Motor Learning Mechanisms are not modified by Feedback Manipulations in a Real-World Task

Federico Nardi[1,2,3], A. Aldo Faisal[1,3,4,5†], Shlomi Haar[2,6*†]

[1]UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London, United Kingdom.
[2]UK Dementia Research Institute – Care Research & Technology, Imperial College London, London, United Kingdom.
[3]Department of Computing, Imperial College London, London, United Kingdom.
[4]Department of Bioengineering, Imperial College London, London, United Kingdom.
[5]Chair in Digital Health & Data Science, University of Bayreuth, Bayreuth, Germany.
[6]Department of Brain Sciences, Imperial College London, London, United Kingdom.

*Corresponding author(s). E-mail(s): s.haar@imperial.ac.uk;
Contributing authors: f.nardi21@imperial.ac.uk; a.faisal@imperial.ac.uk;
†These authors contributed equally to this work.

Abstract

This study examines the distinctiveness of error-based and reward-based mechanisms in motor learning, which are traditionally isolated in laboratory tasks but co-occur in real-world scenarios. Using Embodied Virtual Reality (EVR) of pool billiards - allowing for full proprioception via interaction with the physical pool table, cue stick, and balls - we introduced visual perturbations to a real-world task. Participants underwent sessions learning a visual rotation with either error or reward feedback. While participants corrected the entire rotation with error feedback, only partial correction was observed with reward feedback, highlighting the influence of the feedback regime. However, lag-1 autocorrelation and inter-trial variability decay, indicators of skill learning, showed no significant differences between sessions, suggesting that the provided visual feedback did not exclusively engage specific learning mechanisms. Analysis of post-movement beta rebound (PMBR), a marker of brain activity associated with learning mechanisms, revealed a decrease in PMBR with reward feedback but no consistent trend during error feedback sessions. These findings suggest that while reward feedback was absent in error conditions, participants still engaged in reward-based learning, indicating that feedback manipulation alone may not sufficiently challenge individual learning mechanisms in real-world settings. This study underscores the complexity of motor learning processes and highlights that visual feedback by itself can not elucidate the interplay between error-based and reward-based mechanisms in real-world contexts.

Keywords: Motor control, motor learning, adaptation, reward, PMBR, Virtual Reality

1 Introduction

Motor learning plays a key role in human behaviour throughout our lifespan, from a baby learning to walk to a person living with Parkinson's Disease learning to modify their movement to compensate for their tremor. Motor learning is considered to be based on two learning mechanisms: *error-based* learning, which consists of quantifiable and discernible sensory-prediction errors, and *reward-based* learning, which is driven by reinforcement of model-free successful actions (Wolpert et al., 2011). While playing billiards, for instance, based on the disparity between the pocket (intended outcome) and the actual outcome we try to correct our next shot using *error-based* learning, and after pocketing the ball, the successful actions are positively reinforced via *reward-based* learning, thus increasing the likelihood of the same action being repeated in the future.

This dichotomy in motor learning mechanisms has been the participant of extensive study in the field of neuroscience. *Error-based* learning is considered to be based on adaptation mechanisms in the cerebellum (Kawato and Gomi, 1992; Doya, 2000; Taylor and Ivry, 2014) while *reward-based* learning is often associated with the brain's reward pathways in the basal ganglia and the release of neurotransmitters such as dopamine, which play a crucial role in encoding the value of particular motor actions (Dayan and Balleine, 2002; Niv, 2009; Diedrichsen et al., 2010).

Understanding the interaction and influence of these mechanisms is crucial for developing effective strategies to enhance motor learning and skill acquisition (Krakauer, 2006). Studying their interplay in realistic and diverse settings can provide a deeper understanding of how individuals learn and adapt their movements in daily life (Wolpert et al., 2011).

These two processes are usually studied as distinct and separate mechanisms in simple laboratory tasks that allow for isolating them through the introduction of controlled feedback manipulations (Taylor and Ivry, 2014; Therrien et al., 2016). However, the unparalleled complexity of human movement cannot be fully captured by such simplistic tasks. We recently demonstrated that in a real-world task of playing billiards, error- and reward-based learning coexist, and participants use a combination of both mechanisms while predominantly relying on one or the other (Haar and Faisal, 2020).

Recent advancements in real-world neuroscience have made it possible to study neurobehavioral processes in natural human behaviour, offering new insights into the complexities of motor learning and skill acquisition (Haar et al., 2020; Stout et al., 2021; Krotov et al., 2022; Shamay-Tsoory and Mendelsohn, 2019). However, these approaches also come with limitations. For instance, controlling different types of feedback and learning mechanisms is more challenging compared to classical lab-based tasks. To overcome this limitation, we integrated our billiards task into an Embodied Virtual Reality setup. This setup allows for visual manipulations without compromising the complex movements required by a real-world task (Haar et al., 2021). The Embodied Virtual Reality (EVR) setup synchronises a physical pool table, cue stick, and balls with a virtual environment using optical marker-based motion capture technology. Participants can interact with physical objects simultaneously in both the real and virtual worlds. This capability enables us to manipulate the visual environment while participants engage in realistic movements accompanied by full proprioceptive feedback (Nardi et al., 2023).

In this study, we used our EVR setup to explore the potential of feedback manipulations commonly used in lab-based tasks to differentiate error-based and reward-based learning mechanisms in a real-world motor learning task. We hypothesised that by eliminating specific visual information in the EVR setting, participants would exclusively rely on the relevant learning mechanism, resulting in noticeable differences in behaviour and neural patterns during training sessions. Our analysis of movement data and adjustments in brain activity following feedback manipulations yields new insights into the connection between visual feedback and the distinct characteristics of error-based and reward-based learning mechanisms in real-world human behaviour.

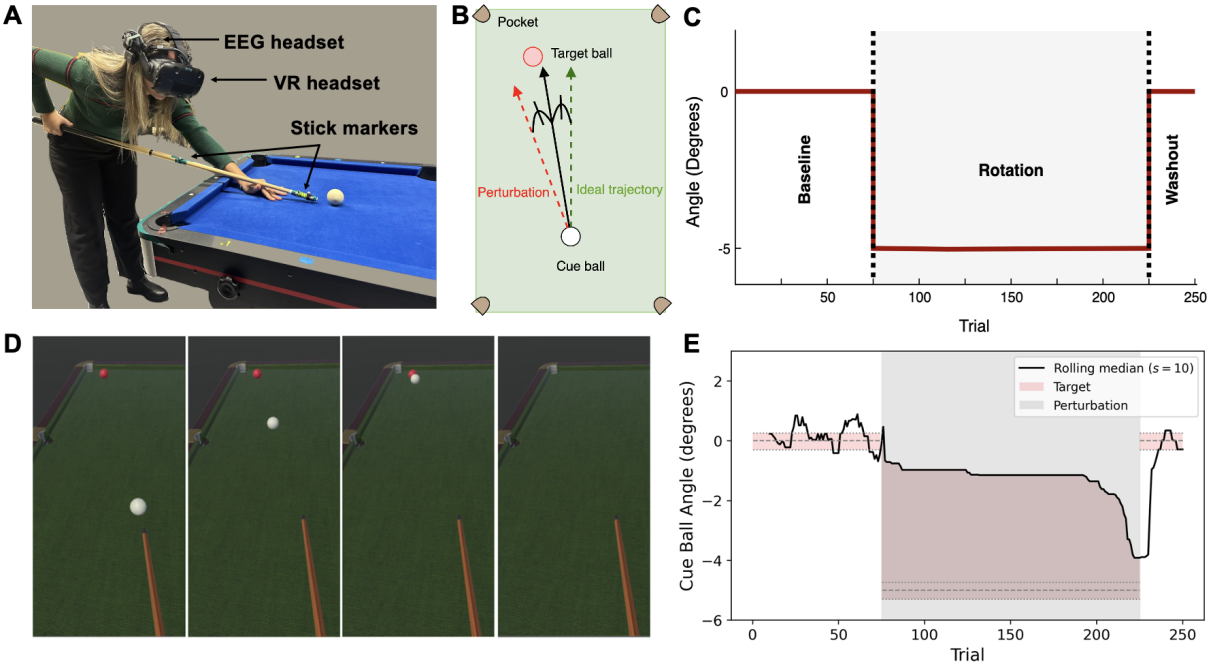

Fig. 1: **(A) Participant playing in the EVR setup**. **(B) Perturbation** applied during the *Rotation* phase: the trajectory of the cue ball is rotated 5° toward the side of the table. **(C) Structure of the experiment**: after 75 trials of full feedback and no perturbation (*Baseline*), a visuomotor rotation is applied on the cue ball trajectory (*Rotation*) for 150 trials; follows a final phase of 25 trials without perturbation and full feedback (*Washout*). The grey area represents the presence of perturbation, whereas the solid red line represents the ideal angle. **(D) Visual feedback for the error task**. At collision, the cue ball and the target ball disappear to make the participant correct only based on the trajectory before the hit. **(E) Reward zone** of an example participant. The grey area represents the presence of perturbation, while the solid black line represents the median of the last 10 rewarded trials. The pink area represents the zone that rewards the participant when the shot is directed in any part of the area: during *baseline* and *washout* it corresponds to the pocketing funnel, while during *rotation* the reward zone is inflated to direct the participant towards the new ideal trajectory. The dotted lines show the range of successful angles, in all the phases of the experiment.

2 Results

32 healthy volunteers took part in an experiment using an Embodied Virtual Reality setup that consisted of a physical pool table, cue stick, and cue ball. Participants played the game while receiving personalised visual feedback in the VR (see fig.1a).

Each participant attended two sessions, 24 to 48 hours apart. Each session included 10 blocks of 25 trials under three conditions: a *Baseline* phase comprising of 3 blocks with full visual feedback; a *Rotation* phase consisting of 6 blocks with a visuomotor rotation of 5° to the trajectory of the cue ball (see fig.1b) and partial visual feedback (error or reward); and a *Washout* phase including one block with no rotation and full feedback (see fig.1c).

During one session, the visual feedback showed only errors without any rewards (cue ball rolling to the target ball, both disappearing after colliding before pocketing, fig.1d). In the other session, we provided reward-only visual feedback by giving pocketing feedback for successful trials and withholding continuous ball feedback for unsuccessful ones. The order of sessions and rotation directions were counterbalanced between participants. Importantly, trials were considered successful and received reward feedback if they fell within a

"reward zone" (Therrien et al., 2016), defined as being better than the median of the previous 10 successful trials (fig.1e).

2.1 Learning and success are strongly modulated by visual feedback manipulations

Participants successfully learned to correct for some of the rotation under both error and reward conditions. However, their learning patterns varied significantly due to feedback manipulation in the EVR setup (see fig. 2a), confirming the preliminary results (Nardi et al., 2023).

During the baseline phase, with full visual feedback and no perturbation, participants displayed similar performance in both sessions. However, during the rotation phase with partial visual feedback and a perturbation on the cue ball's trajectory, participants learned the rotation faster with error-only feedback. This resulted in a larger aftereffect during washout. The trial-by-trial directional errors in the error-feedback session demonstrated a fast learning phase followed by a sustained slow one, best fitted by a double exponential curve ($\tau_{\text{fast|error}} = 6$, $\tau_{\text{slow|error}} = 296$). In contrast, learning through reward-only feedback exhibited slower progress best fitted by a linear curve ($b_0 = 0.31$, $b_1 = -0.014$).

The block averages displayed a similar pattern (see fig. 2b), with a significant difference between the two learning trends during the perturbation phase ($F_{\text{mode|pert}}(1, 62) = 96.25, p < 0.001$; $F_{\text{int|pert}}(5, 310) = 3.94, p < 0.01$). Despite learning at a slower rate in the reward condition, participants managed to compensate for at least some of the visuomotor rotation in both conditions.

Slow learning with reward-only feedback can be partially explained by the use of a "reward zone", leading to increased perceived success in the early blocks of the rotation. In an error feedback session, a successful shot can simply be defined as when the trajectory was heading towards pocketing (though the pocketing was not visible). With reward-only feedback, there are two types of success: *real* and *perceived*. While real success is equivalent to that in the error feedback condition, where the trajectory was heading towards pocketing, perceived success is based on received rewards. Early on in rotations, rewards are given whenever the direction is closer to reaching for a rotated pocket than for an original one.

Accordingly, success rates were similar between sessions in the baseline and washout blocks ($F_{\text{mode|base}}(1, 62) < 0.01, p = 0.99$; $F_{\text{mode|wash}}(1, 62) = 2.57, p = 0.12$). However, during the perturbation phase we observed very different success profiles (fig. 2c). The real and perceived success rates during reward-only blocks showed opposite trends ($F_{\text{int|pert}}(5, 310) = 7.54, p < 0.001$), and the success rates during error-only blocks were different from both perceived ($F_{\text{int|pert}}(5, 310) = 10.11, p < 0.001$) and real reward ($F_{\text{mode|pert}}(1, 62) = 52.33, p < 0.001$).

We then analyse the variability in shot direction under different conditions. To address potential learning trends within a block, we calculate the *corrected variability* by determining the standard deviation of residuals from a block-level linear regression model applied to the cue ball angle data (Haar et al., 2020). The participants' performance evolution closely corresponded with their perceived success rate (fig. 2d), exhibiting similar patterns during the baseline and washout phases ($F_{\text{mode|base}}(1, 62) = 0.43, p = 0.52$; $F_{\text{mode|wash}}(1, 62) = 0.43, p = 0.52$), but demonstrating opposite dynamics during the perturbation phase over time ($F_{\text{int|pert}}(5, 310) = 6.90, p < 0.001$).

2.2 Behavioural Metrics suggest mixture in Learning Mechanism

To assess the contribution of reward-based reinforcement learning and error-based adaptation in the two feedback conditions, we looked at metrics of skill and adaptation learning, which are traditionally linked to reward and error mechanisms, respectively.

We initially analysed the lag-1 autocorrelation [ACF(1)] of the directional error of the cue ball. The decay of ACF is considered a characteristic of skill learning, related to reward-based learning rather than error-based adaptation (van Beers et al., 2013). To quantify the change in ACF, we examined its decay between the two halves of the perturbation phase. Interestingly, a statistically significant decrease was found in the error-only task, tested with a Wilcoxon rank-sum test ($W = 45, p < 0.001$), indicating skill enhancement even

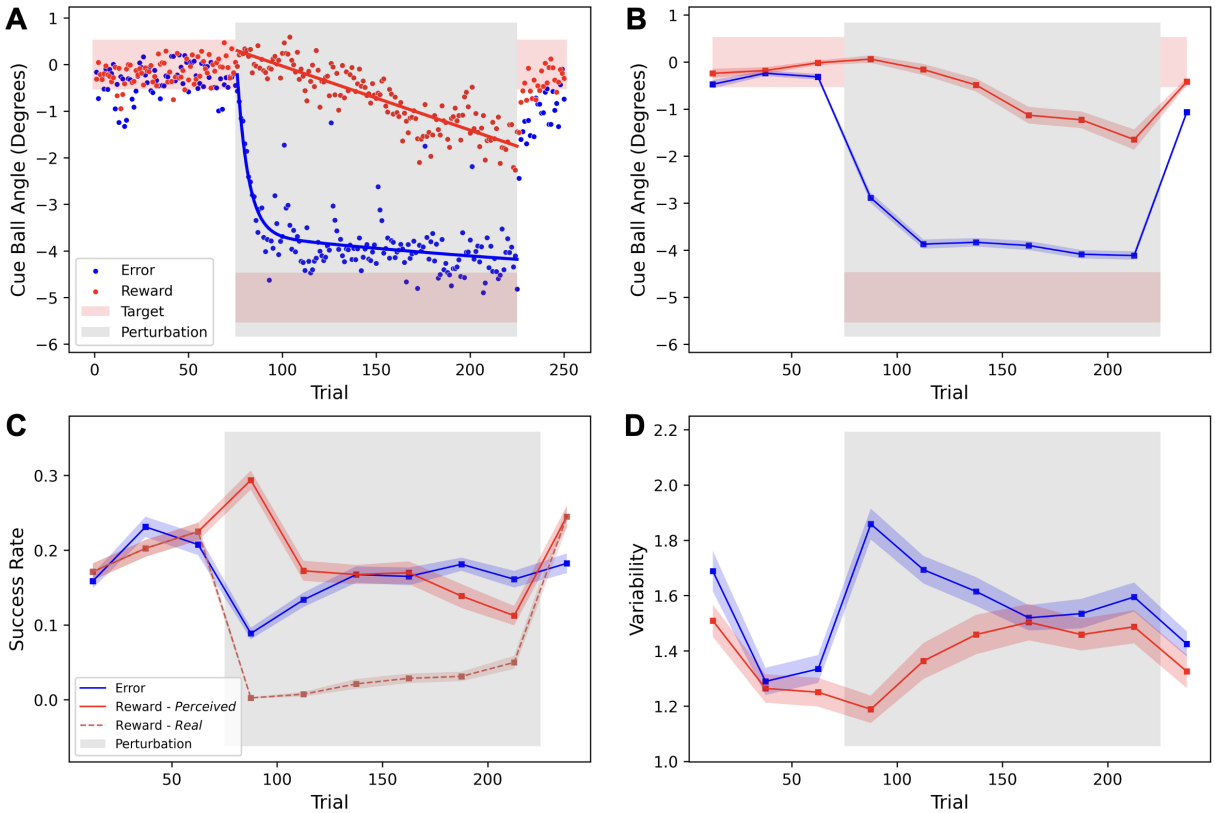

Fig. 2: Quantities in (B)/(D) are calculated within blocks of 25 trials for error (blue) and reward (red) groups, while in (A) at a trial level. Shaded areas represent inter-participant variability (standard error of the mean), grey area represents the presence of perturbation, whereas pink area indicates the range of successful angles. $N = 32$.
(A) Average cue ball trial-by-trial angle. Double exponential fit is in bold. **(B)** Average cue ball angle over blocks. Each block is represented by its distribution, and the means over participants are connected by the bold lines. **(C)** Average perceived success rate, which includes the entire reward zone during *rotation* phase for the *reward* group. **(D)** Average corrected standard deviation, after removal of linear trends.

without visible reward feedback. In contrast, there was no significant change in ACF for participants during the reward-only task as learning progressed from early to late stages ($W = 224, p = 0.47$).

The pattern observed from baseline (trials 1-75) to early learning (trials 76-150) remained consistent across both sessions, aligning with the perturbation provision and the decline in learning seen in the behavioural results. Additionally, a significant difference was detected between feedback typologies, using a 2-sample Mann-Whitney test ($U = 777, p < 0.001$).

The second metric that we investigated is the decay in inter-trial variability, which is known to be a distinctive feature of reward-based skill learning (Müller and Sternad, 2004; Cohen and Sternad, 2009; Shmuelof et al., 2012; Sternad, 2018; Krakauer et al., 2019). We quantified it as the difference in corrected variability between the first and last blocks of the perturbation phase, as representative of the uncertainty throughout the learning period.

Surprisingly, in line with lag-1 autocorrelation, we observed a positive decay when providing error feedback ($W = 142, p = 0.02$), while the opposite trend was displayed for the reward feedback task ($W = 122, p < 0.001$). We also obtained a significant result when testing for differences

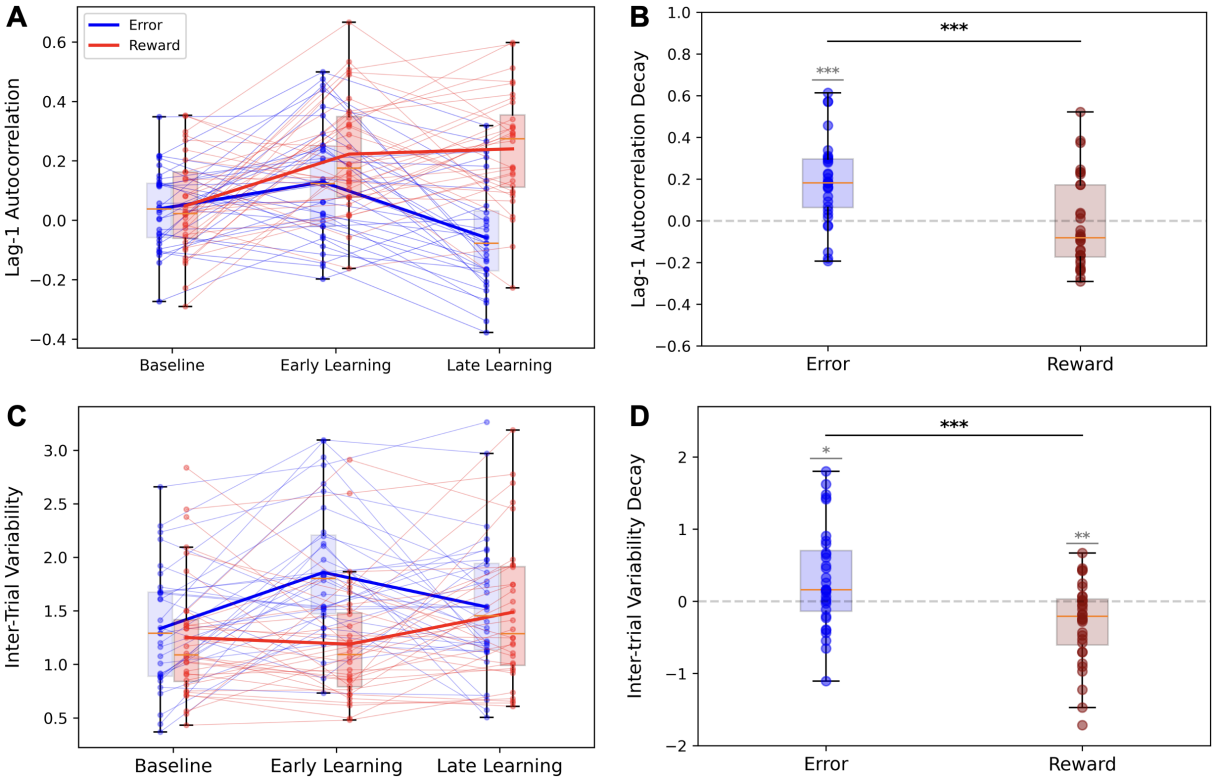

Fig. 3: All quantities are calculated for error (blue) and reward (red) groups and (B) and (D) were tested for significant differences against 0 (in grey) and between groups (in black). $N = 32$.
(A) Lag-1 autocorrelation during *Baseline* (Trials 1-75), *Early Learning* phase (trials 76-150) and *Late Learning* (trials 151-225); **(B)** Decay of lag-1 autocorrelation between *early* and *late* learning, during the *rotation* phase; **(C)** Inter-trial variability in the last baseline block (*Baseline* - trials 51-75), first block of rotation block (*Early Learning* - trials 76-100) and last rotation blocks (*Late Learning* - trials 201-225); **(D)** Inter-trial variability decay from *Early* and *Late* learning.
$^{*}: p < 0.05$, $^{**}: p < 0.01$, $^{***}: p < 0.001$

between the two conditions ($U = 767, p < 0.001$), whereas no differences were detected during the final baseline block ($U = 500, p = 0.87$).

2.3 EEG Neural Metrics suggest mixture in Learning Mechanism

Throughout the experiments, we recorded the brain activity of the players to measure trends in Post-Movement Beta Rebound (PMBR) over the course of their learning. PMBR is defined as the peak in beta activity after a movement ceases (see Fig. 4a). It is a well-established marker of sensorimotor cortex activation and is associated with motor learning, with its trend during learning expected to depend on the learning mechanism (Tan et al., 2014, 2016; Torrecillos et al., 2015; Kranczioch et al., 2008; Haar and Faisal, 2020) . The increase in beta power was evident in the data (see fig. 4c) for all individual sessions.

Both error and reward tasks exhibit similar PMBR trends during the baseline, with a consistent decrease over time. However, there is a distinct difference in the PMBR trend during the rotation phase (fig.4b). The reward sessions show a continuous decline ($b_{\text{block}} = -1.955$, $p < 0.001$, 95% CI: [-2.863, -1.047]), while the participants in the error task do not show a significant trend ($b_{\text{block}} = -1.185$, $p = 0.09$, 95% CI: [-2.534,

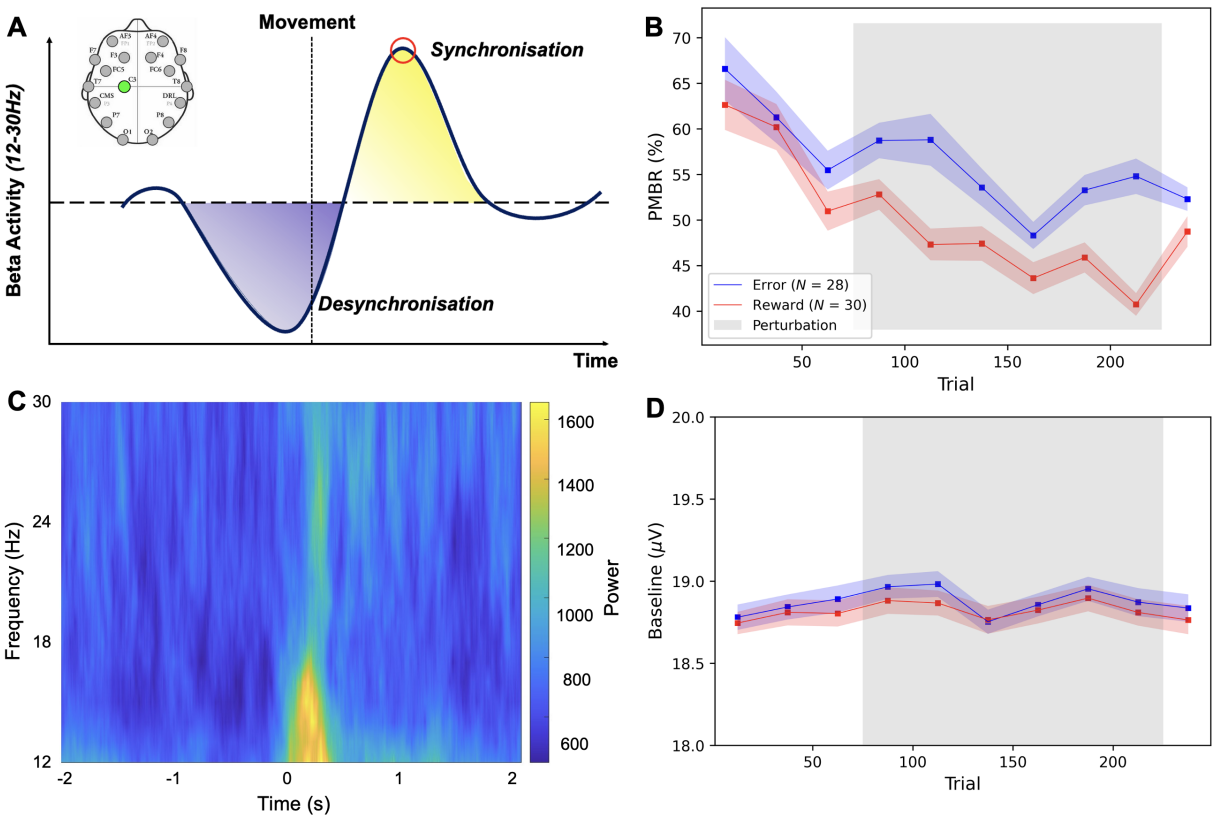

Fig. 4: All quantities are calculated within blocks of 25 trials over *error-based* participants (blue) and *reward-based* participants (red). The grey area represents the presence of perturbation, and the shaded areas represent the variability between the participants (standard error of the mean). **(A)** Graphical representation of β-band *Event-Related Desynchronisation* (in blue) and *Event-Related Synchronisation* (in yellow). Circled in red the *Post-Movement Beta Rebound* peak considered for the analysis. In the upper left angle, the *Emotiv* EPOC+ electrode map within green C3, the channel used for the PMBR derivation; **(B)** Average *Post-Movement Beta Rebound*; **(C)** Time-frequency plot for an example session; **(D)** Average median power used as baseline for the PMBR calculation.

0.164]). During the washout, the average PMBR remains relatively consistent between conditions without significant differences ($F_{\text{mode}}(1, 56) = 0.69,\ p = 0.41$).

No significant baseline differences were identified between feedback types ($F_{\text{mode}}(1, 56) = 0.07, p = 0.79$) or trends linked to the experimental phase ($F_{\text{int}}(9, 504) = 0.24, p = 0.99$), ensuring that changes in PMBR are not driven by underlying shifts (fig. 4d).

3 Discussion

In this study, we explored the impact of feedback manipulations on motor learning and the use of different learning mechanisms in a real-world task. We used an Embodied Virtual Reality setup to introduce visuomotor perturbation and manipulate feedback for a real-world task. Our findings indicate that the use of error-only or reward-only visual feedback is insufficient to induce learning solely based on error-based adaptation or reward-based skill acquisition. Both behavioural and neural data suggest that even when the reward feedback (ball pocketing) was removed in

the error-only condition, participants still engaged in reward-based skill learning during this task. It appears that participants found sufficient rewards in observing what appeared to be the correct ball trajectory – similar to how professional basketball players celebrate successful shots before seeing if they go into the net.

Our results showed a noticeable difference in how individuals adapted to the visuomotor perturbation when receiving error-only versus reward-only feedback. Participants adapted more quickly to cue ball rotation in the error feedback condition compared to the reward feedback condition, aligning with previous research suggesting that adaptation to errors occurs faster than adaptation based on rewards (Galea et al., 2015; Song et al., 2020).

The definition of the reward zone played a major role here, as the increase in the received reward during the early rotation phase with reward-only feedback was not encouraging exploration and learning, as is also evident in the reduced intertrial variability in those blocks. The dynamic reward zone which decreased over time, made the task more challenging, leading to a decrease in perceived success and an increase in variability and exploration over the rotation blocks of the reward-only condition.

The dynamic reward provision was designed to guide participants on different paths without excessive or insufficient rewards, following established paradigms in a lab-based motor learning task (Therrien et al., 2016). However, due to its nature, participants did not experience an abrupt perturbation and noticed changes slowly, possibly affecting the way they perceived the rotated trajectory.

As a result, in the reward-only feedback condition, the reduction in directional error presented a slow and linear learning curve. The *real* success rate in the reward task showed the same slow and linear improvement observed in the reduction of directional errors. On the contrary, in the error-only feedback condition, the learning curve of the decrease in directional error followed a double exponential curve in a pattern consistent with typical adaptation tasks (e.g., Taylor et al., 2014) and with performance in the real-world billiard task (Haar et al., 2020).

While in lab-based tasks the feedback manipulations of error and reward are considered to induce different learning mechanisms, our results suggest a more complex picture.

The Lag-1 autocorrelation [ACF(1)] of the error has been suggested as a marker of expertise; for novices, it is expected to be significantly different from zero and decrease over time (in absolute value), as values close to zero are linked to greater skill (van Beers et al., 2013). The rationale behind this metric is that as skills develop, individuals become less influenced by noise from their previous actions and therefore the trials would be less correlated with the previous ones. In terms of the duality of learning mechanisms, ACF(1) is expected to decay during reward-based skill learning and not to change during error-based adaptation (van Beers et al., 2013).

During baseline, ACF(1) did not differ from zero in both conditions. In the early learning phase, ACF(1) increased and was significantly different from 0 in both conditions. Moreover, contrary to initial expectations, the error-only feedback condition exhibited a decrease in ACF towards zero during the perturbation phase, indicating reduced reliance on previous performance. This measure demonstrated an improvement in the reliability of the shot by the participants, suggesting that the error-based adaptation was not the only mechanism used for movement correction. The reinforcement feedback may have been obtained from the expected outcome of the trajectory or through additional sensorimotor information from other sources, such as the proprioceptive feedback experienced when hitting the ball.

On the contrary, the reward feedback condition displayed relatively consistent levels of ACF(1) over the rotation phase, implying a steady dependence on past results despite evolving task demands. This result can be attributed to the limited amount of learning in this condition or the continuous reliance on limited binary information.

The decay in inter-trial variability is another measure that is traditionally linked to skill learning and not to adaptation (Müller and Sternad, 2004; Cohen and Sternad, 2009; Shmuelof et al., 2012; Huber et al., 2016). Our findings regarding inter-trial variability revealed interesting differences between the error and reward conditions. In the error condition, we observed a decay in inter-trial variability from the beginning to the end of the perturbation phase. Conversely, in the reward condition, we noted a more fluctuating and diverse

pattern of inter-trial variability, with no decay trend across participants. This indicates a continuous exploration of different movement strategies as might be expected in light of the directional error and success rates in reward-based feedback conditions, which suggest that by the end of the rotation phase participants are still learning.

Overall, all of the above results converged to the idea that participants did not exclusively use the learning mechanism related to the specific feedback, especially in the error task, where there are clear signatures of reward-based skill learning.

To verify the absence of clear separation between learning mechanisms despite the feedback provided and to validate the difference in behavioural measures of learning in real-world scenarios with lab-based tasks, we investigated as well the neural activity of the participants for markers of different learning mechanisms. Specifically, we studied the Post-Movement Beta Rebound (PMBR), the increase in power of neural beta oscillations (12-30Hz) at the end of the movement.

PMBR is known to increase over learning in error-based adaptation tasks (Tan et al., 2014; Torrecillos et al., 2015; Tan et al., 2016), whereas in skill-learning tasks PMBR is expected to decrease (itself or its magnetic resonance spectroscopy correlate) over the learning (Floyer-Lea et al., 2006; Kranczioch et al., 2008; Kolasinski et al., 2019). This divergence was attributed to the projections of the GABA activity for error- and reward-based learning, respectively from the cerebellum and basal ganglia (Doyon et al., 2003; Doyon and Benali, 2005)

Our results indicate a high initial PMBR that decreases during baseline in both sessions, as observed in skill-learning tasks (e.g. Kranczioch et al. (2008)). This is expected as participants learn the EVR environment and the pool task. During the rotation phase, after introducing perturbation and partial feedback manipulation, overall PMBR continued to decay for both feedback conditions. In the reward-only condition, there was a clear decay, which is expected in a reward learning task with no error-based adaptation. For the error-only condition, there was a mixed PMBR trend with a decay over the early learning suggesting the use of reward-based reinforcement skill learning, and it increased over the later part of the rotation phase, where presumably the errors got too small for predicting reward from the ball trajectory and only implicit error-based adaptation was taking place then. This implicit component is evident in the aftereffects during washout.

Our findings suggest the combined use of error- and reward-based learning in real-world scenarios, even while withdrawing error or reward feedback. Both behavioural and neural results indicate some interaction between mechanisms, despite attempts to isolate the two through feedback. The richness of stimuli and experience in a real-world task limits the effect of feedback manipulations. This is strongly evident in the error-only conditions, where the participants are probably experiencing enough information from the cue ball trajectory to preserve success and use reward-based reinforcement. In the reward-only condition, we are withdrawing visual sensory-error feedback, but the participants are experiencing full proprioceptive feedback which potentially can guide some error-based adaptation in this condition.

While our study provides valuable insights into the interaction between error- and reward-based learning mechanisms in a real-world scenario, it is important to acknowledge certain limitations that may have influenced the findings. Firstly, the sample size of participants could have an impact on the generalisability of the results. A larger and more diverse participant pool could offer a broader perspective on the dynamics of error- and reward-based learning in motor tasks. Secondly, the consideration of Post-Movement Beta Rebound (PMBR) as a neural marker for the learning mechanisms may not have fully captured the complex neural processes associated with error- and reward-based learning. Future studies could consider a more comprehensive assessment of neural activity across different brain regions to provide a more comprehensive understanding of the underlying mechanisms towards the identification of patterns beyond PMBR. Thirdly, the way the reward feedback was provided did not allow the participants to correct fully for the perturbation, due to the scarce information retrievable. Different methods to provide optimal reward information need to be further explored, as the current reward zone method suffices for lab-based tasks but could not be sufficiently informative in the case of real-world tasks. Finally, while the use of EVR

provided a valuable context of real-world scenarios for studying error- and reward-based learning, it also introduced additional variables and complexities, such as the use of Virtual Reality, that may have influenced the observed dynamics.

The findings from this study challenge the notion that learning mechanisms can be isolated solely through feedback manipulation. The observed differences in ACF(1) and variability decay and in neural activity between error-only and reward-only tasks underscore the intricate nature of motor learning and adaptation in dynamic environments. While feedback manipulation certainly shapes the learning process and influences motor performance, it is clear that participants do not exclusively rely on single learning mechanisms even while receiving partial visual feedback. Instead, they engage in a holistic learning process that incorporates various environmental cues and internal computations.

In conclusion, this study provides valuable insights into the complexities of human motor learning and performance, emphasising the need for a nuanced understanding of the interaction between error- and reward-based learning mechanisms. These findings have implications for the development of more effective learning interventions and rehabilitation strategies. Further research in this area has the potential to advance our knowledge of motor learning processes and inform the design of interventions that better align with the complexities of human motor adaptation.

4 Methods

4.1 Experimental Setup

Our experimental setup involves the use of a physical pool table to link real-world experiences with Embodied Virtual Reality (EVR) to obtain accurate measurements and adjustments. Participants play an actual pool game while wearing an HTC Vive headset and receiving customised visual feedback in the virtual environment, related to the learning mechanism aimed to be engaged. The physical cue stick used in the game is replicated in VR using its geometric characteristics and marker positions, streamed with high precision (accuracy of ±0.2mm) through Optitrack cameras, allowing participants to perceive it as a virtual object as well.

This embodiment enables participants to physically interact with real-world objects, experiencing complete somatosensory and proprioceptive feedback. Haptic feedback is maintained as participants shoot a physical cue ball along with realistic collision sounds; however, artificial sounds represent collisions between the cue ball and target balls that are solely virtual but programmed to adhere to real physics similar to other objects in the virtual environment (for validation see Haar et al., 2021). Pre-experiment calibration via four cameras and VR controllers ensures alignment between dimensions of virtual and real tables for enhanced realism.

The entire EVR paradigm was developed using Unity software coded in C#. An electroencephalogram device (eMotiv EPOC+) has been integrated into the setup for recording brain activity during experiments; this 14-channel wireless device can be easily worn alongside the VR headset without interference.

4.2 Experimental Design

The experiment included two sessions which took place on different days. Each session (fig. 1c) consisted of a baseline phase (75 trials, divided into 3 blocks of 25 trials), a perturbation phase (150 trials, comprised of 6 blocks of 25 trials) and a final washout block (25 trials). Each trial involved one attempt where participants had to shoot a cue ball from a fixed starting position aiming to pocket the target ball placed near the far-left or far-right corner of the pool table, without any time pressure. As a result, the duration of the trial ranged from several seconds up to more than ten seconds as per individual pacing.

32 healthy volunteers (20 men and 12 women) aged 20 to 25 with normal or corrected-to-normal vision and little to no experience in playing pool, participated in an experiment. Each participant took part in two sessions. During the perturbation phase, a 5° rotation was applied to the trajectory of the cue ball (Fig. 1b), requiring participants to adjust their aim toward the centre of the table to successfully pocket it. During this phase of rotation, specific feedback related to the condition was provided through virtual reality. To compel participants to learn using a particular method, we

structured the task in such a way that visual information that contributes to the opposite method was concealed.

In the error-based feedback condition, the ball trajectories were hidden after colliding with each other, requiring individuals to rely solely on cue ball trajectory errors for progression as they could not see the target ball pocketing. On the other hand, in the reward feedback condition, balls were hidden after shooting the cue ball. Successful shots resulted in participants being shown an artificial successful trajectory as a form of reinforcement, according to reward criteria well established in motor learning literature (Therrien et al., 2016, 2018; van Mastrigt et al., 2023).

This reward region was established using two specific criteria: (i) shooting within a successful zone around the pocket, or (ii) when the accuracy of the shot exceeded the median of the previous 10 successful attempts. The defined successful trials assisted participants in making necessary adjustments without error feedback.

A previous pilot study aimed to find the best level for providing reinforcement. Giving too many rewards would make participants reach the correct path quickly, but may reduce their interest. On the other hand, setting the reward threshold too high would make it difficult for participants to adjust in a reasonable time due to a lack of guidance. To verify this, several participants took part in pilot experiments where trials were considered "successful" if better than a specific percentage (ranging from 30% to 70%) of earlier rewarded trials. The optimum threshold was found by surpassing both the 40th percentile and median performance levels, which were close; therefore, we chose to prioritise meeting the median requirement for easier explanation without significant compromise.

4.3 Behavioural Data Analysis

The data were pre-processed to address extreme values: a threshold was established to remove outliers in terms of cue ball directional errors during the baseline ($\mu \pm 3\sigma$) and applied in the perturbation learning zone.

All primary results were evaluated by grouping trials into blocks of 25 to enhance the robustness and reliability of detected trends. No significant difference between target pockets was observed, enabling aggregation of samples (aiming to the left or right pockets) for increased statistical power.

The main focus of the behavioural data was on the directional error made by the participants in each shot. After defining the ideal successful trajectory (0°), the angle of the cue ball was measured by centring the range around that value and analysing the deviations in the direction of the shot.

The trial-by-trial directional error was used to measure learning. When the visuomotor rotation was introduced during the rotation phase, the "ideal" angle changed from 0° to -5° due to rotation, and all results are displayed accordingly (fig.1c). To choose the best fit between linear and double exponential for the learning curves of the two feedback, we performed an F-test comparing the residual sum of squares of the two curves. The parameters of the curves are reported in the results and referred as τ_{fast} and τ_{slow} respectively for the fast and slow components of the exponential curve. The linear fit is reported through b_0 and b_1, i.e. the intercept and slope of the linear model.

Another game feature that was taken into account was the success rate, which included two different components: *real* success rate and *perceived* success rate. During the baseline and washout phases, when no changes were made, these rates matched. In the phase of perturbation for the error feedback task, they aligned because participants did not have visual information about whether their shots were successful or not; instead, for the reward session the two differed significantly, as perceived success consisted in the rewards given based on the reward zone while real success referred to the actual instances when the target ball was pocketed.

Finally, to account for uncertainty in the shots and maintain consistency with previous research, the variability of the shots was taken into consideration. In particular, the focus was placed on measuring the corrected variability - that is, the standard deviation of the residuals after removing any linear trend from the directional error in shots at a block level. This approach remained valid even when an overall exponential trend was present at the session level, as there was a sufficiently linear trend within blocks of 25 trials.

In terms of methodology, mixed-effects ANOVA and mixed-effect linear models were used to analyse the repeated measurements from each

participant. Non-parametric tests were utilised for sample testing and comparisons to comply with non-normal distributions of the samples. Specifically, Wilcoxon-Rank tests (reported as W) were used to compare single samples against a 0 median, while paired Mann-Whitney tests (reported as U) were used to compare the different samples.

All the mixed effects ANOVAs tests in the results sections refer to baseline ($F_{\bullet|\text{base}}$), perturbation ($F_{\bullet|\text{pert}}$) or washout ($F_{\bullet|\text{wash}}$) and have been executed between different modes ($F_{\text{mode}|\bullet}$) or interaction between mode and time ($F_{\text{int}|\bullet}$).

4.4 Behavioural Measures of Learning

To further validate the behavioural results obtained with the two individual mechanisms, two well-known measures of motor learning were used to demonstrate differences between the feedback: lag-1 autocorrelation and inter-trial variability.

The lag-1 autocorrelation (ACF) is calculated by correlating the directional error time series of one session with itself with a one-time step lag. To address potential bias in autocorrelation estimation from short time series (Arnau and Bono, 2001), we analyse autocorrelation functions for sets of trials during the learning session at different stages: baseline (blocks 1-3), early learning (blocks 4-6), and late learning (blocks 7-9) rather than within each set of 25 trials.

Inter-trial variability (ITV) was defined at a block level as the corrected variability over the 25 trials within each block, again considering the baseline (last block - trials 51 to 75), early learning (first block - 76 to 100), and late learning (last block - 201 to 225).

We evaluated the change in ACF and ITV during the learning process by calculating the decay of ACF based on the difference between the halves of the rotated phase, and similarly, we assessed the decay of ITV as derived from the difference between initial and final perturbation blocks.

4.5 EEG Acquisition and Pre-processing

EEG data was collected using an Emotiv EPOC+ headset, a wireless 14-channel system positioned under the VR headset. The data was sampled at 256Hz and aligned with the VR data through the trial timestamps. The whole pre-processing pipeline was executed in MATLAB using the EEGlab (Delorme and Makeig, 2004) and Fieldtrip (Oostenveld et al., 2011) toolboxes.

After applying high-pass and low-pass filtering with a FIR filter at 1Hz and 45Hz respectively, the data was cleaned using the EEGlab *clean_rawdata* function. This function automatically removes artefacts from signals, including bad portions of data and problematic channels, using the *Artifact Subspace Reconstruction* algorithm (Kothe, 2013). Specifically, it identifies channels with a large amount of noise based on their standard deviation and poorly correlated channels. The default rejection threshold for channel correlation was set to 0.8 but in order to retain as much information as possible, we gradually decreased it by steps of 0.05 until no more than three channels were removed or with a minimum correlation of 0.6. This approach aims to minimise the loss of informative data while also preventing excessive removal of channels or retention of non-informative data.

Furthermore, signal windows were removed based on thresholds for the maximum standard deviation of bursts and the maximum fraction of contaminated channels tolerated in the final output data of each considered window. The specific parameters were selected through manual examination of sessions' time-frequency plots across 4 combinations of parameter settings, with increased levels of strictness.

We split the data into epochs, focusing on the shots as a valid reference for the movement performed by the participants, coherently with a previous *Event-Related Potential* (ERP) and event-related spectral perturbation (ERSP) evaluations and Haar and Faisal (2020). We set the epochs to be 9 seconds long around each shot in order to establish a baseline reference before and after movement which would include larger temporal margins, as well as accurately detect fluctuations in beta activity. However, since PMBR can be detected up to 1000 ms after movement onset (Jurkiewicz et al., 2006), we focused only on that interval to identify the peak in the beta activity.

An *Independent Component Analysis* was then performed on the data to remove muscle-, eye- and heart-induced artefacts, and, afterwards, the C3 channel was interpolated from the other existing

channels, using spherical splines and a topological map coherent with the international 10-20 EEG system that positions C3 over the left motor cortex (Jasper, 1958). The choice of channel placement was informed by previous studies in motor learning (e.g. Tan et al., 2014, 2016) and the right-handedness of all participants.

4.6 EEG Time-Frequency Analysis

Following the preprocessing of the EEG data, we performed a time-frequency analysis to investigate brain activity associated with the movements of the participants during the experiment while providing individual feedback.

The time-frequency domain transformation was performed on each block using convolution with complex Morlet wavelets in 0.8 Hz steps with cycles of 3 milliseconds.

To ensure comparability between various sessions within and between participants, we standardised the data so that all sessions have equal total power in the specific frequency range. We specifically calculated the total power within the 8-45Hz range to determine the normalisation factors for each separate session (Liu et al., 2021). This normalisation method guarantees that variations in beta are more noticeable and accurate, allowing for precise detection of variations that are solely due to the signal itself, as potential trends and patterns would be observable across all frequencies and not just in the beta band. Normalising across sessions also eliminates any trends in the baseline, which we define as the median power over each session, as the correction is done as well between the different blocks of the same session.

The subsequent analysis involved calculating the EEG power change in response to events as a percentage change, normalised relative to the median power evaluated in 1Hz steps over the block. This approach was taken due to the absence of a clear baseline during the task (Torrecillos et al., 2015; Tan et al., 2014, 2016). The process involved subtracting one from the normalised value and then multiplying by 100 to obtain a comparable measure on a percentage scale.

Finally, we calculated the peak on the average signal across the beta-band frequencies and derived one value per trial. To ensure robustness, we then took the median over various trials to obtain a PMBR value for the block.

To ensure data reliability, we established criteria to include or exclude specific sessions in the group analysis. EEG data can have missing values, possibly due to electrode misplacement as the task required significant movements.

Blocks of an individual session were included in the EEG time-frequency analysis only if at least half of the block (13 trials) had data available. Entire sessions were removed if at least 3 blocks were missing to accurately capture trends and do not include noisy information.

After selecting reliable sessions, missing data were imputed, separately between error and reward sessions, using *Probabilistic Principal Component Analysis* (PPCA), which samples from the conditional distribution of missing values given the observed data (Tipping and Bishop, 1999).

The overall pre-processing of the data reduced the number of participants in the EEG analysis from 32 to 28 and 30, respectively for the error and reward tasks. Finally, all block values were aggregated by block and reported by condition.

Declarations

Funding

F.N. is supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1]; A.A.F. acknowledges a UKRI Turing AI Fellowship Grant (EP/V025449/1); S.H. is supported by the Edmond and Lily Safra Fellowship Program and by the UK Dementia Research Institute Care Research & Technology Centre.

Conflict of interest/Competing interests

The authors declare no competing interests.

Ethics approval

All experimental procedures were approved by Imperial College Research Ethics Committee and performed in accordance with the declaration of Helsinki.

Consent to participate

All subjects gave informed consent prior to participating in the study.

Availability of data and materials

Data will be made available on FigShare upon publication.

Authors' contributions

S.H. and A.A.F. conceived and designed the study; F.N., A.A.F. and S.H. developed the experimental setup; F.N. acquired the data; F.N., A.A.F. and S.H. analysed and interpreted the data; F.N. drafted the paper; SH and A.A.F. revised the paper.

References

Arnau, J. and Bono, R. (2001). Autocorrelation and bias in short time series: An alternative estimator. *Quality and Quantity*, 35:365–387.

Cohen, R. and Sternad, D. (2009). Variability in motor learning: relocating, channeling and reducing noise. *Experimental brain research*, 193:69–83.

Dayan, P. and Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36(2):285–298.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Diedrichsen, J., White, O., Newman, D., and Lally, N. (2010). Use-dependent and error-based learning of motor behaviors. *Journal of Neuroscience*, 30(15):5159–5166.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current opinion in neurobiology*, 10(6):732–739.

Doyon, J. and Benali, H. (2005). Reorganization and plasticity in the adult brain during learning of motor skills. *Current opinion in neurobiology*, 15(2):161–167.

Doyon, J., Penhune, V., and Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, 41(3):252–262.

Floyer-Lea, A., Wylezinska, M., Kincses, T., and Matthews, P. M. (2006). Rapid modulation of gaba concentration in human sensorimotor cortex during motor learning. *Journal of neurophysiology*, 95(3):1639–1644.

Galea, J. M., Mallia, E., Rothwell, J., and Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nature neuroscience*, 18(4):597–602.

Haar, S. and Faisal, A. A. (2020). Brain activity reveals multiple motor-learning mechanisms in a real-world task. *Frontiers in human neuroscience*, 14:354.

Haar, S., Sundar, G., and Faisal, A. A. (2021). Embodied virtual reality for the study of real-world motor learning. *Plos one*, 16(1):e0245717.

Haar, S., van Assel, C. M., and Faisal, A. A. (2020). Motor learning in real-world pool billiards. *Scientific reports*, 10(1):20046.

Huber, M. E., Brown, A. J., and Sternad, D. (2016). Girls can play ball: Stereotype threat reduces variability in a motor skill. *Acta Psychologica*, 169:79–87.

Jasper, H. (1958). The 10-20 electrode system of the international federation. *Electroencephalogr Clin Neuropysiol*, 10:370–375.

Jurkiewicz, M. T., Gaetz, W. C., Bostan, A. C., and Cheyne, D. (2006). Post-movement beta rebound is generated in motor cortex: evidence from neuromagnetic recordings. *Neuroimage*, 32(3):1281–1289.

Kawato, M. and Gomi, H. (1992). A computational model of four regions of the cerebellum based on feedback-error learning. *Biological cybernetics*, 68(2):95–103.

Kolasinski, J., Hinson, E. L., Divanbeighi Zand, A. P., Rizov, A., Emir, U. E., and Stagg, C. J. (2019). The dynamics of cortical gaba in human motor learning. *The Journal of physiology*, 597(1):271–282.

Kothe, C. (2013). The artifact subspace reconstruction method. *Accessed: Jul*, 17:2017.

Krakauer, J. (2006). Motor learning: its relevance to stroke recovery and neurorehabilitation. *Current opinion in neurology*, 19 1:84–90.

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., and Haith, A. M. (2019). Motor learning. *Compr Physiol*, 9(2):613–663.

Kranczioch, C., Athanassiou, S., Shen, S., Gao, G., and Sterr, A. (2008). Short-term learning of a visually guided power-grip task is associated with dynamic changes in eeg oscillatory activity. *Clinical Neurophysiology*, 119(6):1419–1430.

Krotov, A., Russo, M., Nah, M., Hogan, N., and Sternad, D. (2022). Motor control beyond reach—how humans hit a target with a whip.

Royal Society Open Science, 9(10):220581.

Liu, S., Shen, J., Li, Y., Wang, J., Wang, J., Xu, J., Wang, Q., and Chen, R. (2021). Eeg power spectral analysis of abnormal cortical activations during rem/nrem sleep in obstructive sleep apnea. *Frontiers in Neurology*, 12:643855.

Müller, H. and Sternad, D. (2004). Decomposition of variability in the execution of goal-oriented tasks: three components of skill improvement. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1):212.

Nardi, F., Haar, S., and Faisal, A. A. (2023). Billevr: An embodied virtual reality framework for reward-and-error-based motor rehab-learning. In *2023 International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:1–9.

Shamay-Tsoory, S. G. and Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspectives on Psychological Science*, 14(5):841–859.

Shmuelof, L., Krakauer, J. W., and Mazzoni, P. (2012). How is a motor skill learned? change and invariance at the levels of task success and trajectory control. *Journal of neurophysiology*, 108(2):578–594.

Song, Y., Lu, S., and Smiley-Oyen, A. L. (2020). Differential motor learning via reward and punishment. *Quarterly J. Exp. Psychology*, 73(2):249–259.

Sternad, D. (2018). It's not (only) the mean that matters: variability, noise and exploration in skill learning. *Current opinion in behavioral sciences*, 20:183–195.

Stout, D., Chaminade, T., Apel, J., Shafti, A., and Faisal, A. A. (2021). The measurement, evolution, and neural representation of action grammars of human behavior. *Scientific Reports*, 11(1):13720.

Tan, H., Jenkinson, N., and Brown, P. (2014). Dynamic neural correlates of motor error monitoring and adaptation during trial-to-trial learning. *Journal of Neuroscience*, 34(16):5678–5688.

Tan, H., Wade, C., and Brown, P. (2016). Postmovement beta activity in sensorimotor cortex indexes confidence in the estimations from internal models. *Journal of Neuroscience*, 36(5):1516–1528.

Taylor, J. A. and Ivry, R. (2014). Cerebellar and prefrontal cortex contributions to adaptation, strategies, and reinforcement learning. *Progress in brain research*, 210:217–53.

Taylor, J. A., Krakauer, J. W., and Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *J. Neurosci.*, 34(8):3023–3032.

Therrien, A. S., Wolpert, D. M., and Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*, 139(1):101–114.

Therrien, A. S., Wolpert, D. M., and Bastian, A. J. (2018). Increasing motor noise impairs reinforcement learning in healthy individuals. *eneuro*, 5(3).

Tipping, M. E. and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622.

Torrecillos, F., Alayrangues, J., Kilavik, B. E., and Malfait, N. (2015). Distinct modulations in sensorimotor postmovement and foreperiod β-band activities related to error salience processing and sensorimotor adaptation. *Journal of Neuroscience*, 35(37):12753–12765.

van Beers, R. J., van der Meer, Y., and Veerman, R. M. (2013). What autocorrelation tells us about motor variability: insights from dart throwing. *PloS one*, 8(5):e64332.

van Mastrigt, N. M., Tsay, J. S., Wang, T., Avraham, G., Abram, S. J., van der Kooij, K., Smeets, J. B., and Ivry, R. B. (2023). Implicit reward-based motor learning. *Experimental Brain Research*, 241(9):2287–2298.

Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature reviews neuroscience*, 12(12):739–751.