Advances in Health Sciences Education (2023) 28:1579–1592
https://doi.org/10.1007/s10459-023-10232-z

RESEARCH

Pathologists aren't pigeons: exploring the neural basis of visual recognition and perceptual expertise in pathology

Sarah J. Anderson[1,2] · Amy L. Warren[1] · Nia Abdullayeva[3] · Olav Krigolson[4] · Kent G. Hecker[1,5]

Received: 27 August 2022 / Accepted: 21 April 2023 / Published online: 15 May 2023

Abstract

Visual (perceptual) reasoning is a critical skill in many medical specialties, including pathology, diagnostic imaging, and dermatology. However, in an ever-compressed medical curriculum, learning and practicing this skill can be challenging. Previous studies (including work with pigeons) have suggested that using reward-feedback-based activities, novices can gain expert levels of visual diagnostic accuracy in shortened training times. But is this level of diagnostic accuracy a result of image recognition (categorization) or is it the acquisition of diagnostic expertise? To answer this, the authors measured electroencephalographic data (EEG) and two components of the human event-related brain potential (reward positivity and N170) to explore the nature of visual expertise in a novice-expert study in pathology visual diagnosis. It was found that the amplitude of the reward positivity decreased with learning in novices (suggesting a decrease in reliance on feedback, as in other studies). However, this signal remained significantly different from the experts whose reward positivity signal did not change over the course of the experiment. There were no changes in the amplitude of the N170 (a reported neural marker of visual expertise) in novices over time. Novice N170 signals remained statistically and significantly lower in amplitude compared to experts throughout task performance. These data suggest that, while novices gained the ability to recognize (categorize) pathologies through reinforcement learning as quantified by the change in reward positivity, increased accuracy, and decreased time for responses, there was little change in the neural marker associated with visual expertise (N170). This is consistent with the multi-dimensional and complex nature of visual expertise and provides insight into future training programs for novices to bridge the expertise gap.

Keywords Visual reasoning · Pathology · Electroencephalography · Neuroeducation · Feedback/reinforcement based learning · N170 · Reward positivity

Sarah J. Anderson and Amy L. Warren: Co-first authors

Extended author information available on the last page of the article

Introduction

With competing and ongoing demands in health professions curricula, the time afforded to students to develop visual reasoning skills is increasingly compressed (Rockarts et al., 2020). This has prompted work exploring the use of reinforcement-based learning methods where students iteratively learn to identify anatomic structures or formulate diagnoses from clinical data, through the repetitive presentation of data, selecting an appropriate answer from a list, and providing immediate feedback (Anderson et al., 2018, 2019; Williams et al., 2017). There has even been work with pigeons that uses reinforcement-learning to train pigeons to distinguish between benign and malignant breast masses through histopathology and diagnostic imaging (Levenson et al., 2015). These studies demonstrated that the time to respond and accuracy of novices (and pigeons) reached similar levels to experts in relatively short periods of time, something that may otherwise take years to develop. However, this is based on assumptions that perceptual recognition is acquired through reinforcement learning, that it equates to perceptual expertise, that it does not account for other aspects of experience, knowledge, context, and metacognition that are commonly cited components of clinical reasoning expertise (Jensen et al., 2008). Put another way, while pigeons and novices may become proficient in image recognition in reinforcement-learning activities, pathologists are not pigeons, and perceptual expertise is multifaceted that involves other components, including but not limited to deliberate practice, clinical exposure, learning from misdiagnoses, and further learning (Ericsson & Harwell, 2019; Pusic et al., 2011). This statement appears intuitive, almost laughable, but what evidence, specifically biological evidence, exists to support this assumption?

To address this in a medical training context, we studied visual/perceptual recognition and expertise within a reinforcement-based learning paradigm in both novice and expert populations using behavioural metrics–response time and accuracy–and electroencephalographic (EEG) event-related potentials (ERPs; measured through EEG). Specifically, we sought to examine the relationship between the reward positivity (recognition) and N170 (visual expertise) ERP components in both novices and experts to identify if there is a difference between image recognition and visual expertise in pathology.

ERPs are evoked responses to visual, auditory, or somatosensory stimuli and are posited to reflect various cognitive processes (Luck, 2014). Here, we are interested in magnitude of neural response (either positive [*P*] or negative [*N*] measured in microvolts, *Y*-axis of a graph) and time of response (milliseconds, *X*-axis of a graph) to the stimuli. For example, two ERP components associated with neural leaning and decision making have informed our understanding of expertise: the visual N170 ERP component and the feedback evoked reward positivity ERP component (Anderson et al., 2018; Krigolson et al., 2009; Krigolson et al., 2013). The N170 is an ERP component related to expert object recognition that occurs approximately 170 ms after visual stimulus, onset over the occipitotemporal regions (Scott, 2011; Tanaka & Curran, 2001) where larger amplitudes (peaks) are associated with perceptual expertise. This ERP component has also been identified as a marker of human visual expertise in several image-based studies including identifying dogs, birds, and cars (Scott, 2011; Scott et al., 2006). In a health profession context, Rourke et al., demonstrated increased N170 amplitude in experts when interpreting visual data specific to their expertise (in this case using chest radiographs and electrocardiogram (EKG) tracings) (Rourke et al., 2016). Considering the

large repertoire of pathology image characteristics that pathologists are trained to recognize and categorize during their training, it is likely that pathologists will have similar kinds of visual processing expertise that evoke similar N170 responses.

The reward positivity ERP component has been used in other domains to track visual learning in novices and learning in general as it is produced by external feedback provided during trial-and-error reinforcement learning and is measured over the medial frontal cortex (Krigolson et al., 2013). In a medical training context, reinforcement learning can occur after a decision (for example, a diagnosis) is followed by feedback as to whether the decision is correct or incorrect (Anderson et al., 2018). For instance, in previous work in medical education, reinforcement learning has been used to teach students neuroanatomy (Anderson et al., 2018, 2019). In the study by Anderson et al. (2018), students were provided with a neuro-anatomical image and a correct or incorrect label that they then needed to indicate whether there was a match between image and label, after which, feedback was given as to whether they were able to correctly match image and label (Anderson et al., 2018). EEG tracing from such experiments evokes a reward positivity ERP component (a neural correlate of a reinforcement learning prediction error) that is initially large as learners rely on the outcome of the feedback and decreases as students become more proficient in visual (or diagnostic) categorization (Anderson et al., 2018; Williams et al., 2017).

Using a reinforcement-learning paradigm, we hypothesized that experts would not be particularly reliant on feedback to correct or corroborate their results (and thus we would see a reduced reward positivity) while novices would rely more on feedback initially (and have a larger amplitude reward positivity) and this reliance would decrease over time. In other words, we would expect that, with learning in a visual categorization task, initial differences in reward positivity amplitude would converge by the end of the reinforcement-based learning task, indicating that novices had become proficient at the visual categorization/recognition task. Indeed, our own studies in neuroanatomy and medical reasoning learning have shown similar changes in reward positivity amplitude in novices (Anderson et al., 2019; Williams et al., 2017).

Concurrently, we hypothesized that N170 amplitude, a marker of visual expertise, would be significantly lower for novices compared to experts over the entire course of the learning paradigm, suggesting that recognition (marked by changes in reward positivity) does not equate to expertise. Previous studies suggest that N170 does not change within novice participants over learning task time (Anderson et al., 2018) and N170 amplitude in the domain of expertise remains constant over the course of a visual categorization task (Rourke et al., 2016; Tanaka & Curran, 2001). In summary, based on previous findings in the perceptual expertise domain, we were interested in whether we could tease out the components of visual expertise by combining the use of reinforcement learning paradigm (training to recognize objects) with novices and experts in pathology.

Materials and methods

Participants

Participants were recruited by email from the Faculty of Veterinary Medicine, University of Calgary (experts) and the Bachelors of Health Science, Bachelor of Science (Neurosciences program) and Doctor of Veterinary Medicine (first-year students only) programs at the University of Calgary (novices). Participation was voluntary and informed consent was obtained

per the Declaration of Helsinki. All experimentation was approved by the University of Calgary Conjoint Health Ethics Board (Ethics ID: REB 16–0925).

Experimental procedure

The experimental paradigm we employed was adapted from Krigolson et al. (2009) where they used a reinforcement based task in order to teach categorization of blobs, and Anderson et al. (2018) were they used a reinforcement based task to teach identification of neuroanatomical structures. Modifications are outlined below.

In our study, participants were shown images of bovine hepatic pathology with a paired incorrect or correct diagnosis and asked to indicate if this pairing was coherent or not while their brain activity was recorded using electroencephalographic (EEG) recording equipment. For each bovine pathology image, four incorrect diagnoses were identified that had a similar visual appearance to the diagnosis (e.g., the diagnosis of hepatic lipidosis was paired with hepatic lymphoma, centrilobular necrosis, periportal fibrosis, and normal). For each image, the correct label was shown 50% of the time. Participants were shown 288 paired bovine pathology image-diagnosis (correct or incorrect) questions across twelve blocks. For each of the 24 pathology image-diagnosis questions in a block, a sequence of four steps were used–a white fixation cross in the center of the screen (for 500 ms), the image of one of the twelve bovine hepatic pathologies (1500 ms), the same image with an added label of a correct or incorrect diagnosis (maximum 2000 ms) [– after which a respondent had to indicate if the image and diagnosis were correctly (yes) or incorrectly (no) matched. Finally, after selecting "yes" or "no", participants were given feedback as to whether their answer was correct (✓), incorrect (*x*) or if they took too long to answer (!). Between each of the twelve blocks, participants were free to take an optional break for as long as they needed to mitigate fatigue.

Bovine hepatic pathology cases

Twelve images of bovine liver pathologies were sourced from the Charles Louis Davis and Samuel Wesley Thompson DVM Foundation Noah's Arkive collection of veterinary pathology gross images (https://noahsarkive.cldavis.org/) and the Cornell University Dr. John M. King's Necropsy Show and Tell (https://secure.vet.cornell.edu/nst/nst/asp.) open-source pathology image collections (David Thompson Foundation, 2021; King et al., 2021). Images were modified to remove identifying markers, to ensure they were of similar magnification, and to make all of the backgrounds consistent (black). The twelve images represented common bovine liver pathologies consisting of the normal hepatic parenchyma, infectious necrotic hepatitis (black disease; *Clostridium novyi*), acute hepatic abscess (*Trueperella pyogenes*), chronic hepatic abscess (also *Trueperella pyogenes* infection with significant peri-abscess fibrosis), hepatic necrobacilliosis (*Fusobacterium necrophorum*), Fascioliasis (*Fasciola hepatica*), hepatic granulomas (*Mycobacterium bovis*), centrilobular necrosis, mild periportal fibrosis, severe periportal fibrosis and nodular hyperplasia (cirrhosis), hepatic lymphoma, and hepatic lipidosis.

EEG data acquisition

After filling out a pre-experiment survey that included questions on demographics, training, and prior experience, participants were fitted with an EEG cap and the EEG system

was setup. Participants were seated in an electromagnetically shielded, sound-proofed room in front of a 17″ ASUS laptop computer for the reinforcement-based learning task. MATLAB (Release 2018a, Math Works, Natick, MA, USA) was used with a script modified from Anderson et al. (2018) in conjunction with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997) to present the questions. For the EEG, an appropriately sized thirty-two electrode cap (acti CAP slim, Brain Products, GmbH, Munich, Germany; FP1, Fp2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, O1, Oz, O2, FCz:) was fitted using in a standard 10–20 layout (Luck, 2014). Brain Vision Recorder Software (Version 1.21, Brain Products, Munich, Germany) was used to acquire EEG data. Impedances for all electrodes were below 20 kΩ. Data were sampled at a rate of 500 Hz and a LiveAmp (Version 4.60, Brain Products, GmbH, Munich, Germany) was used to amplify data which was connected to the data acquisition computer via a Sensor and Trigger extension (Version 1.14, Brain Products, GmbH, Munich, Germany).

Behavioural data acquisition and analysis

Behavioural data of response accuracy and reaction time (in ms) were obtained for each participant using MATLAB. When the participant's response time was > 2000 ms, the answer was coded as incorrect. Accuracy was determined by the ability of the participant to correctly identify whether each bovine pathology image and its associated diagnosis were coherent or incoherent. Accuracy and reaction time was recorded for each pathology image-diagnosis question, and the mean accuracy (%) and mean reaction time (in ms) were calculated across the 24 questions for each of the twelve blocks.

EEG data processing and analysis

For further information regarding EEG processing please refer to Luck (2014). Brain Vision Analyzer 2 software (Version 2.1, Brain Products, GmbH, Munich, Germany) was used to analyze EEG data. Data were down-sampled to 250 Hz (for data size management), re-referenced to linked TP9 and TP10 electrodes (to allow for comparisons amongst electrodes), and filtered using a phase-shift Butterworth filter with 0.1–30 Hz passband and 60 Hz notch filter (to remove artifactual data like electrical noise). Independent component analysis was performed to correct known ocular artifacts (blinks and horizontal eye movements) from the continuous data.

Discrete ERP waveforms representing responses to stimuli (epochs) are calculated by averaging segments of the EEG signal time locked to the onset of the stimuli of interest (for example, pathology images or positive feedback). Epochs were baseline corrected using the mean voltage 200 ms preceding the event of interest. Epochs containing non-ocular artifacts were removed using a 10 μV/ms gradient and 100 μV absolute difference criteria. If greater than 20% of the data for a participant was removed due to artifacts, the participant was removed from the study.

Epochs for each ERP component of interest were performed as described in Anderson et al. (2018). This included the appearance of the pathology image (to time lock the N170 response to identify object recognition activity indicating visual expertise) and the provision of correct and incorrect feedback (to time lock the reward positivity).

The N170 ERP component was observed at the O1 and O2 electrode sites. Here, the mean voltage activity was extracted from the overall ERP waveforms (epoch) by

identifying the first and second positive going peaks in the waveform following the onset of the pathology image, identifying the negative going peak between these positive peaks, and extracting mean voltage from this negative peak; this process is similar to previous literature (Tanaka & Curran, 2001; Scott, 2006, Scott et al., 2008). This N170 peak was observed at 148 ms and a ± 10 ms window of mean voltage activity surrounding this peak was extracted for both novice and expert participants for each block.

The reward positivity ERP component was observed at the FCz electrode site. Here, similar to the N170, mean voltages were extracted from the negative peak bounded by the two positive peaks following stimulus, however this epoch was time-locked to the provision of feedback. The identification of the general timing of this peak is confirmed through a difference wave analysis where ERP waveforms in response to correct vs. incorrect feedback are compared to confirm differences exist. Next, since the reward positivity ERP component is thought to be a modulator of this negative peak, when the reward positivity is large, the amplitude of this negative peak will be dampened. Similarly, when reward positivity decreases we expect to see an increase in the amplitude of this negative peak. Across the blocks, the reward positivity was measured at this negative peak ranging from 244 to 308 ms in experts and 264–300 ms in novices. A ± 5 ms window of mean voltage activity surrounding this peak was extracted for both novice and expert participants for each block.

Statistical data analysis

Statistical analyses of the behavioral data (response time and accuracy) and neural data were performed using SPSS (Version 28) and Excel (Version 2002). Changes in the behavioural data (accuracy and reaction time) and the brain activity data (reward positivity and N170 waveforms) were examined over the 12 blocks as outlined above of the experiment and analyzed using repeated-measures analysis of variance (RM-ANOVA) with post hoc analysis to identify specific differences between novice and expert groups within and between blocks. As assumptions of statistical sphericity were violated for these analyses, a Greenhouse-Geiser correction was applied to adjust the degrees of freedom. An alpha level of 0.05 was assumed for all statistical tests.

Results

Participants

Twenty-six undergraduate students (novices) were recruited. The novices had no prior experience with bovine pathology. In the novices, twenty-three (88.5%) were female and three were male (11.5%); there was an age range of 18–24 years with a mean age of 19.8 years (SD = 1.50). Experts all had a DVM (or DVM equivalent [BVSc]) and post-DVM training in pathology (five were board-certified by the American College of Veterinary Pathologists (ACVP) in anatomic pathology, two were board-certified by the ACVP in clinical pathology, one had a Masters of Veterinary Studies in clinical pathology, and one was an intern in veterinary anatomic pathology at the University of Calgary). This constituted the population of veterinary pathologists in Calgary, Alberta, Canada. There were six female (66.7%) and three male (33.3%) experts. There was an age range in experts of 30–76 years with a mean age of 44.1 years (SD = 13.66). Ratios of women and men in both novice and expert groups are consistent with current

demographics in veterinary training programs and veterinary medicine. Data from one expert was removed due to issues during data collection. One novice participant was excluded from the data analysis due to an artifact rejection rate greater than 20% from the N170 ERP component. Therefore, there were data from 25 novices and 8 experts in the final analyses.

All participants in the novice and expert groups had normal or corrected to normal vision and no declared neurological impairments.

Behavioural data

Accuracy

Novices and expert percent accuracy by experiment block are shown in Fig. 1a. Novices started with near 50% accuracy (guessing) in the first block of the task ($M = 52.83\%$, SD = 10.19) and progressed to near expert levels of accuracy by the final block ($M = 97.00\%$, SD = 4.58). Experts began with 68.23% accuracy (SD = 20.16) in block 1 and 91.67% accuracy (SD = 4.98) by block 4. There was a significant main effect on accuracy between novices and experts (F (1,31) = 9.01, $p = 0.005$) and post hoc tests identified significant differences between novices and experts for blocks 1, 2, 4, 5, 7, and 8 ($p < 0.05$).

Reaction time

Reaction times for both novices and experts decreased throughout the course of the experiment (Fig. 1b), beginning with a mean reaction time of 1.13 s (SD = 0.23 s) for novices and 1.40 s (SD = 0.18 s) for experts in the first block, and reducing to 0.63 s (SD = 0.15 s) and 0.72 s (SD = 0.11 s) by the twelfth block. While there was no significant main effect in reaction times between novices and experts (F (1, 31) = 2.44, $p = 0.129$), post hoc tests identified that experts took significantly longer for block 1.

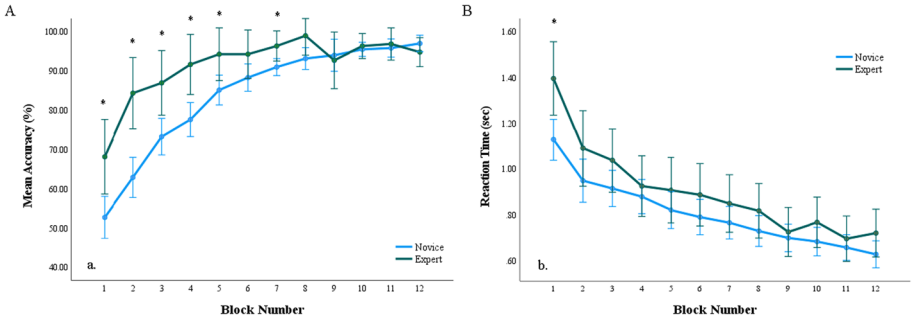

Fig. 1 Mean accuracy (**a**. percent correct) and reaction time (**b**. seconds) of novices (26) and experts (8) participants as they progress through 12 blocks of the reinforcement-learning exercise. Error bars represent the 95% confidence intervals, asterisk represent significant differences between novice and experts

EEG data

The reward positivity

The reward positivity ERP component difference wave analyses comparing neural responses to correct vs. incorrect feedback for novices and experts are shown in Fig. 2. This component was measured maximally at the FCz electrode located over the medial frontal cortex. There was a significant main effect between novices and experts ($F(1,31) = 14.23$, $p < 0.001$) in reward positivity amplitude measured at FCz. Post hoc analyses revealed that these differences were present across the majority of blocks, the only non-significant difference between novice and expert occurred in Block 11 (Fig. 3). There were significant changes in reward positivity amplitude over blocks within groups ($F(11, 341) = 8.43$, $p < 0.001$). Post hoc analyses showed differences over blocks within novices but not within experts. Significant differences were found between Blocks 1–5 (as novices relied more heavily on feedback) and 6 through 12 (where novices did not rely as heavily on feedback). Differences between novices and experts persisted until Block 12, meaning that while novices relied less on feedback in the later blocks and their reward positivity amplitude plateaued, there were still significant differences in reward positivity amplitude between the two groups.

The N170

The N170 ERP component measured for both novices and experts at O1 and O2 is shown in Fig. 4B, C. The scalp distribution map shows greater negativity for experts vs. novices in the occipital region at 148 ms (± 10 ms) post image presentation (Fig. 4A). There were significant main effect differences between novices and experts at both O1 ($F(1,31) = 7.59$,

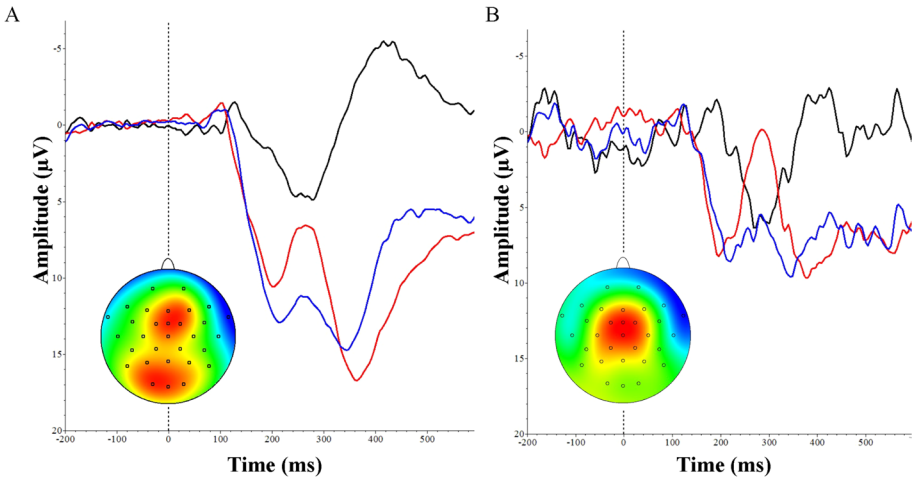

Fig. 2 Mean scalp potential (µV) at the FCz electrode in response to correct (blue) and incorrect (red) feedback; difference waves shown in black. **A** Novices ($n = 25$). **B** Experts ($n = 8$). These event-related potential is time locked to feedback onset (dashed line) and the extracted peak occurred at approximately 275 ms after feedback onset. Scalp distributions show a positivity (incorrect–correct feedback responses) occurring in the medial frontal region for both novices and experts. Negative is plotted up

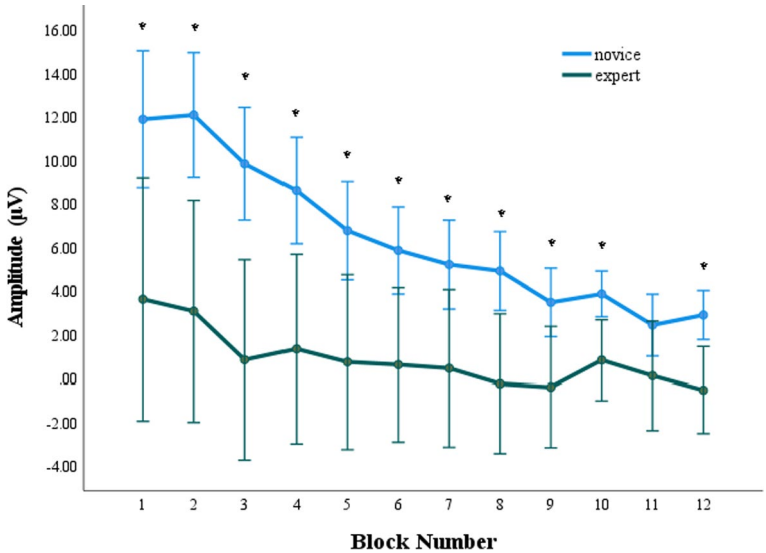

Fig. 3 Reward positivity mean amplitude measured at the FCz electrode site. Error bars represent the 95% confidence intervals, asterisk represent significant differences between novice ($n=25$) and experts ($n=8$)

$p=0.01$) and O2 ($F(1,31)=8.15$, $p=0.008$) electrode positions indicating that novices had significantly smaller N170 amplitudes than experts. Post hoc analyses revealed that these differences were present across most blocks in O1, with the only non-significant blocks, block 7 and 10. All blocks in O2 were significantly different (Fig. 4D, E). There were no significant changes in N170 amplitude over blocks within groups for novices or exerts for O1 or O2 (O1: $F(11,341)=0.725$, $p=0.715$; O2: $F(11,341)=0.688$, $p=0.750$).

Discussion

In this paper, we demonstrated novice and expert differences in neural signatures in a reinforcement-based learning task that gives insight into both perceptual learning and expertise. Our data suggest that while we found that novices can attain some level of expertise via reinforcement learning (specifically image recognition and the reward positivity ERPs), we still saw different neural responses within experts (N170) suggesting that the visual learning gained through the reinforcement-based learning task reflects only one dimension of perceptual expertise, or potentially that more training and deliberate practice was needed (Ericsson & Harwell, 2019).

We have demonstrated that, like other studies (Anderson et al., 2018; Krigolson et al., 2009; Levenson et al., 2015; Williams et al., 2017), during a reinforcement-based learning task (in this case using veterinary pathology images), novices can respond as quickly as experts while still maintaining diagnostic accuracy. Taken alone, this could suggest that reinforcement-based learning was successful in creating perceptual expertise in our novices. There were some unexpected findings in the behavioural data though. Experts took significantly longer in the first block than novices, and their initial expert diagnostic accuracy was lower than expected (68%, improving over the 12 blocks). The mean age of experts compared to novices (44 years vs. 20 years) and inexperience with gaming

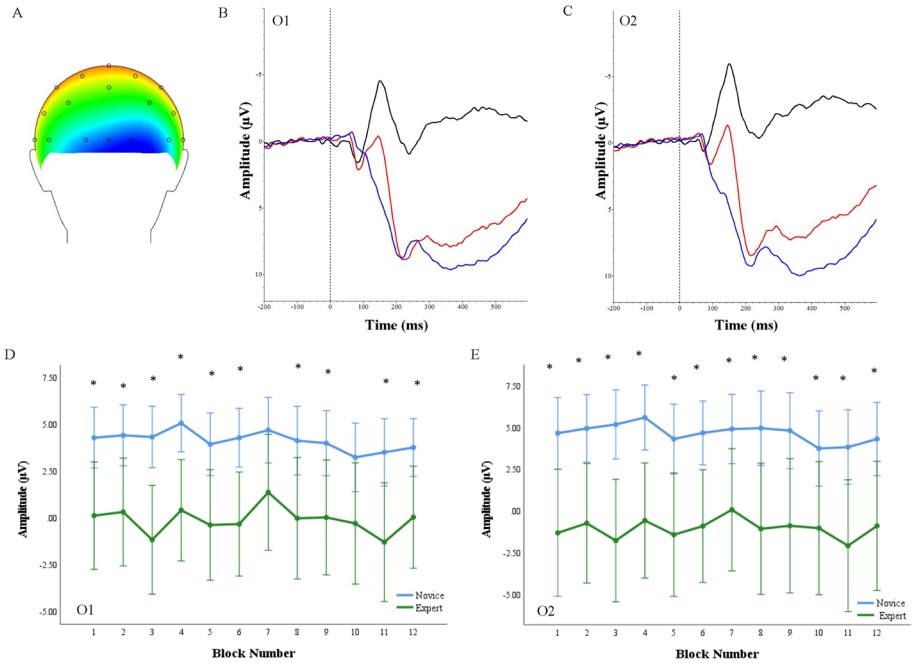

Fig. 4 **A** Scalp distribution map showing difference (novice-expert), indicating that experts have greater negativity compared to novices in the occipital region at 148 ms (±10 ms). **B** Mean amplitude (µV) at the O1 electrode time locked to image onset (dashed line) for all questions for novices (blue, $n=25$) and experts (red, $n=8$). Difference wave shown in black. C) Mean amplitude (µV) at the O2 electrode time locked to image onset (dashed line) for all questions for novices (blue, $n=25$) and experts (red, $n=8$). Difference wave shown in black. D) Mean amplitude at O1. Error bars represent the 95% confidence intervals, asterisk represent significant differences between novice and experts. E) Mean amplitude at O2. Error bars represent the 95% confidence intervals, asterisk represent significant differences between novice and experts

equipment particularly for older experts, may explain some differences in response time (Thompson et al., 2014). Post-experiment questioning from experts also revealed some expert disagreement on diagnosis, particularly for images with very similar visual pathologies (e.g., hepatic lipidosis vs. hepatic lymphoma), which may also account for initial lower diagnostic accuracy in the expert group. Finally, the format in which images were presented differs from how experts would likely encounter the lesions in an animal, which may also contribute to potential differences in diagnoses. However, general response time and diagnostic accuracy results were as expected with steep improvements in novice accuracy and time to diagnosis over the course of the reinforcement-based learning exercise.

Our analysis of the reward positivity for pathology image recognition in novices also supported perceptual learning over the course of the training. As in other studies, reward positivity amplitude in novices declined initially, then plateaued. In previous experiments, this was interpreted as gaining of expertise and demonstration of learning, akin to gaining of expertise (Anderson et al., 2018; Krigolson et al., 2009; Williams et al., 2017).

A notable novel contribution from this study, was our comparison of differences and changes in the reward positivity between novices and experts in the same learning task. Differences between novices and experts persisted through Block 12 (Fig. 3), meaning that while novices relied less on feedback in the later blocks and their reward positivity

amplitude plateaued, there were still differences in reward positivity amplitude between the two groups. Expert reward positivity amplitudes remained low and did not change over the course of the training, suggesting that as a result of their expertise they were much less reliant on feedback.

In our N170 data, novices and experts were both static in their amplitude with significantly higher amplitudes observed in experts that persisted over the course of the exercise. Given that the N170 has been shown to be a marker of perceptual expertise, with expertise associated with higher N170 amplitudes (Rourke et al., 2016; Tanaka & Curran, 2001), our data supports this in pathology image expertise too. The fact that N170 amplitude did not change for novices over the course of learning, and remained significantly smaller than experts, indicates a missing component of expertise not attained in reinforcement-based learning paradigms without sufficient practice–pathologists are not pigeons.

Taken together, the significantly higher reward positivity in novices, and the unchanging, smaller N170 amplitude, suggests that novices, while improving performance and reducing reliance on feedback, do not develop the same neural component for visual expertise within this learning paradigm (or at least with the amount of practice they were exposed to here). This is important to tease apart as training may allow for recognition and assist with initial training within novices, however, it may not facilitate expertise. So, while reinforcement-based learning is important for the acquisition of recognition there is a limitation in claiming the acquisition of expertise. What is interesting to explore further is the acquisition of expertise over time and what teaching/learning interventions could affect the change in N170 (knowledge, exposure, time etc.).

The findings of this work have important implications to both our understanding of expertise as a multi-faceted construct and in designing training programs in perceptual expertise. First, is that the study of expertise and expertise development, or for that matter learning, can be further explored using brain imaging-based methods, this is the focus of the emerging field of 'neuroeducation'. Clearly, there is much work to be done in terms of what research questions should be explored, what theories should be tested, what experimental designs should be employed, which imaging modalities should be used, which analytical methods could be called upon to analyze and interpret the results, and how these results can be incorporated (if at all) in our educational training programs. Studies in clinical reasoning (e.g., Durning et al., 2015; Hruska et al., 2016a, 2016b; Williams et al., 2017), visual expertise (e.g., this study, Biliac et al., 2016; Rourke et al, 2016), acquisition of psychomotor skills (e.g., Crewther et al., 2016; Toy et al., 2023) as well as use of neurostimulation (e.g., Ciechanski et al., 2019; Gao et al., 2021) have been reported and are challenging some of our current assumption of expertise and educational practice. What remains to be determined are the claims regarding the science and the influence it may or may not have on our educational practices. How can we use our understanding of the neural markers of learning/expertise/decision making as a mechanism to inform or optimize educational approaches in our training programs?

Second, much visual learning is unstructured, either relying on students being shown visual exemplars of disease in static images or fixed specimens in laboratories (Warren & Donnon, 2013). We previously have used reinforcement-based learning as a just-in-time teaching tool where students use reinforcement-learning as a pre-class "warm-up" learning activity (Anderson et al., 2018). Just-in-time teaching incorporates pre-class activities to establish foundational knowledge that then allows instructors to focus on more complex concepts and higher levels of learning during class time (Marrs & Novak, 2004; Novak, 2011). In the context of pathology, providing students pre-class with a reinforcement-learning activity (just-in-time teaching) to learn the visual representation of disease processes,

would allow class time to be more efficiently used to contextualize those disease patterns in terms of diagnosis, disease pathogenesis, treatment options and disease outcomes, foster increased active learning, and promote deeper learning (Novak, 2011)–perhaps providing the 'secret sauce' of perceptual expertise that separated the novices and experts in our study.

While this work provides evidence for the use of reinforcement-based tasks for internalizing recognition, we would caution against the interpretation that this equates to expertise. Comparing these results to the original Tanaka and Curran paper (2001) and the later Rourke et al. paper (2016), there is the potential for the N170 to be used as a neural marker of expertise within health professions, but how it is developed and potential for interventions through both behavioural based and neural-based studies requires sustained efforts to determine where how best and where to use these neural markers and provide a basis as to what can be inferred from these studies (Rourke et al., 2016; Tanaka & Curran, 2001).

Acknowledgements The authors would like to thank Dr. Martin Pusic and the editors for their thoughtful comments on this manuscript. This work was funded by KGH's Canada Foundation for Innovation infrastructure grant.

Author contributions A.W., S.A. and K.G.H. wrote the manuscript text, S.A. analyzed the neural data and prepared the figures, K.G. H. analyzed the behavioral data. All authors assisted with experimental design and all reviewed the manuscript.

Declarations

Competing interests K.G.H. is also the Chief Assessment Officer for the International Council for Veterinary Medicine.

References

Anderson, S. J., Hecker, K. G., Krigolson, O. E., & Jamniczky, H. A. (2018). A reinforcement-based learning paradigm increases anatomical learning and retention-a neuroeducation study. *Frontiers Human Neuroscience, 12*, 38. https://doi.org/10.3389/fnhum.2018.00038

Anderson, S. J., Jamniczky, H. A., Krigolson, O. E., Coderre, S. P., & Hecker, K. G. (2019). Quantifying two-dimensional and three-dimensional stereoscopic learning in anatomy using electroencephalography. *Nature Partner Journal: Science of Learning, 4*, 10. https://doi.org/10.1038/s41539-019-0050-4

Bilalic, M., Grottenthaler, T., Nagele, T., & Lindig, T. (2016). The faces in radiological images: Fusiform face area supports radiological expertise. *Cerebral Cortex, 26*(3), 1004–1014.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Ciechanski, P., Kirton, A., Wilson, B., Williams, C. C., Anderson, S. J., Cheng, A., Lopushinsky, S., & Hecker, K. G. (2019). Electroencephalography correlates of transcranial direct-current stimulation enhanced surgical skill learning: A replication and extension study. *Brain Research.* https://doi.org/10.1016/j.brainres.2019.146445

Crewther, B. T., Shetty, K., Jarchi, D., Selvadurai, S., Cook, C. J., Leff, D. R., Darzi, A., & Yang, G. Z. (2016). Skill acquisition and stress adaptations following laparoscopic surgery training and detraining in novice surgeons. *Surgical Endoscopy, 30*(7), 2961–2968.

David Thompson Foundation (2021). *Noah's Arkive: Davis-Thompson Foundation*. Retrieved November 30, 2021, from: https://davisthompsonfoundation.org/noahs-arkive/

Durning, S. J., Costanzo, M. E., Artino, A. R., Graner, J., van der Vleuten, C., Beckman, T. J., Wittich, C. M., Roy, M. J., Holmboe, E. S., & Schuwirth, L. (2015). Neural basis of nonanlytical reasoning expertise during clinical evaluation. *Brain and Behavior, 5*(3), e00309.

Ericsson, K. A., & Harwell, K. W. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.02396

Gao, Y., Cavuoto, L., Dutta, A., Kruger, U., Yan, P., Nemani, A., Norfleet, J. E., Makled, B. A., Silvestri, J., Schwaitzberg, S., Intes, X., & De, S. (2021). Decreasing the surgical errors by neurostimulation of primary motor cortex and the associated brain activation via *Neuroimaging*. *Frontiers in Neuroscience, 15*, 651192. https://doi.org/10.3389/fnins.2021.651192

Hruska, P., Hecker, K. G., Coderre, S., McLaughlin, K., Cortese, F., Doig, C., Beran, T., Wright, B., & Krigolson, O. (2016b). Hemispheric differences in novices and experts during decision making. *Advances in Health Sciences Education*. https://doi.org/10.1007/s10459-015-9648-3

Hruska, P., Krigolson, O., Coderre, S., McLaughlin, K., Cortese, F., Doig, C., Beran, T., Wright, B., & Hecker, K. G. (2016a). Working memory, reasoning, and expertise in medicine–insights into their relationship using functional neuroimaging. *Advances in Health Sciences Education*. https://doi.org/10.1007/s10459-015-9649-2

Jensen, G., Resnik, L., & Haddad, A. (2008) Expertise and clinical reasoning. In Higgs J, Jones M, Loftus S, Christensen N (Eds.), *Clinical Reasoning in the Health Professions* (3rd ed.) Philadelphia, PA: Butterworth Heinemann Elsevier.

King, J.M. (2021) *Dr. John M. Kings' Necropsy Show and Tell: Cornell College of Veterinary Medicine.* Retrived December 1, 2021, from https://secure.vet.cornell.edu/nst/nst.asp.

Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2013). How we learn to make decisions: Rapid propagation of reinforcement learning prediction errors in humans. *Journal of Cognitve Neuroscience, 26*(3), 635–644.

Krigolson, O. E., Pierce, L. J., Holroyd, C. B., & Tanaka, J. W. (2009). Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Journal of Cognitve Neuroscience, 21*(9), 1834–1841.

Levenson, R. M., Krupinski, E. A., Navarro, V. M., & Wasserman, E. A. (2015). Pigeons (Columba livia) as Trainable Observers of Pathology and Radiology Breast Cancer Images. *PLoS ONE, 10*(11), e0141357.

Luck, S. J. (2014). *An Introduction to the Event-related Potential Technique* (2nd ed.). MITPress.

Marrs, K. A., & Novak, G. (2004). Just-in-time teaching in biology: Creating an active learner classroom using the internet. *Cell Biology Education, 3*(1), 49–61.

Novak, G. (2011). Just-in-time teaching. *New Directions for Teaching and Learning., 128*, 63–73.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pusic, M., Pecaric, M., & Boutis, K. (2011). How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. *Academic Medicine, 86*(6), 731–736.

Rockarts, J., Brewer-Deluce, D., Shali, A., Mohialdin, V., & Wainman, B. (2020). National survey on Canadian undergraduate medical programs: The decline of the anatomical sciences in Canadian medical education. *Anatomical Sciences Education, 13*(3), 381–389.

Rourke, L., Cruikshank, L. C., Shapke, L., & Singhal, A. (2016). A neural marker of medical visual expertise: Implications for training. *Advances in Health Sciences Education, 21*(5), 953–966.

Scott, L. S. (2011). Face perception and perceptual expertise in adult and developmental populations. *Oxford Handbook of Face Perception., 28*, 195–213.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitve Neuroscience, 18*(9), 1453–1465.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, *1210*, 204–215.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psycholological Science, 12*(1), 43–47.

Thompson, J. J., Blair, M. R., & Henrey, A. J. (2014). Over the hill at 24: persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. *PLoS ONE, 9*(4), e94215. https://doi.org/10.1371/journal.pone.0094215

Toy, S., Ozsoy, S., Shafiei, S., Antonenko, P., & Schwengel, D. (2023). Using electroencephalography to explore neurocognitive correlates of procedural proficiency: A pilot study to compare experts and novices during simulated endotracheal intubation. *Brain and Cognition, 165*, 105938. https://doi.org/10.1016/j.bandc.2022.105938

Warren, A. L., & Donnon, T. (2013). Optimizing biomedical science learning in a veterinary curriculum: A review. *Journal of Veterinary Medical Education, 40*(3), 210–222.

Williams, C. C., Hecker, K. G., Paget, M. K., Coderre, S. P., Burak, K. W., Wright, B., et al. (2017). The application of reward learning in the real world: Changes in the reward positivity amplitude reflect learning in a medical education context. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2017.10.010

Authors and Affiliations

Sarah J. Anderson[1,2] · Amy L. Warren[1] · Nia Abdullayeva[3] · Olav Krigolson[4] · Kent G. Hecker[1,5]

✉ Kent G. Hecker
kghecker@ucalgary.ca

1 Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada

2 Department of Cell Biology and Anatomy, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

3 Department of Biological Sciences, Faculty of Science, University of Calgary, Calgary, AB, Canada

4 Neuroscience Graduate Program, University of Victoria, Victoria, BC, Canada

5 Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada