct and incorrect feedback irrespective of phase and time in learning. B: correct feedback waveforms binned into early and late conditions to demonstrate learning effects. This refers to the averaged correct feedback for the first twenty trials (early) and last twenty trials (late) of all completed phases. The bar graph within B is the mean peak amplitude ± 25 ms centered around the peak latency, 292 ms, of the corresponding waveforms with 95% confidence intervals.

the reward positivity was reduced in amplitude for the late condition relative to the early condition, paralleling the aforementioned behavioural changes that we observed. As the magnitude of the positive deflection of the reward positivity is theorized to be proportional to the amplitude of reward prediction error signals reaching the ACC (Holroyd and McClure, 2015), the reduction in component amplitude is suggestive that the neural system underlying this component was no longer computing prediction errors. In other words, early in learning participants' expectations of outcomes were not precise and as such in a theoretical framework a prediction error would be computed (i.e., in our data we observed a reward positivity early in learning). However, after participants learned to classify diagnostic disease types their expectations would be the same as their outcomes. Thus, reinforcement learning theory would predict that the amplitude of the prediction error would be reduced. Supporting this, in our data we saw a related decrease in the amplitude of the reward positivity with learning. In other words, our data suggest that the amplitude of the reward positivity is an index of degree of learning.

Our task also provided insight as to whether the reward positivity is involved in real-world learning, rather than being an artifact of simple

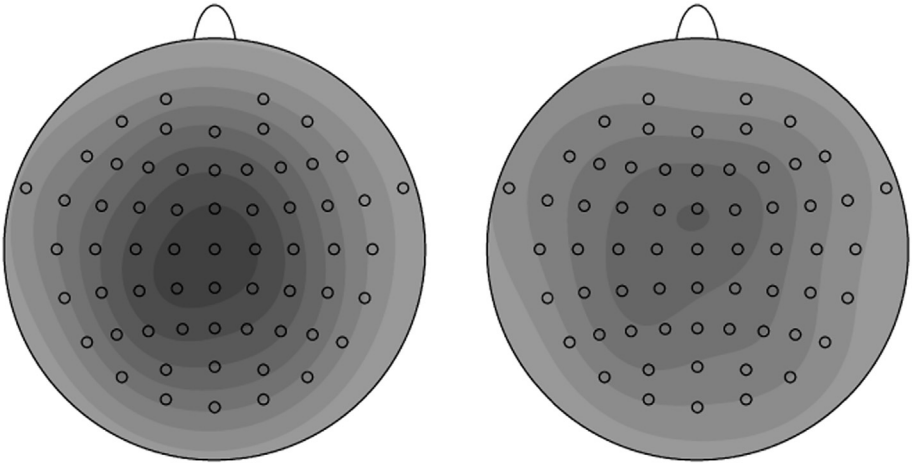

Fig. 4. Topographic maps for early condition (left) and late condition (right). Early and late conditions refer to the averaged correct feedback for the first twenty trials (early) and last twenty trials (late) of all completed phases, respectively. The potential scale ranges from − 10 μV (white) to 10 μV (dark grey) in 16 steps.

experimental stimulus-response paradigms. This is an important consideration because all studies exploring the change in reward positivity amplitude across learning use artificial stimulus-response mappings. For example, Krigolson et al. (2014) had participants determine which of two coloured squares were rewarding. Although they showed that participants could learn this, and that learning was reflected by the reward positivity, this scenario is unlike anything we encounter in everyday life. As predicted, our findings are congruent with this and other studies, supporting the account that the reward positivity reflects a generic learning system responsible for feedback learning in all contexts – confirming a long-standing assumption of the generalized function underlying the reward positivity and the ACC (e.g., Holroyd and Coles, 2002). Further, this also elucidates to the fact that this associative learning system continues to drive learning across simple and complex material. With this knowledge, future educational interventions can focus more heavily on reinforcement learning techniques, even when regarding difficult material.

In sum, we presented evidence that a reinforcement learning paradigm is effective when learning complex material. Further, we supported claims that the reward positivity reflects reward learning systems as its amplitude diminished with learning. Lastly, our paradigm allowed us to generalize what is known about this component to real world contexts. Thus, the reward positivity reflects a generic learning system.

References

Anderson, S.J., Krigolson, O.E., Jamniczky, H.A., Hecker, K.G., 2016. Learning anatomical structures: a reinforcement-based learning approach. Med. Sci. Educ. 26 (1), 123–128.

Bannister, S.L., Au, H., Forbes, K.L., Zucker, M., Weiler, G., et al., 2016. A Card Is Worth a Thousand Questions: Leveraging the Power of Key Features and Semantic Qualifiers in Writing Clinical Questions. Council on Medical Student Education in Pediatrics, St. Louis, MO (2016 Annual Meeting).

Bellebaum, C., Colosio, M., 2014. From feedback-to response-based performance monitoring in active and observational learning. J. Cogn. Neurosci. 26 (9), 2111–2127.

Bellebaum, C., Daum, I., 2008. Learning-related changes in reward expectancy are reflected in the feedback-related negativity. Eur. J. Neurosci. 27 (7), 1823–1835.

Bellebaum, C., Kobza, S., Thiele, S., Daum, I., 2010. It was not MY fault: event-related brain potentials in active and observational learning from feedback. Cereb. Cortex 20, 2874–2883.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10 (4), 433–436.

Burak, K.W., McLaughlin, K.J., Coderre, S.P., Busche, K.D., Raman, M., 2015. The Flipped Classroom Improves Exam Performance in Undergraduate Medical Education. Presented at the Office of Health and Medical Education Scholarship Symposium, Calgary, AB.

Coderre, S., Mandin, H., Harasym, P., Fick, G., 2003. The effect of diagnostic reasoning on diagnostic success. Med. Educ. 37, 695–703.

Eppinger, B., Kray, J., Mock, B., Mecklinger, A., 2008. Better or worse than expected? Aging, learning, and the ERN. Neuropsychologia 46 (2), 521–539.

Eppinger, B., Mock, B., Kray, J., 2009. Developmental differences in learning and error processing: evidence from ERPs. Psychophysiology 46, 1043–1053.

Foti, D., Weinberg, A., Dien, J., Hajcak, G., 2011. Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: temporospatial principal components analysis and source localization of the feedback negativity. Hum. Brain Mapp. 32 (12), 2207–2216.

Gershman, S.J., Niv, Y., 2010. Learning latent structure: carving nature at its joints. Curr. Opin. Neurobiol. 20 (2), 251–256.

Groen, Y., Wijers, A.A., Mulder, L.J.M., Minderaa, R.B., Althaus, M., 2007. Physiological correlates of learning by performance feedback in children: a study of EEG event-related potentials and evoked heart rate. Biol. Psychol. 76, 174–187.

Hämmerer, D., Li, S.C., Müller, V., Lindenberger, U., 2010. Life span differences in electrophysiological correlates of monitoring gains and losses during probabilistic reinforcement learning. J. Cogn. Neurosci. 23, 579–592.

Holroyd, C.B., Coles, M.G., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109 (4), 679–709.

Holroyd, C.B., Krigolson, O.E., 2007. Reward prediction error signals associated with a modified time estimation task. Psychophysiology 44 (6), 913–917.

Holroyd, C.B., McClure, S.M., 2015. Hierarchical control over effortful behavior by rodent medial frontal cortex: a computational model. Psychol. Rev. 122 (1), 54–83.

Holroyd, C.B., Yeung, N., 2012. Motivation of extended behaviors by anterior cingulate cortex. Trends Cogn. Sci. 16 (2), 122–128.

Holroyd, C.B., Pakzad-Vaezi, K.L., Krigolson, O.E., 2008. The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. Psychophysiology 45 (5), 688–697.

Horrey, K., Keegan, D., Paget, M.K., Tan, A., 2016. Creating open access FM micro-cases for online medical student learning. In: 2016 Conference on Medical Student Education. The Society of Teachers of Family Medicine, Phoenix, AZ.

Kazoleas, K., 2016. Med students practice clinical problem solving by playing cards. In: UCalgary Medicine Magazine Winter. 13 Retrieved from. http://cumming.ucalgary.ca/w2016-playing-cards.

Krigolson, O.E., Pierce, L., Tanaka, J., Holroyd, C.B., 2009. Learning to become an expert: reinforcement learning and the acquisition of perceptual expertise. J. Cogn. Neurosci. 21 (9), 1834–1841.

Krigolson, O.E., Hassall, C., Handy, T.C., 2014. How we learn to make decisions: the rapid propagation of reinforcement learning prediction errors in humans. J. Cogn. Neurosci. 26 (3), 635–644.

Liu, Y., Fu, X., 2007. How does distraction task influence the interaction of working memory and long-term memory? In: Harris, D. (Ed.), Engineering Psychology and Cognitive Ergonomics. Springer, Berlin, pp. 366–374.

Luck, S.J., 2014. An Introduction to the Event-related Potential Technique, 2nd ed. MIT Press, Cambridge, MA.

Luque, D., López, F.J., Marco-Pallarés, J., Càmara, E., Rodríguez-Fornells, A., 2012. Feedback-related brain potential activity complies with basic assumptions of associative learning theory. J. Cogn. Neurosci. 24, 794–808.

Miltner, W.H., Braun, C.H., Coles, M.G., 1997. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. J. Cogn. Neurosci. 9 (6), 788–798.

Minear, M., Park, D.C., 2004. A lifespan database of adult facial stimuli. Behav. Res. Methods Instrum. Comput. 36, 630–633.

Nieuwenhuis, S., Ridderinkhof, K.R., Talsma, D., Coles, M.G., Holroyd, C.B., Kok, A., et al., 2002. A computational account of altered error processing in older age: dopamine and the error-related negativity. Cogn. Affect. Behav. Neurosci. 2 (1), 19–36.

Niv, Y., Daniel, R., Geana, A., Gershman, S.J., Leong, Y.C., et al., 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. J. Neurosci. 35 (21), 8145–8157.

Proudfit, G.H., 2015. The reward positivity: from basic research on reward to a biomarker for depression. Psychophysiology 52 (4), 449–459.

Sailer, U., Fischmeister, F.P.S., Bauer, H., 2010. Effects of learning on feedback-related brain potentials in a decision-making task. Brain Res. 1342, 85–93.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., et al., 2005. Effects of value and reward magnitude on feedback negativity and P300. Neuroreport 16 (4), 407–411.

Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and

reward. Science 275 (5306), 1593–1599.
Skinner, B.F., 1958. Reinforcement today. Am. Psychol. 13 (3), 94–99.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
Tang, A., Kwan, E., Paget, M., Coderre, S., Burak, K., et al., 2016. Cards: A Novel, Case-based Method for Undergraduate Medical Students to Learn Key Concepts in Geriatrics. Presented at the 36th Annual Scientific Meeting of the Canadian Geriatrics Society. (Vancouver, BC).
Walsh, M.M., Anderson, J.R., 2012. Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neurosci. Biobehav. Rev. 36 (8), 1870–1884.
Wilson, R.C., Niv, Y., 2012. Inferring relevance in a changing world. Front. Hum. Neurosci. 5 (189), 1–14.
Xu, B., Rourke, L., Robinson, J.K., Tanaka, J.W., 2016. Training melanoma detection in photographs using the perceptual expertise training approach. Appl. Cogn. Psychol. 30 (5), 750–756.
Yeung, N., Sanfey, A.G., 2004. Independent coding of reward magnitude and valence in the human brain. J. Neurosci. 24 (28), 6258–6264.