Punishment Leads to Greater Sensorimotor Learning But Less Movement Variability Compared to Reward

Adam M. Roth, [a] **Rakshith Lokesh,** [b] **Jiaqiao Tang,** [f] **John H. Buggeln,** [b] **Carly Smith,** [b] **Jan A. Calalo,** [a] **Seth R. Sullivan,** [b] **Truc T. Ngo,** [b] **Laura St. Germain,** [f] **Michael J. Carter** [f,*,1] **and Joshua G. A. Cashaback** [a,b,c,d,e,f,*,1]

[a] *Department of Mechanical Engineering, University of Delaware, United States*

[b] *Department of Biomedical Engineering, University of Delaware, United States*

[c] *Kinesiology and Applied Physiology, University of Delaware, United States*

[d] *Interdisciplinary Neuroscience Graduate Program, University of Delaware, United States*

[e] *Biomechanics and Movement Science Program, University of Delaware, United States*

[f] *Department of Kinesiology, McMaster University, Canada*

Abstract—When a musician practices a new song, hitting a correct note sounds pleasant while striking an incorrect note sounds unpleasant. Such reward and punishment feedback has been shown to differentially influence the ability to learn a new motor skill. Recent work has suggested that punishment leads to greater movement variability, which causes greater exploration and faster learning. To further test this idea, we collected 102 participants over two experiments. Unlike previous work, in Experiment 1 we found that punishment did not lead to faster learning compared to reward (n = 68), but did lead to a greater extent of learning. Surprisingly, we also found evidence to suggest that punishment led to less movement variability, which was related to the extent of learning. We then designed a second experiment that did not involve adaptation, allowing us to further isolate the influence of punishment feedback on movement variability. In Experiment 2, we again found that punishment led to significantly less movement variability compared to reward (n = 34). Collectively our results suggest that punishment feedback leads to less movement variability. Future work should investigate whether punishment feedback leads to a greater knowledge of movement variability and or increases the sensitivity of updating motor actions. © 2024 IBRO. Published by Elsevier Inc. All rights reserved.

Keywords: reward, punishment, movement variability, reinforcement, motor learning, sensorimotor adaptation.

1. INTRODUCTION

From hitting the bullseye in a game of darts to mistakenly striking a sharp note on a guitar, success and failure feedback is integral to learning a new motor skill. Indeed, such positive reward and punishment feedback has been shown to enhance distinctly different features of sensorimotor adaptation, such as the rate of learning (Galea et al., 2015; Song and Smiley-Oyen, 2017; Song et al., 2020; Wächter et al., 2009; Wu et al., 2014) and retention of recently acquired motor actions (Abe et al., 2011; Shmuelof et al., 2012; Galea et al., 2015; Song and Smiley-Oyen, 2017; Song et al., 2020; Vassiliadis

et al., 2021). These behavioural differences may be a result of distinct neural pathways governing reward-based and punishment-based processes, evidenced by dissociable physiological responses at the neural (den Ouden et al., 2013; Frank et al., 2004; Hill et al., 2021; Hill et al., 2020; Hamel et al., 2023; Robinson et al., 2010; Wächter et al., 2009) and behavioural (Galea et al., 2015) levels. Understanding how different forms of feedback influence sensorimotor adaptation may be beneficial for enhancing the rate of learning, extent of learning, and retention of newly acquired motor actions in a variety of contexts, including when rehabilitating those with a neurological disorder.

Studying the influence of reward feedback and punishment feedback has a rich history in psychology and more recently has been investigated in sensorimotor neuroscience. Both an increased sensitivity (Ernst et al., 2002; Galea et al., 2015) or an increase in movement variability (Song et al., 2020) in response to punishment feedback have been suggested

to increase learning rate, when compared to reward feedback. Theoretically, a greater sensitivity to punishment feedback can lead to a stronger update to motor commands and faster learning. Likewise, it has been shown that additional movement variability leads to greater exploration, allowing the sensorimotor system to more quickly find successful motor actions and consequently faster learning (Cashaback et al., 2019; Therrien et al., 2016; Wu et al., 2014). Behavioural work by Galea and colleagues (2015) examined the influence of punishment feedback and reward feedback on sensorimotor adaptation. Participants were instructed to reach through a target while using online visual feedback of their hand via a small cursor (i.e., error-based feedback). The cursor position was rotated relative to participant hand position, about the center of the start position, during a block of adaptation trials. This visuomotor rotation required participants to adjust the direction of their reach to hit the target. Concurrently, participants also received either reward feedback (small monetary gain) or punishment feedback (small monetary loss). In one of their experiments the magnitude of reward feedback or punishment feedback was dependent on their distance from the target. Following the adaptation trials, they withheld feedback to assess the retention of the recently acquired reaching behaviour. Interestingly, Galea and colleagues (2015) found a double dissociation in learning rate and retention given reward or punishment feedback. Specifically, they found that participants receiving punishment feedback had a faster learning rate to counteract the visuomotor rotation compared to participants receiving reward feedback. Further, they also found that participants receiving reward feedback had greater retention compared to those receiving punishment feedback. The authors attributed greater learning rates to punishment feedback increasing the sensitivity to negative outcomes in the cerebellum (Ernst et al., 2002) that enhanced visual error-based learning (Hester et al., 2010). A different yet potentially complementary explanation is that punishment feedback increased movement variability to boost exploration and increase the rate of learning.

Indeed, past work has suggested that elevated levels of movement variability have been associated with faster sensorimotor adaptation (Cashaback et al., 2019; Dhawale et al., 2017; Therrien et al., 2016; Wu et al., 2014). Recently, Song and colleagues (2020) asked whether punishment feedback leads to greater movement variability that is used to enhance exploration and accelerate learning. In their task, participants were required to counteract a visuomotor rotation but they did not have visual error-based feedback of a cursor during adaptation trials. They were given scalar punishment or reward feedback in the form of a score based on the participant's distance from the target. Similarly, they found participants displayed faster learning with punishment feedback compared to reward feedback during adaptation. They also found that participants displayed greater trial-by-trial movement variability when receiving punishment feedback compared to reward feedback. The authors suggested that increased adaptation with punishment feedback is a result of exploration via greater trial-by-trial movement variability. In their study, trial-by-trial movement variability was defined as the difference in reach position (i.e., reach angle) between successive reaches. Yet it remains unclear whether greater trial-by-trial movement variability with punishment feedback was the cause or byproduct of faster adaptation. That is, a greater difference in reach angle between successive trials when participants received punishment feedback could have been caused by greater movement variability or faster adaptation. Thus it is unclear whether punishment feedback leads to greater movement variability that enhances sensorimotor adaptation.

The goal of this paper was to investigate the influence of punishment feedback and reward feedback on movement variability and sensorimotor adaptation. In **Experiment 1**, we adapted our previous experimental protocol (Cashaback et al., 2019) and provided 68 participants either reward feedback or punishment feedback during a sensorimotor adaptation task. Unexpectedly, and unlike previous work, we did not observe a difference in learning rate between punishment and reward feedback. However, participants receiving punishment feedback displayed a greater extent of learning than participants receiving reward feedback. Interestingly, our metrics of movement variability pointed to the idea that punishment feedback may actually decrease movement variability. While greater adaptation has been associated with greater movement variability (Dhawale et al., 2017; Song et al., 2020; Van Mastrigt et al., 2020; Wu et al., 2014), it is difficult to parse whether trial-by-trial changes in reach behaviour are from movement variability or updates in intended reach aim during tasks that require adaptation (Cashaback et al., 2019; Therrien et al., 2016; Therrien et al., 2018; Van Mastrigt et al., 2020; van Mastrigt et al., 2021). To address whether punishment feedback leads to different magnitudes of movement variability compared to reward feedback, in **Experiment 2** we utilized a sensorimotor task that does not require adaptation to be successful (Roth et al., 2023; van Beers et al., 2013). Thus, we were able to more readily observe the influence of reward feedback and punishment feedback on movement variability, while mitigating the influence of adaptive processes. Aligned with the results in the first experiment, again we found that participants displayed decreased movement variability with punishment feedback compared to reward feedback. Collectively, our results support the idea that punishment feedback leads to reduced movement variability compared to reward feedback, which may be linked to sensorimotor adaptation.

2. EXPERIMENTAL PROCEDURES

2.1. Experiments 1 and 2 participants

Across both experiments we collected 102 participants (n = 68 in **Experiment 1** and n = 34 in **Experiment 2**, age: 18–30 yr). Participants reported they were right-handed and free of neuromuscular disease. All participants provided written informed consent to participate and the procedures were approved by McMaster University's Research Ethics Board

(**Experiment 1**) and the University of Delaware's Institutional Review Board (**Experiment 2**).

2.2. Apparatus

For both experiments, participants grasped the handle of a robotic manipulandum (Fig. 1A and Fig. 1C, KINARM, BKIN Technologies, Kingston, ON, Canada) and made reaching movements in the horizontal plane. A semi-silvered mirror blocked vision of both the participant's upper-limb and the robotic manipulandum. Images (start position, targets) from an LCD screen were projected onto the semi-silvered mirror. Hand position was recorded at 1000 Hz and stored offline for analysis.

2.3. General experimental protocol: experiment 1 and 2

With punishment feedback, participants were told they would hear an unpleasant sound (**Supplementary E**), the target would expand and change color (red), and a small amount of money was taken away from their compensation each time they missed the target. With reward feedback, participants were told they would hear a pleasant sound (**Supplementary F**), the target would expand and change color (blue), and they would receive a small amount of money when they hit the target.

Participants in the punishment group were told that base compensation was $10.00 and that they could lose up to $5.00 based on task performance. Participants in the reward group were told that the base compensation was $5.00 and that they could earn up to an additional $5.00 bonus based on task performance. All participants received the full $10.00 after completing the experiment, irrespective of task performance.

2.4. Experiment 1 protocol

The goal of **Experiment 1** was to test whether punishment feedback and reward feedback differentially influence sensorimotor adaptation and movement variability. For this experiment, we utilized our previous motor adaptation task (Cashaback et al., 2019). Participants were presented with images of a start position (white circle, radius = 0.5 cm) aligned with the sagittal plane and a target (white circle, radius = 0.5 cm) located 20 cm forward of the start position. A 30 cm wide white finish line was located 2 cm forward of the target. Participants were instructed to "hit the target" without vision of their hand. For each trial, participants began at the start position, passed through or near the target, and stopped their hand after passing through the finish line that disappeared once crossed. The start position turned yellow after a short, randomized delay (250–750 ms) to signal the beginning of the trial. This small delay allowed us to examine reaction times. We calculated the participant's reach angle relative to the line connecting the start position and target once their hand was 20 cm from the start position (Fig. 1A). After 250 ms, the robotic arm returned the participant's hand to the start position using a minimum jerk trajectory.

Participants performed 450 reaching movements. Participants received no feedback during baseline reaches (trials 1–50). During experimental reaches (trials 51–400), participants received either binary reward feedback or binary punishment feedback. Unknown to participants, the probability of receiving reward feedback or punishment feedback was a function of their reach angle according to their assigned reward landscape or punishment landscape (Fig. 1B; see section below for details). Participants received no feedback during the washout trials (trials 401–450).

2.5. Reward and punishment landscapes

During experimental trials (trials 51–400), participants were exposed to either a reward landscape or a punishment landscape (Fig. 1B). As in our previous work (Cashaback et al., 2019), to generate these landscapes we manipulated the probability of receiving feedback as a function of their reach angle for a given trial. Thus, participants had to change their reach angle to maximize their probability of success (i.e., avoid punishment feedback or receive reward feedback). Binary punishment feedback or reward feedback is advantageous because it limits the ability to form a vectored error signal over multiple trials, unlike scalar punishment feedback or reward feedback that varies in magnitude according to the distance from a target.

The punishment landscape or reward landscape experienced by a participant was normalized to their baseline variability (Cashaback et al., 2019). Reach angle was defined as the angle made between the hand and the vertical line connecting the start position and the target, where the start position was the center of rotation (Fig. 1A). On each trial, reach angle was calculated when the hand was 20 cm away from the start position. Reach angles were normalized to the last 25 baseline trials and expressed as a z-score (Cashaback et al., 2019). A z-score of 0.0 corresponded to the participant's average baseline reach angle. A z-score of 1.0 or −1.0 indicated that a reach angle was ± 1 standard deviation away from their average baseline reach angle in the clockwise or counterclockwise direction, respectively.

The direction of each feedback landscape represents the direction towards the optimal reach angle. The clockwise reward (R^{CW}) landscape can be summarized as

$$R(\theta)^{CW} = p(r = 1|\theta_i)^{CW} = \begin{cases} \frac{\theta_i}{6} + \frac{1}{2}; & -3 \leqslant \theta_i \leqslant 3 \\ 1; & 3 < \theta_i \leqslant 6 \\ 0; & \text{otherwise} \end{cases} \quad (1)$$

Where p is the probability of receiving reward and θ_i is the reach angle for trial i. Here, $r = 1$ denotes a successful reach. Similarly, the clockwise punishment (P^{CW}) landscape can be summarized as

$$P(\theta)^{CW} = p(r = 0|\theta_i)^{CW} = \begin{cases} \frac{1}{2} - \frac{\theta_i}{6}; & -3 \leqslant \theta_i \leqslant 3 \\ 0; & 3 < \theta_i \leqslant 6 \\ 1; & \text{otherwise} \end{cases} \quad (2)$$

Where p is the probability of receiving punishment feedback, and θ_i is the reach angle for trial i. Here, $r = 0$ denotes an unsuccessful reach.

Experiment 1
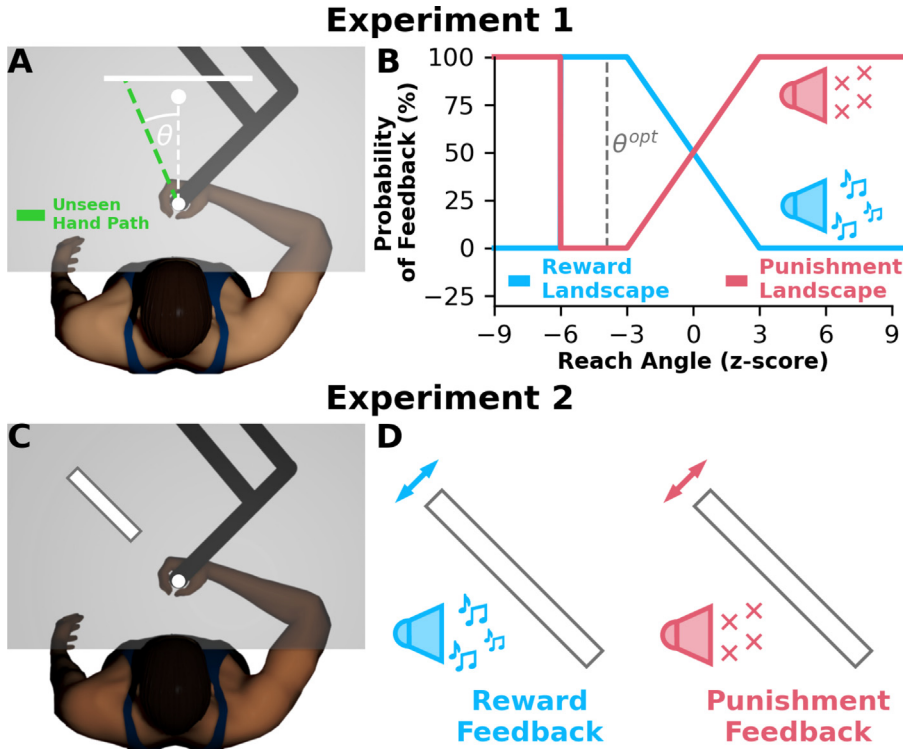
Experiment 2

Fig. 1. **Experiment Design. (A, C)** In both **Experiment 1** and **2**, participants grasped the handle of a robotic manipulandum and made reaching movements in the horizontal plane. An LCD display projected images (start position, targets) onto a semi-silvered mirror that occluded vision of the hand and upper arm. **(A)** The goal of **Experiment 1** was to examine how reward feedback and punishment feedback influence sensorimotor adaptation and movement variability. Participants were instructed to reach from the start position (close white circle) and hit the target circle (far white circle). A long white bar positioned above the target disappeared after participants reached through it, signaling the end of the trial. Participants that experienced the reward landscape received reward feedback (pleasant sound, target expands, monetary reward) if they hit the target. Participants that experienced the punishment landscape received punishment feedback (unpleasant sound, target expands, monetary loss) if they missed the target. We recorded their reach angle (θ) on each trial. Reach angles were normalized to individual baseline movement variability and expressed as a z-score. **(B)** Unbeknownst to participants, we directly manipulated the probability of receiving reward feedback or punishment feedback (y-axis) based on a participant's normalized reach angle (z-score; x-axis) according to the assigned reward landscape (blue) or punishment landscape (red). Reward and punishment landscapes promote participants to change their reach angle to maximize success (θ^{opt}) by respectively maximizing positive reward or minimizing punishment. **(C)** The goal of **Experiment 2** was to examine how reward feedback and punishment feedback influenced movement variability while mitigating the influence of adaptation. Accordingly, we used a motor task that did not require changes in average movement behaviour to successfully complete the task. Participants were told to reach from the start position (white circle) and stop anywhere within the virtually displayed target (white rectangle). **(D)** Participants received only reward feedback in a block of experimental trials and only punishment feedback in the other block of experimental trials. In the reward block of experimental trials, participants received reward feedback if they successfully stopped within the target. In the punishment block of experimental trials (red), participants were told they would receive punishment feedback if they missed the target. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

The counterclockwise reward (R^{CCW}) and counterclockwise punishment (P^{CCW}) landscapes are mirror images of the clockwise landscapes, reflected about the average baseline reach angle (0.0 z-score). The counterclockwise landscapes are summarized as

$$R(\theta)^{CCW} = p(r = 1|\theta_i)^{CCW} = \begin{cases} \frac{1}{2} - \frac{\theta_i}{6}; & 3 \geqslant \theta_i \geqslant -3 \\ 1; & -3 > \theta_i \geqslant -6 \\ 0; & \text{otherwise} \end{cases} \quad (3)$$

$$P(\theta)^{CCW} = p(r = 0|\theta_i)^{CCW}$$
$$= \begin{cases} \frac{\theta_i}{6} + \frac{1}{2}; & 3 \geqslant \theta_i \geqslant -3 \\ 0; & -3 > \theta_i \geqslant -6 \\ 1; & \text{otherwise} \end{cases} \quad (4)$$

Landscape direction, clockwise or counterclockwise, was counterbalanced within each group.

2.6. Experiment 2 protocol

The goal of **Experiment 2** was to determine whether punishment feedback and reward feedback influenced movement variability. We used a motor task that does not require participants to change their reach angle (Roth et al., 2023), thus limiting the influence of adaptive changes in hand position that would artificially increase estimates of movement variability. Participants were presented with virtual images of a start position (white circle, radius = 0.75 cm) that was aligned with the sagittal plane and approximately 15 cm away from their body. The center of the displayed target was located 45 degrees to the left of the sagittal plane and 15 cm away from the start position (Fig. 1C). Rectangular targets were rotated so that their major axis aligned with movement from the start position to the target. For each trial, participants began from the start position and were instructed to "reach and stop inside the target." The start position turned yellow after a short, randomized delay (250–1000 ms) to signal the beginning of the trial. Final hand position was defined as the participant's hand location after their hand velocity went below 0.045 cm/s for 100 ms. One second after stopping, the robot used a minimum jerk trajectory to return their hand to the start position.

Participants performed 50 baseline trials, 200 experimental trials, 50 washout trials, and another 200 experimental trials. During baseline and washout trials, participants reached towards and attempted to stop within a small white circle (radius = 0.5 cm). Participants saw a small yellow dot (radius = 0.25 cm) at their final hand position for the first 40 baseline and washout trials. No feedback was given for the final 10 baseline or washout trials. During the first and second block of experimental trials, participants reached towards and attempted to stop

within a large rectangular target. The major axis of the target was 12 cm (Roth et al., 2023; van Beers et al., 2013). The minor axis length (0.65σ, 0.99 $\pm$ 0.34 cm) was proportional to each participant's lateral movement variability during the last 10 baseline trials (Roth et al., 2023). Participants received only punishment feedback or reward feedback for the first block of experimental trials. If a participant received punishment feedback for missing the target during the first block of experimental trials (trials 51–250), they would receive reward feedback when they hit the target during the second block of experimental trials (trials 301–500). Conversely if they received reward feedback on the first block of experimental trials, they would receive punishment feedback on the second block of experimental trials. The ordering of punishment feedback or reward feedback on the first and second block of experimental trials was counterbalanced.

2.7. Data analysis

We performed data analysis using custom Python 3.8.13 scripts. For **Experiment 1**, we performed all analyses on reach angles (z-score). For **Experiment 2**, final hand position coordinates were projected onto a rotated coordinate system that was aligned with the major and minor axes of the rectangular target. Thus the x-axis and y-axis of the rotated system were aligned with the minor and major axes of the long rectangular target, respectively. The origin of this rotated coordinate system was the center of the rectangular target.

2.8. Quantifying adaptive behaviour

2.8.1. Changes in reach angle. As in our previous work (Cashaback et al., 2019), for **Experiment 1** we compared participant behaviour by averaging group reach angles during early learning (trials 51–100), late learning (trials 351–400), and washout (trials 401–450). Comparing participant behaviour during these time windows provides a direct way to analyze learning rate (early learning), learning extent (late learning), and retention (washout). To perform comparisons across landscape directions, we multiplied normalized reach angles by -1 for all participants that experienced a counterclockwise feedback landscape (Acerbi et al., 2014; Cashaback and Cluff, 2015; Cashaback et al., 2019).

2.8.2. Optimal reach angle. We define optimal aim point θ_{opt}^{aim} as the angle that the participants should aim to maximize the utility U.

$$\theta_{opt}^{aim} = \underset{\theta^{aim} \in \Theta}{\operatorname{argmax}}\{E[U(\theta^{aim}, \sigma^2)]\} \qquad (5)$$

where, θ^{aim} is the reach aim and Θ is the set of all possible reach angles. The expected utility $E[U(\theta^{aim}, \sigma^2)]$ for any given unbiased aim point θ^{aim} is calculated as

$$E[U(\theta^{aim}, \sigma^2)] = \int p(\theta|\theta^{aim}, \sigma^2)L(\theta)d\theta \qquad (6)$$

where, $L(\theta)$ is either the reward landscape $R(\theta)$ or punishment landscape $P(\theta)$. Here we are interested in

the expected utility over all possible reach angles ($\int d\theta$). The probability of reaching at any angle θ with aim point θ^{aim} was modeled as a normal distribution

$$p(\theta|\theta^{aim}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\theta-\theta^{aim}}{2\sigma^2}} \qquad (7)$$

where σ^2 is the reach angle variance. σ^2 was estimated by considering both motor (execution) variance σ_m^2 and exploration variance σ_e^2. Pekny and colleagues (2015) proposed that the magnitude of exploration variability is inversely related to the probability of hitting the target. We manipulated the probability of hitting the target $p(r = 1|\theta)$ as a function of reach angle according to the assigned feedback landscape. Thus, we approximated σ^2 by considering two potential sources of movement variability and the probability of hitting the target.

$$\sigma^2 = \sigma_m^2 + [1 - p(r = 1|\theta)]\sigma_e^2) \qquad (8)$$

Here, motor (execution) variance is constant and applied on every trial. The exploration variance is scaled inversely with the probability of hitting the target. The values of $\sigma_m^2(0.87)$ and $\sigma_e^2(1.06)$ were obtained using the method from Cashaback et al., 2019. Using these values, we obtained an optimal reach angle ($\theta_{opt}^{aim} = 3.90$).

2.9. Quantifying movement variability

2.9.1. Movement variability: trial-by-trial difference. We (Cashaback et al., 2019; Roth et al., 2023; van der Kooij et al., 2023) and others (Chen et al., 2017; Holland et al., 2018; Pekny et al., 2015; Therrien et al., 2016; Sidarta et al., 2018; Sidarta et al., 2022; van der Kooij and Smeets, 2019; Van Mastrigt et al., 2020) have found that movement variability is modulated by task success. For **Experiment 1**, as an estimate of movement variability we calculated the standard deviation of the trial-by-trial difference in reach angle. For **Experiment 2**, we calculated the standard deviation of the trial-by-trial differences in final hand position separately along the dimensions aligned with the minor and major axes of the long rectangular target during baseline and experimental conditions.

For both experiments, baseline trial-by-trial differences (Δ) were calculated during the final 10 trials where no feedback was given to the participant (Eq. 9). Experimental condition movement variability was calculated independently for successful (i.e. hitting the target, Eq. 10) and unsuccessful trials (i.e. missing the target, Eq. 11).

$$\Delta X^{baseline} = X_{t+1} - X_t \qquad (9)$$

$$\Delta X^{r=1} = X_{t+1} - X_t^{r=1} \qquad (10)$$

$$\Delta X^{r=0} = X_{t+1} - X_t^{r=0} \qquad (11)$$

Here, X represents the participant's reach angle (**Experiment 1**), final hand position along the major axis of the displayed rectangular target (**Experiment 2**), or final hand position along the minor axis of the displayed rectangular target (**Experiment 2**). t is the trial number,

and r represents if a trial was successful (r = 1) or unsuccessful (r = 0).

2.9.2. Movement variability: detrended trial-by-trial differences. Past work has suggested that punishment feedback induces greater trial-by-trial differences in reach position (Eq. (9)–(11)), which is commonly used as a metric of movement variability (Song et al., 2020). However, such trial-by-trial differences as a metric of movement variability would also be influenced by changes in behavior caused by adaptation. That is, adaptive processes that update reach aim would also lead to increased trial-by-trial differences independent of isolated movement variability caused by motor noise (Faisal et al., 2008; Jones et al., 2002; van Beers et al., 2004) or exploratory noise (Cashaback et al., 2019; Pekny et al., 2015). To mitigate the influence of adaptation on trial-by-trial differences, we also analyzed trial-by-trial movement variability after detrending the reach angles. Detrending the reach angles during the experimental condition provided a way to observe trial-by-trial movement variability while mitigating the influence of reach aim updates altering behaviour.

To detrend the data, we used a central moving average-subtraction method (Hyndman et al., 2018) with a bin size of 15 trials. Specifically, for every trial, we took an average of a small window of 15 trials centered about a particular trial. We subtracted this average from this particular trial before sliding the window to the next trial. We repeated this for every trial in the learning block (trials 51–400). Note that as a result of this technique, we lose the first and last 7 trials due to lack of data points to calculate the average. We then calculated movement variability for the early learning (trials 51–100) and late learning (trials 351–400). We also tested different bin sizes used for the moving average (5–25 trials). Using too short of a window to calculate the moving average risks removing too much of the trial-to-trial variability. Conversely, using too long of a window to calculate the moving average may not remove adaptive processes that update reach aim.

2.9.3. Movement variability: distribution of final hand positions (IQR). Past studies (Buzzi et al., 2019; Cusumano and Cesari, 2006; Latash et al., 2002; Lokesh and Ranganathan, 2019; Bernstein, 1967; Scholz and Schoner, 1999; Roth et al., 2023) have quantified movement variability differences between task-irrelevant ('uncontrolled manifold' or 'null space') and task-relevant ('orthogonal dimension'), which in this study respectively correspond to the major-axis and minor-axis of the rectangular target. Similarly and inline with our past work (Cashaback and Cluff, 2015; Roth et al., 2023), in **Experiment 2** we quantified the magnitude of movement variability of final hand position by calculating the interquartile range (IQR). IQR is known to be a robust measure of variability (Kaltenbach, 2012) because it is not heavily influenced by outliers. We did not examine IQR in **Experiment 1** because IQR would be conflated with adaptation. We calculated IQR as the difference between the 25th and 75th percentiles. We took the IQR ratio between conditions to describe the relative movement variability in final hand position. A value of one represents equal movement variability of final hand position between conditions. We calculated the IQR ratio between the major (task-redundant) and minor (task-relevant) axes between the reward and punishment conditions (Roth et al., 2023). Statistical comparisons were made between the mean IQR ratio to a value of one.

2.9.4. Quantifying exploratory behaviour with lag-1 autocorrelation. We quantified the level of exploration along the solution manifold by calculating lag-1 autocorrelations on trial-by-trial final hand position (Roth et al., 2023; van Beers et al., 2013). Here, a larger lag-1 autocorrelation suggests greater exploratory behaviour. For each condition in **Experiment 2**, we performed lag-1 autocorrelation analysis separately along the major and minor axes of the rectangular target.

2.9.5. Statistical analyses. Non-parametric bootstrap hypothesis tests (1,000,000 iterations) were used for follow-up mean comparisons (Cashaback et al., 2019; Calalo et al., 2023; Cashaback et al., 2017; Gribble and Scott, 2002; Lokesh et al., 2022; Roth et al., 2023). We used directional tests when testing theory-driven predictions, and nondirectional tests otherwise. Spearman Rank correlation was used for all correlation analyses to capture monotonic relationships (Hauke and Kossowski, 2011). We computed common language effect sizes ($\hat{\theta}$) for all comparisons (Calalo et al., 2023; Cohen, 1988; McGraw and Wong, 1992; Roth et al., 2023; Lokesh et al., 2023). Statistical tests were considered significant at $p < 0.05$.

3. RESULTS

3.1. Experiment 1

The goal of **Experiment 1** was to observe whether binary punishment feedback and binary reward feedback differentially influence sensorimotor adaptation. For each trial, participants began at a start position and attempted to pass their hand through a small circular target (Fig. 1A). The trial ended when a white line beyond the target disappeared once crossed and the robot brought the participant's hand back to the start position.

Participants were placed in either the reward group (n = 32) or punishment group (n = 32). Participants completed 50 baseline trials with no feedback, 350 experimental trials with feedback, and 50 washout trials with no feedback. Participants that experienced the punishment landscape were told they would receive punishment feedback (unpleasant sound, target would expand, small monetary loss) if they missed the target and no feedback if they hit the target. Participants that experienced the reward landscape were told they would receive reward feedback (pleasant noise, target expands, small monetary gain) if they successfully passed through the target and no feedback if they missed the target.

3.2. Individual behaviour

Reach angle over trials for an individual experiencing a reward landscape and another participant experiencing a punishment landscape are shown in Figs. 2A and **2B**, respectively. Both participants learned to adjust their reach angle to increase the probability of success by approaching the optimal reach angle (θ^{opt}). For the participant experiencing the punishment landscape, changing reach angle to approach the optimal reach angle decreased the probability of punishment feedback. Conversely for the participant experiencing a reward landscape, changing reach angle to approach the optimal reach angle increased the probability of reward feedback. Each participant displayed clear between trial changes in reach angle, which would reflect both adaptive changes and movement variability. For the displayed participants, the participant experiencing the punishment landscape had comparatively less between trial changes in reach angle compared to the participant experiencing the reward landscape.

3.3. Group behaviour

3.3.1. Adaptation. At the group level, we found that participants experiencing either the punishment landscape and reward landscape adapted their reach angle to respectively decrease or increase the probability of punishment feedback and reward feedback (Fig. 2C).

We compared average reach angles between participants experiencing the punishment landscape and reward landscape during early learning (trials 51–100), late learning (trials 351–400), and washout (trials 401–450). Unlike past work (Galea et al., 2015; Song et al., 2020), we did not find significant differences in reach angle during early learning between participants experiencing the punishment landscape and reward landscape ($p = 0.454$, $n = 34$ per group). This suggests that the reward and punishment groups adapted at a similar rate during early learning.

Participants experiencing the punishment landscape displayed significantly greater average reach angles during late learning compared to participants experiencing the reward landscape ($p < 0.001$), suggesting that punishment feedback leads to a greater extent of learning that is closer to the optimal solution. This finding was robust when comparing the median reach angles between groups during late learning ($p < 0.001$). We also found that participants experiencing the punishment landscape had significantly greater reach angles during washout ($p = 0.018$) compared to participants experiencing the reward landscape. However, when normalizing washout reach angle to average late learning reach angle there was no difference in retention ($p = 0.372$) between participants that experienced the punishment landscape or reward landscape. Thus, our work does not support the previous finding that reward feedback leads to greater retention compared to punishment feedback (Galea et al., 2015).

As noted above, we found that participants experiencing the punishment landscape displayed a greater extent of learning during late learning trials. Past work has suggested that punishment feedback leads to greater movement variability and subsequently faster learning (Song et al., 2020), and more generally that greater movement variability leads to faster learning (Wu et al., 2014). While we did not find faster early adaptation between participants experiencing punishment and reward landscapes, we did observe a greater extent of learning. It is possible that greater movement variability induced by punishment feedback could have led to a greater extent of learning during late learning trials. Thus, we then examined whether there were differences in movement variability between participants that experienced the punishment landscape and reward landscapes.

3.3.2. Movement variability. Past literature (Cashaback et al., 2019; Chen et al., 2017; Holland et al., 2018; Roth et al., 2023; Pekny et al., 2015; Therrien et al., 2016; Sidarta et al., 2018; Sidarta et al., 2022; van der Kooij and Smeets, 2019; Van Mastrigt et al., 2020; van der Kooij et al., 2023) has shown that movement variability is modulated by task outcome. Specifically, movement variability is greater after an indicated target miss compared to after an indicated target hit. As a reminder, participants experiencing the punishment landscape were told they would receive punishment feedback for missing the target and no feedback for hitting the target. Conversely, participants experiencing the reward landscape were told they would receive no feedback for missing the target and reward feedback for hitting the target. As expected, we found that participant movement variability increased after an indicated target miss relative to an indicated target hit for participants experiencing the reward landscape ($p < 0.001$) and participants experiencing the punishment landscape ($p < 0.001$).

More importantly, we wanted to compare movement variability between participants experiencing the punishment landscape and reward landscape. During early learning (Fig. 3A), we found no difference in trial-by-trial movement variability between participants experiencing the punishment landscape and those experiencing the reward landscape with an indicated target hit ($p = 0.960$) or target miss ($p = 0.696$). Likewise during late learning, we did not find differences in trial-by-trial movement variability between participants experiencing the punishment landscape and those experiencing the reward landscape with an indicated target hit ($p = 0.232$) or target miss ($p = 0.250$). A lack of trial-by-trial movement variability differences between participants experiencing the punishment landscape and reward landscape does not support the previously suggested idea that punishment feedback leads to greater movement variability that enhances learning (Song et al., 2020).

However, as noted above, examining trial-by-trial movement variability would be influenced by both
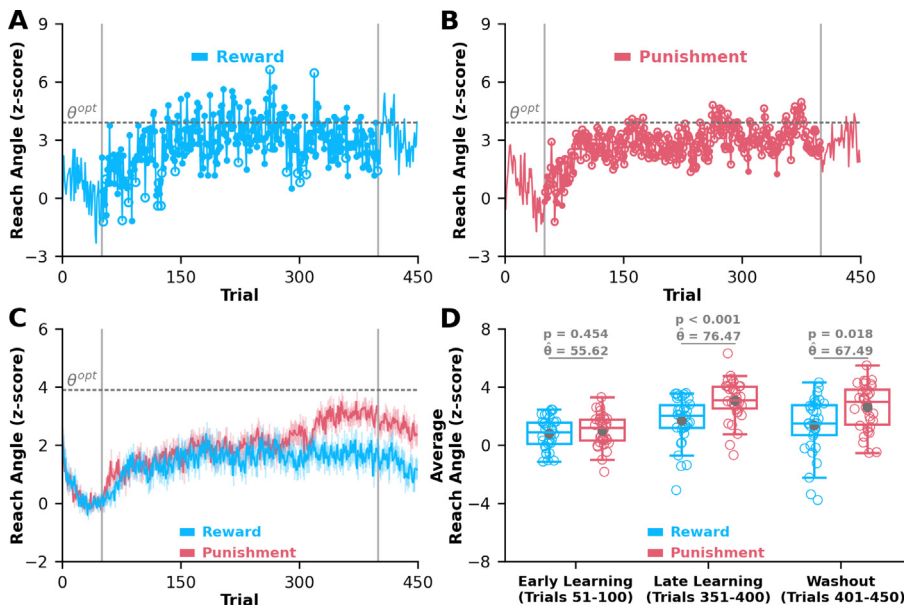
Fig. 2. Sensorimotor Adaptation in Experiment 1. Here we show reach angle (y-axis) per trial (x-axis) for a single participant that experience the **(A)** reward landscape (blue), a single participant that experienced the **(B)** punishment landscape (red). Solid grey lines separate baseline trials (trials 1–50), experimental trials (trials 51–400) and washout trials (trials 401–450). Solid circles represent trials where participants received reward or punishment feedback. Unfilled circles represent trials where participants did not receive feedback. **(C)** Here we show the average reach angle across participants for each group. Shaded areas represent ± 1 SE. Dashed horizontal lines (gray) represent the optimal reach angle (θ^{opt}) that maximizes reward or minimizes loss. **(D)** We characterized learning rate, learning extent, and retention by calculating the average reach angle (y-axis) within the respective 50 trial block (x-axis): early learning (trials 51–100), late learning (trials 350–400), and washout (trials 401–450). Unfilled circles represent individual data. Solid grey circles represent mean reach angle for the group. Boxplots represent the 25th, 50th, and 75th percentiles. We found that participants experiencing the punishment landscape displayed significantly greater reach angles during the late learning block (p < 0.001), where this behaviour carried over to the washout block (p = 0.018). Our finding shows that punishment feedback leads to a greater extent of sensorimotor learning. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

adaptation and movement variability. Thus, despite no differences in trial-by-trial movement variability (the difference in reach angle between successive trials), this finding does not completely remove the possibility that punishment feedback differentially influences movement variability compared to reward feedback. For example, one could get similar levels of trial-by-trial movement variability if one of the forms of feedback caused larger (or smaller) movement variability with smaller (or larger) trial-level adaptation.

Next we examined whether there was a relationship between movement variability and adaptation. Specifically, we analyzed the relationship between trial-by-trial movement variability following an indicated target miss with either average learning in the early learning (trials 51–100) or late learning (trials 351–400) blocks. We focused on the movement variability following an indicated target miss, where we would expect minimal influence of adaptation between these successive trials (Cashaback et al., 2019; Roth et al., 2023; Therrien et al., 2016; Therrien et al., 2018). For those experiencing the reward landscape, we found no relationship between trial-by-trial movement variability following an indicated target miss and early learning (p = 0.099) or late learning

(p = 0.114). For those experiencing the punishment landscape, we found a relationship between trial-by-trial movement variability following an indicated target miss and early learning (p = 0.045, $\rho = 0.346$). However, as a reminder, we did not find significant differences in average reach position between the reward and punishment groups in the early learning block. We found a significant relationship in the punishment group between trial-by-trial movement variability following an indicated target miss and late learning (p = 0.020, $\rho = 0.409$; Fig. 3E). Specifically, learning extent monotonically increased with trial-by-trial movement variability following an indicated miss. Although movement variability following an indicated target miss is not different between the reward group and punishment group, it is possible that the sensorimotor system has more knowledge of the movement variability and or there is greater sensitivity to updating reach aim following a successful trial in the presence of punishment feedback. That is, participants experiencing the punishment landscape may have better utilized movement variability to update reach aim following an indicated target hit compared to those experiencing the reward landscape, which resulted in greater adaptation during the late learning trials.

As a way to mitigate the influence of adaptation on a metric of movement variability, we detrended each participant's reach angles. Specifically, we detrended each participant's reach angles with a central moving average subtraction method. Fig. 3C shows the average detrended data (bin size = 15 trials) across participants that experienced the punishment landscape and reward landscape. For both the early learning and late learning trials, we took the standard deviation of the detrended reach angles. As before, we found no significant difference in detrended movement variability during the early learning trials (p = 0.88). However, for the late learning trials, we found significantly less detrended movement variability for those experiencing the punishment landscape compared to those experiencing the reward landscape (p = 0.043; bin size = 15; Fig. 3D). Additionally and aligned with the analysis above, we found that detrended movement variability was related to adaptation during the late learning trials for the participants experiencing the punishment landscape (p = 0.033; $\rho = 0.366$; Fig. 3F) but not those experiencing the reward landscape (p = 0.115; $\rho = 0.275$), further supporting the idea that the sensorimotor system may have more knowledge of
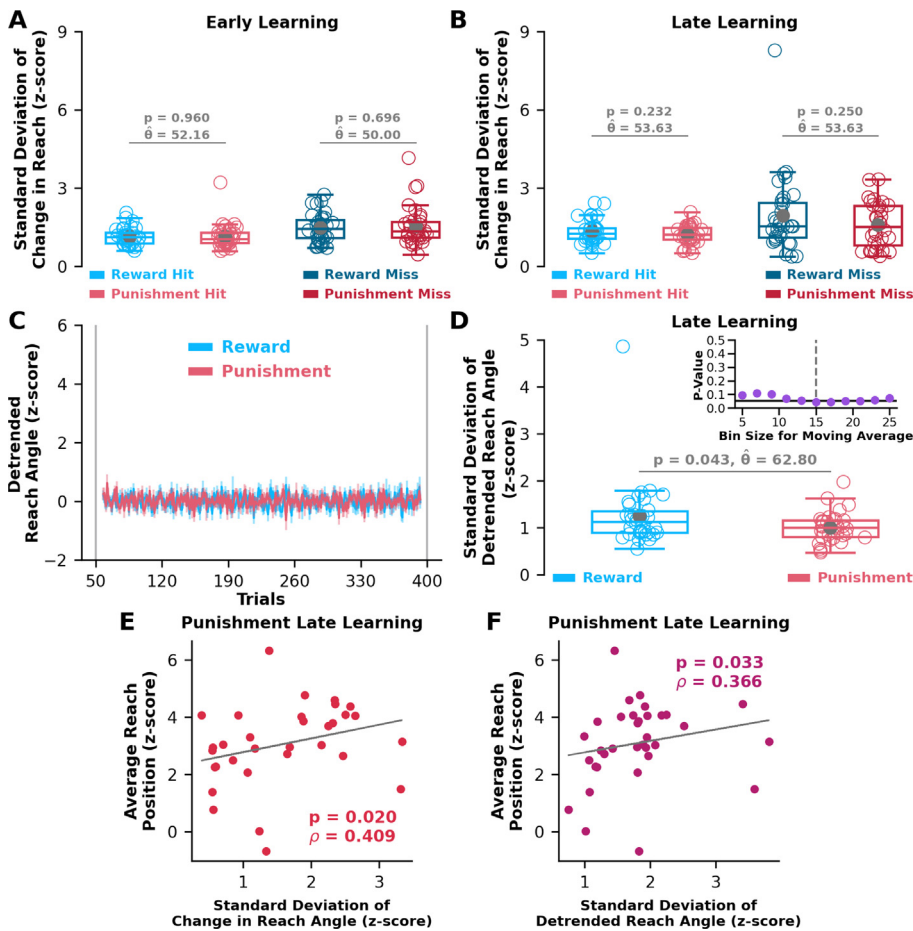
Fig. 3. Movement Variability in Experiment 1. We calculated the standard deviation between changes in reach angle separately following successful (lighter shades) and unsuccessful (darker shades) trials to assess movement variability. **A)** We found no difference in trial-by-trial movement variability between reward feedback (blue) and punishment feedback (red) following a hit (p = 0.960) or miss (p = 0.696) in early learning trials. **B)** Likewise, we found no difference in trial-by-trial movement variability between reward feedback (blue) and punishment feedback (red) following a hit (p = 0.232) or miss (p = 0.250) in late learning trials **(C)** Detrended reach angle (y-axis) per trial (x-axis) for the detrended moving average when using the reach angles shown in ((**A**). Specifically, here we used a central moving average subtraction (15 trial bin size) to detrend individual data, to limit the influence of changes in reach aim due to adaptive behaviour. **D)** As a proxy of movement variability, we measured the standard deviation of the detrended reach angles separately for participants that experienced a reward landscape (blue) or a punishment landscape (red). With the detrended data (15 trial bin size), participants that experienced a punishment landscape displayed significantly lower movement variability in the late learning block than participants that experienced a reward landscape (p = 0.043). The inset displays the p-value (y-axis) when using different bin sizes (x-axis) for the moving average. All but one bin size was below a p-value of 0.1, with several below 0.05. These results do not support the hypothesis that punishment feedback leads to faster learning by increasing movement variability. **(E)** We found a significant positive correlation (p = 0.020, ρ = 0.409) between trial-by-trial movement variability (x-axis) and average reach position (y-axis) during the late learning block of the punishment group. **(F)** This monotonically increasing trend held when using the detrended reach angle (p = 0.033, ρ = 0 366). Note that the trend line shown in **E, F** are only visual and are not meant to suggest a linear relationship between average reach position and movement variability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

movement variability and or there is greater sensitivity to updating reach aim following a successful trial in the presence of punishment feedback. For the detrending analysis, we then considered the influence of bin size (5–25 trials). In general, we found that all but one bin size was below a p-value of 0.1, with several below 0.05 (Fig. 3D, inset). Note that we are unable to partition the standard deviation of the detrended reach angles based

on indicated target hits and misses, due to a loss of trials from the bounds of the moving average when performing the detrending process (see **Supplementary D1**). Thus, while not conclusive, the results of the detrended movement variability analysis point to the idea that punishment feedback may actually decrease movement variability.

We did find a potential link between movement variability and the extent of learning when given punishment feedback. However, the results in **Experiment 1** do not support the idea that punishment feedback leads to comparatively greater movement variability than reward feedback to enhance sensorimotor adaptation. Rather, our findings point to the notion that punishment feedback may decrease movement variability.

Recent literature has also examined the effects of reinforcement (reward) feedback and task success on movement vigor (Mazzoni et al., 2007; Panigrahi et al., 2015; Shadmehr et al., 2019; Sukumar et al., 2021; Summerside et al., 2018). These studies use reaction times and movement times as proxies of movement vigor. We found that participants in the punishment group displayed significantly slower reaction times (p = 0.045) and movement times (p < 0.023) compared to the reward group after an indicated target miss. That is, participants displayed lower movement vigor after receiving punishment feedback (see **Supplementary C**).

3.4. Experiment 2

Contrary to recent findings in the literature (Galea et al., 2015; Song et al., 2020), in **Experiment 1**, we did not find faster adaptation with punishment feedback compared to reward feedback. However, we did find a greater extent of learning with punishment feedback. As suggested by previous work (Song et al., 2020), it is possible the punishment feedback leads to greater movement variability that can enhance learning. Yet we did not find significant differences in trial-by-trial movement variability between participants experiencing punishment and reward landscapes. Rather, when detrending the data to mitigate the potential influence of adaptation on a metric

of movement variability, the results were more suggestive of punishment feedback leading to reduced movement variability. The goal of **Experiment 2** was to better isolate the influence of punishment feedback and reward feedback on movement variability by using a motor task that does not require adaptation.

In **Experiment 2**, for each trial participants began their reach in a start position and attempted to stop within a virtually displayed target without vision of their hand. For each reach, we recorded their final hand position when they stopped within or outside the virtually displayed target (Roth et al., 2023). Participants performed 50 baseline trials, 200 experimental trials, 50 washout trials, and then another 200 experimental trials. During each of the experimental trial blocks, participants were informed that they would receive punishment feedback (unpleasant sound, target expands, monetary loss) for missing the target or reward feedback (pleasant sound, target expands, monetary gain) for hitting the target. If a participant received punishment feedback (or punishment feedback) during the first block of experimental trials (trials 51–250), they would receive reward feedback (or reward feedback) during the second block of experimental trials (trials 301–500).

3.4.1. Individual behaviour. Final hand positions for a participant when receiving punishment feedback and reward feedback are shown in Figs. 4A. Fig. 4B shows the corresponding final hand positions along the minor axis of the target. This participant displayed less movement variability (smaller spread of final hand position) along the minor axis of the target when receiving punishment feedback. **Supplementary A** shows all the results of **Experiment 2** along the major axis of the target.

3.4.2. Group behaviour. As in the first experiment and past work (Cashaback et al., 2019; Chen et al., 2017; Holland et al., 2018; Roth et al., 2023; Pekny et al., 2015; Sidarta et al., 2018; Sidarta et al., 2022; Therrien et al., 2016; van der Kooij and Smeets, 2019; Van Mastrigt et al., 2020), we found that movement variability was greater after a target miss compared to a target hit with reward feedback (p < 0.001) and punishment feedback (p < 0.001). Crucially in **Experiment 2**, and in support of the detrended reach angle analysis in **Experiment 1**, we found that participants displayed significantly less trial-by-trial movement variability with punishment feedback compared to reward feedback (Fig. 4C), following either a target hit (p = 0.016) or a target miss (p = 0.022).

As an additional metric of movement variability, we then calculated the interquartile range to examine the distribution of final hand positions along the minor axis of the target. Specifically, we found the ratio of interquartile range between the punishment and reward feedback conditions as a measure of their relative movement variability (Fig. 4D). Participants displayed an interquartile range ratio significantly greater than 1 (p = 0.034), again suggesting that punishment feedback leads to less movement variability compared

to reward feedback. Finally, as in our recent work (Roth et al., 2023), we examined lag-1 autocorrelations as a metric of motor exploration, but did not find any differences when participants received reward and punishment feedback (Fig. 4E). We also did not find a difference in lag-1 autocorrelation between groups when partitioning final hand positions based on indicated target hits or misses (Roth et al., 2023; **Supplementary B)**.

As in **Experiment 1**, we wanted to look at the effects of reward and punishment on reaction and movement times. Aligning with the results of **Experiment 1**, we found that participants in the punishment condition displayed significantly slower reaction times (p = 0.005) compared to the reward condition after an indicated target miss (see **Supplementary C**).

Collectively, the results across both **Experiment 1** and **2** suggest that punishment feedback leads to less movement variability compared to reward feedback. Below we further discuss the potential influence of punishment feedback and reward feedback on both movement variability and sensorimotor adaptation.

4. DISCUSSION

In this paper, we investigated how reward feedback and punishment feedback influence motor learning and movement variability. Contrary to recent findings, we did not find that punishment compared to reward leads to a faster learning rate (Galea et al., 2015; Song et al., 2020) that is linked to greater movement variability (Song et al., 2020). Rather, our findings point to the notion that punishment feedback decreases movement variability. As discussed below, a greater extent of learning may have been caused by greater knowledge of movement variability and or increased sensitivity to updating reach aim following punishment.

In **Experiment 1**, participants learned to adjust their reach aim in response to an imposed punishment or reward landscape. We found no difference in reach angles during early learning trials, suggesting punishment feedback and reward feedback cause a similar learning rate. However, participants experiencing punishment feedback did have a significantly greater reach angle that was closer to the optimal reach angle during the late learning trials, suggesting that punishment feedback leads to a greater extent of learning. When detrending the data, we found that punishment feedback appeared to cause lower movement variability, which was related to the extent of learning. However, it is difficult to parse whether trial-by-trial changes in reach behaviour are from movement variability or updates in intended reach aim during tasks that require adaptation. To better isolate the influence of punishment feedback and reward feedback on movement variability, for **Experiment 2** we used a motor task that does not require adaptation. Participants reached to a large rectangular target while receiving only binary punishment feedback or reward feedback. Our **Experiment 2** results show that participants decreased their movement variability when receiving punishment feedback compared to reward feedback.
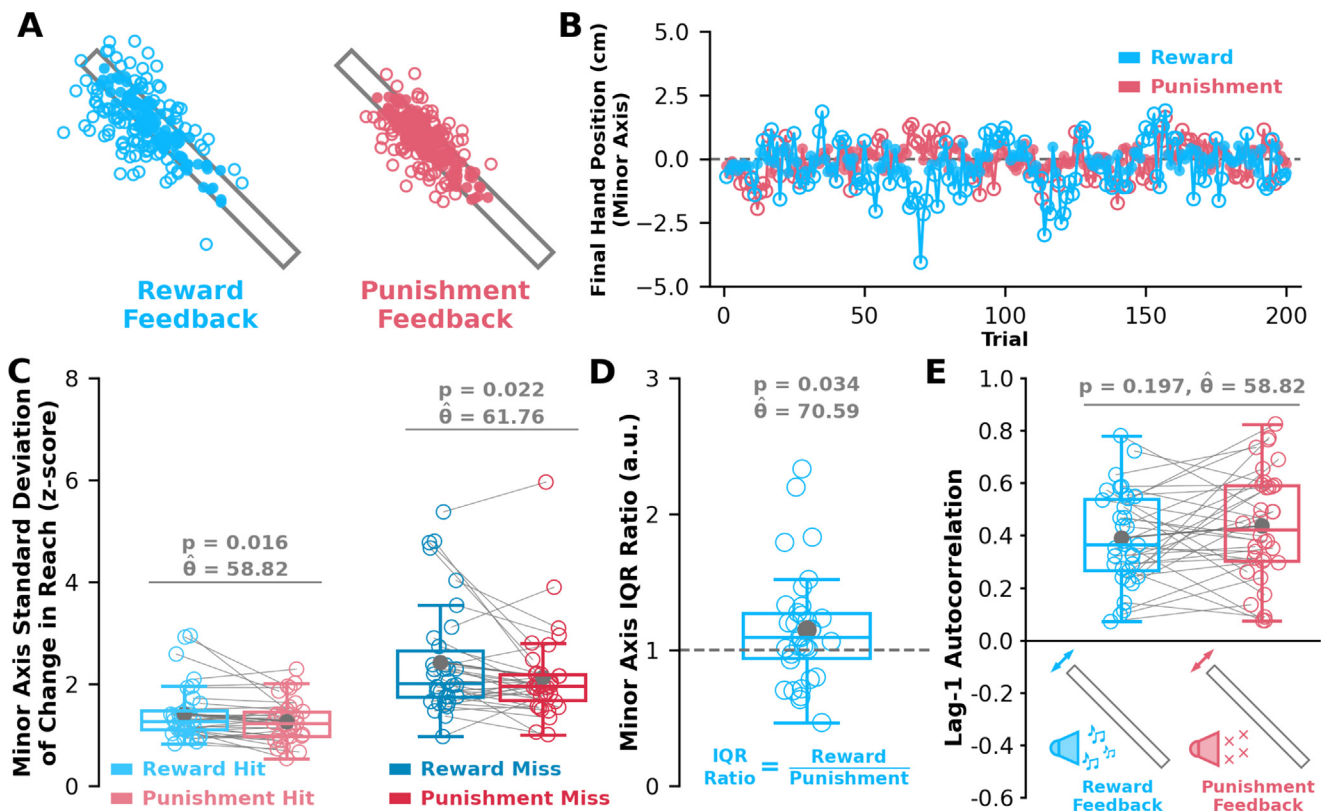
Fig. 4. Movement Variability in Experiment 2. A) Successful (filled circle) and unsuccessful (unfilled circle) reaches by an individual participant performing the reward feedback (blue) and punishment feedback (red) conditions. **B)** Corresponding final hand position coordinates (y-axis) along the minor axis of the target for each trial (x-axis). **C)** We calculated movement variability in each condition separately following successful trials (Hit, dark colours) and unsuccessful trials (Miss, light colours). Final hand position was normalized to baseline and expressed as a z-score. We defined movement variability as the standard deviation of the trial-by-trial change in final hand position. Participants displayed significantly lower movement variability with punishment feedback (red) following either a hit ($p = 0.016$) or a miss ($p = 0.022$) compared to reward feedback (blue). **D)** We calculated the interquartile range (IQR) of final hand positions for each condition. Here we show the IQR ratio between conditions (y-axis). An IQR ratio greater than one (dashed grey line) indicates lower movement variability when given punishment feedback compared to reward feedback. Participants displayed significantly lower movement variability ($p = 0.034$) along the minor axis of the target with punishment feedback. **E)** Participants did not display differences in lag-1 autocorrelation between conditions ($p = 0.197$), suggesting that reward feedback and punishment feedback have a similar effect on sensorimotor exploration. Solid circles and connecting lines represent mean data for each condition. Hollow circles and connecting lines represent individual data. Box plots represent the 25th, 50th, and 75th percentiles. Taken together, these results suggest that punishment feedback suppresses movement variability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Taken together, the results in both experiments suggest that punishment feedback reduces movement variability that may be linked to a greater extent of learning.

Behavioural work by Galea and colleagues (2015) found that punishment feedback leads to faster adaptation while reward feedback leads to greater retention. Subsequent work by Song and colleagues (2020) replicated the finding that punishment feedback leads to faster adaptation, but did not find greater retention with reward feedback. Contrary to both these experiments, and despite using a larger sample size (n = 68), we did not replicate faster adaptation with punishment feedback. Aligned with the findings of Song and colleagues (2020), we also did not find greater retention with reward feedback when controlling for the extent of learning. There were some notable differences in punishment feedback, reward feedback, and error-based feedback between all three studies. Both Galea et al. (2015) and Song et al. (2020) used scalar punishment and reward feedback that changed in magnitude given the distance to the target (e.g., −5 points vs. −10

points), whereas our study used binary punishment feedback and binary reward feedback that was a constant magnitude. Additionally, our study used probabilistic binary feedback compared to the deterministic feedback used by Galea et al. (2015) and Song et al. (2020). In the present study, participants received probabilistic feedback of success and failure that would limit how quickly one would approach the optimal reach angle. Thus, the probabilistic feedback by nature could reduce the ability to see differences between groups during early learning.

Further, Galea and colleagues (2015) used error-based feedback (small cursor mapped to hand position), whereas Song and colleagues (2020) and our study did not use error-based feedback. Indeed, past work has shown evidence that participants can optimally account for movement variability to adjust motor planning in the presence of visual feedback (Trommershauser et al., 2005). However, this ability to optimally account for additional, externally provided movement variability is decreased when provided binary reinforcement feedback (Therrien et al., 2018). Scalar punishment and

reward can be used to form an error signal over just two trials. For example, if one receives −10 points for straight reach and then receives −5 points for reaching to the right, then it is possible to deduce a vectored error signal (direction and magnitude) and realize that reaching further right may result in no punishment (i.e., 0 points). Conversely, binary feedback (success or fail) does not make an error signal readily available. It is possible that scalar punishment feedback, with (Galea et al., 2015) or without (Song et al., 2020) online feedback of a hand position, is needed to elicit faster adaptation when compared to reward feedback. Additionally, the perturbation used by Song and colleagues (2020) was considerably larger than that of the present study. A larger perturbation in combination with scalar punishment feedback may have induced faster adaptation via explicit motor learning processes. Future studies should examine if there are fundamentally different mechanistic processes associated with binary and scaler reward or punishment feedback. Given current inconsistencies in the literature, future work examining the differential roles of reward and punishment on sensorimotor adaptation should be well-powered and meta-analyses would also likely be insightful.

Past work has suggested that punishment feedback may lead to greater movement variability to enhance learning. Interestingly, across several metrics of movement variability in **Experiments 1** and **2** we found evidence of punishment feedback leading to *less* movement variability. In **Experiment 1**, there was significantly less movement variability about detrended reach angles with punishment feedback compared with reward feedback for several bin sizes, with all but one bin size having a p-value below 0.1. In **Experiment 2**, participants experienced punishment feedback or reward feedback in a task that mitigated the influence of adaptive processes that would lead to changes in reach aim along a particular direction. Using a within experimental design, we found both significantly less trial-by-trial movement variability and less dispersion of final hand position (i.e., IQR ratio). Our finding is in opposition to the assumption that greater movement variability leads to enhanced learning (Song et al., 2020; Wu et al., 2014), and recent findings to suggest that punishment feedback leads to comparatively more movement variability than reward feedback (Song et al., 2020). However, it is important to note that movement variability is often decomposed into sensorimotor noise and exploratory movement variability. While some research has suggested that sensorimotor noise in some contexts may impede adaptation during error-based tasks (He et al., 2016), others have suggested that the sensorimotor system has knowledge of exploratory movement variability that can facilitate adaptation (Cashaback et al., 2019; Therrien et al., 2016; van Beers, 2009). Several different sources of movement variability have been proposed in the literature. Motor noise, planned noise, and exploratory noise have been proposed as sources of movement variability that respectively arise from stochastic neuromuscular processes (Faisal et al., 2008; Jones et al., 2002; van Beers et al., 2004), the dorsal premotor cortex (Churchland et al., 2006; Sutter et al., 2021; van Beers

et al., 2013), and the basal ganglia (Cashaback et al., 2019; Olveczky et al., 2005; Pekny et al., 2015). Of these potential noise sources, the sensorimotor system is proposed to have knowledge of planned noise (van Beers et al., 2013; van Beers, 2009; Van Der Vliet et al., 2018) and exploratory noise (Therrien et al., 2018; Cashaback et al., 2019; Roth et al., 2023). It is largely thought that the sensorimotor system has knowledge of movement variability arising from exploratory noise following an unsuccessful motor action, which is subsequently used to update the intended reach aim following a successful motor action (Therrien et al., 2018; Cashaback et al., 2019; Roth et al., 2023). Our findings suggest that punishment feedback leads to lower movement variability compared with reward feedback. However, it is possible that the sensorimotor system has more knowledge of exploratory movement variability following punishment feedback.

The idea that punishment decreases exploratory movement variability while increasing knowledge of movement variability can be explained using a class of reinforcement-based learning models (Cashaback et al., 2019;Roth et al., 2023). In this framework, there would be an increase in the term associated with movement updates (indicating greater knowledge of movement variability) as well as a decrease in terms associated with exploratory movement variability. A decrease in exploratory movement variability would lower a lag-1 autocorrelation while an increase in knowledge of exploratory movement variability would increase the lag-1 autocorrelation in **Experiment 2**, which could result in no net change of the lag-1 autocorrelation. However, a greater knowledge of exploratory movement variability following punishment, despite having a comparatively smaller magnitude of movement variability compared to reward, could explain our finding that punishment feedback led to a greater extent of learning in **Experiment 1.**

As explained by prospect theory, a common observation across species is that organisms tend to avoid decisions where the probability of success is low (Tversky and Kahneman, 1981; Niv et al., 2012; McDougle et al., 2016; Kahneman and Tversky, 1979; Nioche et al., 2019; Isett et al., 2023; Harder and Real, 1987; Dener et al., 2016). Such risk aversion has been shown to influence motor planning (Nagengast et al., 2011). In **Experiment 2**, and likely to some extent in **Experiment 1**, participants displayed reduced movement variability with punishment feedback compared to reward feedback. Initially, this appears to contradict risk aversion. One may expect participants to more drastically alter their reaching behaviour to avoid receiving punishment, a behaviour typically seen in decision-making tasks (Kahneman and Tversky, 1979; Tversky and Kahneman, 1981; Worthy et al., 2013). However, radically altering reaching behaviour without visual feedback could increase the likelihood of a negative outcome (monetary loss) on the next trial. The sensorimotor system may have limited drastic changes in reach behaviour to avoid increasing the probability of experiencing punishment feedback, aligning with risk aversion. Indeed, participants may have taken longer in movement preparation to minimize reach variability (Sutter et al., 2021), as evidenced

by the increased reaction times seen with punishment feedback compared to reward feedback across **Experiments 1** and **2**. In **Experiment 1**, we found some evidence of reduced movement variability between participants receiving punishment feedback and reward feedback, as well as a greater extent of learning with punishment feedback. One possible explanation is that with punishment, the sensorimotor system makes smaller, more directed changes in reach angle so as not to move too far from the displayed visual target. Conversely, with reward the sensorimotor system may be more willing to execute larger, more sporadic changes in reach angle when there is no negative outcome associated with missing the target. Additionally, it is possible that the punishment feedback was more salient to participants than the reward feedback, creating a stronger signal with punishment feedback. This stronger signal, particularly in late learning, could have also induced lower variability in the punishment group of **Experiment 1** and subsequently a greater extent of learning. Thus, risk aversion as a mechanism that avoids drastic changes in reach aim may provide an explanation for the observed decreases in movement variability with punishment feedback.

Larger changes in reach aim following an indicated target miss with reward feedback compared to punishment feedback highlights an important distinction. Specifically, the lack of a reward signal may not have the same effect on reach behaviour as the addition of a punishment signal for the same motor error. Indeed, prediction errors for both reward and punishment have been shown to arise in different brain regions (Gueguen et al., 2021). Here a prediction error is the difference between received and expected reward (or punishment). It is possible that risk aversion could directly cause a larger prediction error caused by punishment compared to reward, given the same motor error. A larger prediction error could have induced greater updates to reach aim in our tasks. Future work should explore the relative strength of prediction errors generated by reward and punishment feedback.

We found in **Experiment 1** that trial-by-trial movement variability and detrended movement variability was correlated with the extent of learning. As mentioned, one possibility for this finding is that the sensorimotor system had greater knowledge of movement variability following punishment. Another possibility that we observed a greater extent of learning is that adaptive processes become more sensitive with punishment. Galea and colleagues (2015) had previously found a faster learning rate with punishment feedback during an error-based motor task. They attributed a faster learning rate with the cerebellum becoming more sensitive to sensory prediction errors in the presence of punishment (Ernst et al., 2002). Here we did not use error-based feedback, so it is unlikely that the greater extent of learning we observed in **Experiment 1** was due to increased cerebellar sensitivity. However, it is possible that other reinforcement-based processes had greater sensitivity to punishment feedback that led to updates in reach aim. The basal ganglia is linked to both reward and punishment (Delgado et al., 2000) as well as being implicated

with successful and unsuccessful motor actions for songbirds (Olveczky et al., 2005), mice (Dhawale et al., 2019), and those with Parkinson's disease (Pekny et al., 2015). Further, the indirect pathway of the basal ganglia is associated with punishment signals that reduce motor output (Kravitz and Kreitzer, 2012), which may explain reduced movement variability with punishment feedback. It is possible that punishment may also increase the sensitivity of the basal ganglia to update changes in reach aim. Thus, greater learning extent in the first experiment may have been caused by greater knowledge of movement variability and or increased sensitivity to updating reach aim following punishment.

Reward and punishment have a rich history in psychology and have recently been investigated in sensorimotor neuroscience. Reward and punishment have been shown to enhance distinctly different features of sensorimotor adaptation, such as the rate of learning (Wächter et al., 2009; Wu et al., 2014; Galea et al., 2015; Song and Smiley-Oyen, 2017; Song et al., 2020) and retention of motor actions (Abe et al., 2011; Shmuelof et al., 2012; Galea et al., 2015; Song and Smiley-Oyen, 2017; Song et al., 2020; Vassiliadis et al., 2021). Here, we investigated how punishment feedback and reward feedback differentially influence motor adaptation and movement variability. Unexpectedly, and contrary to recent findings (Galea et al., 2015; Song et al., 2020), with punishment feedback we did not observe faster learning, but did find a greater extent of learning. Further, across two experiments we show evidence to suggest that punishment feedback decreases movement variability. We did find a relationship between movement variability and a greater extent of learning, which could be explained by the sensorimotor system having greater knowledge of movement variability and or increased sensitivity to updating reach aim following punishment. Understanding how reward and punishment feedback differentially influences adaptation may be beneficial for informing neurorehabilitation strategies for a variety of neurological disorders, such as Parkinson's disease (Pekny et al., 2015), cerebellar ataxia (Therrien et al., 2021), and stroke (Therrien et al., 2016; Therrien et al., 2018; Reinkensmeyer et al., 2012; Reinkensmeyer et al., 2016).

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abe M, Schambra H, Wassermann EM, Luckenbaugh D, Schweighofer N, Cohen LG (2011) Reward improves long-term retention of a motor memory through induction of offline memory gains. Curr Biol 21(7):557–562.

Acerbi L, Vijayakumar S, Wolpert DM (2014) On the origins of suboptimality in human probabilistic inference. PLOS Comput Biol 10(6):e1003661.

Bernstein N. The co-ordination and regulation of movement; 1967.

Buzzi J, De Momi E, Nisky I (2019) An uncontrolled manifold analysis of arm joint variability in virtual planar position and orientation telemanipulation. IEEE Trans Biomed Eng 66(2):391–402.

Calalo JA, Roth AM, Lokesh R, Sullivan SR, Wong JD, Semrau JA, Cashaback JG (2023) The sensorimotor system modulates muscular co-contraction relative to visuomotor feedback responses to regulate movement variability. J Neurophysiol 129(4):751–766.

Cashaback JGA, Cluff T (2015) Increase in joint stability at the expense of energy efficiency correlates with force variability during a fatiguing task. J Biomech 48(4):621–626.

Cashaback J, McGregor H, Mohatarem A, Gribble P (2017) Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. PLoS Comput Biol 13(7):1005623.

Cashaback J, Lao C, Palidis D, Coltman S, McGregor H, Gribble P (2019) The gradient of the reinforcement landscape influences sensorimotor learning. PLoS Comput Biol 15(3):1006839.

Chen X, Mohr K, Galea J (2017) Predicting explorative motor learning using decision-making and motor noise. PLoS Comput Biol 13(4):1005503.

Churchland M, Afshar A, Shenoy K (2006) A central source of movement variability. Neuron 52(6):1085–1096.

Cohen J (1988) Statistical power analysis for the behavioural sciences. 2nd edn. New York. NY: Lawrence Erlbaum Associates.[Google Scholar].

Cusumano JP, Cesari P (2006) Body-goal variability mapping in an aiming task. Biol Cybern 94(5):367–379.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 84(6):3072–3077.

Dener E, Kacelnik A, Shemesh H (2016) Pea plants show risk sensitivity. Curr Biol 26(13):1763–1767.

den Ouden HEM, Daw ND, Fernandez G, Elshout JA, Rijpkema M, Hoogman M, Franke B, Cools R (2013) Dissociable effects of dopamine and serotonin on reversal learning. Neuron 80(4):1090–1100.

Dhawale A, Smith M, Olveczky B (2017) The role of variability in motor learning. Annu Rev Neurosci 40:479–498.

Dhawale AK, Miyamoto YR, Smith MA, Ölveczky BP (2019) Adaptive regulation of motor variability. Curr Biol 29(21):3551–3562.e7.

Ernst M, Bolla K, Mouratidis M, Contoreggi C, Matochik JA, Kurian V, Cadet J-L, Kimes AS, London ED (2002) Decision-making in a risk-taking task: a PET Study. Neuropsychopharmacology 26(5):682–691.

Faisal A, Selen L, Wolpert D (2008) Noise in the nervous system. Nat Rev Neurosci 9(4):292–303.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306(5703):1940–1943.

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

Gribble PL, Scott SH (2002) Overlap of internal models in motor cortex for mechanical loads during reaching. Nature 417(6892):938–941.

Gueguen MC, Lopez-Persem A, Billeke P, Lachaux J-P, Rheims S, Kahane P, Minotti L, David O, Pessiglione M, Bastin J (2021) Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. Nat Commun 12(1):3344.

Hamel R, Pearson J, Sifi L, Patel D, Hinder MR, Jenkinson N, Galea J. The neurochemical mechanisms underlying the enhancing effects of rewards and punishments on motor performance. bioRxi; 2023, 2023–03.

Harder LD, Real LA (1987) Why are bumble bees risk averse? Ecology 68(4):1104–1108.

Hauke J, Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest Geogr 30(2):87–93.

He K, Liang Y, Abdollahi F, Bittmann MF, Kording K, Wei K (2016) The statistical determinants of the speed of motor learning. PLOS Comput Biol 12(9):e1005023.

Hester R, Murphy K, Brown FL, Skilleter AJ (2010) Punishing an error improves learning: the influence of punishment magnitude on error-related neural activity and subsequent learning. J Neurosci 30(46):15600–15607.

Hill CM, Stringer M, Waddell DE, Del Arco A (2020) Punishment feedback impairs memory and changes cortical feedback-related potentials during motor learning. Front Human Neurosci 14:294.

Hill CM, Waddell DE, Del Arco A (2021) Cortical preparatory activity during motor learning reflects visuomotor retention deficits after punishment feedback. Exp Brain Res 239(11):3243–3254.

Holland P, Codol O, Galea J (2018) Contribution of explicit processes to reinforcementbased motor learning. J Neurophysiol 119(6):2241–2255.

Hyndman R, Athanasopoulos G. Stationarity and differencing — Forecasting: Principles and Practice (2nd ed). OTexts: Melbourne, Australia. OTexts.com/fpp2; 2018.

Isett BR, Nguyen KP, Schwenk JC, Yurek JR, Snyder CN, Vounatsos MV, Adegbesan KA, Ziausyte U, Gittis AH (2023) The indirect pathway of the basal ganglia promotes transient punishment but not motor suppression. Neuron.

Jones K, Hamilton AC, Wolpert D (2002) Sources of signal-dependent noise during isometric force production. J Neurophysiol 88(3):1533–1544.

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–291.

Kaltenbach H-M (2012) A concise guide to statistics. SpringerBriefs in statistics. 2191–544X. Heidelberg: Springer.

Kravitz AV, Kreitzer AC (2012) Striatal mechanisms underlying movement, reinforcement, and punishment. Physiology 27(3):167–177.

Latash M, Scholz J, Schoner G (2002) Motor control strategies revealed in the structure of motor variability. Exercise Sport Sci Rev 30(1):26–31.

Lokesh R, Ranganathan R (2019) Differential control of task and null space variability in response to changes in task difficulty when learning a bimanual steering task. Exp Brain Res 237(4):1045–1055.

Lokesh R, Sullivan S, Calalo JA, Roth A, Swanik B, Carter MJ, Cashaback JG (2022) Humans utilize sensory evidence of others' intended action to make online decisions. Scient Rep 12(1):8806.

Lokesh R, Sullivan SR, St. Germain L, Roth AM, Calalo JA, Buggeln J, et al. Visual accuracy dominates over haptic speed for state estimation of a partner during collaborative sensorimotor interactions. J Neurophysiol; 2023.

Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27(27):7105–7116.

McDougle SD, Boggess MJ, Crossley MJ, Parvin D, Ivry RB, Taylor JA (2016) Credit assignment in movement-dependent reinforcement learning. Proc Nat Acad Sci 113(24):6797–6802.

McGraw KO, Wong SP (1992) A common language effect size statistic. Psychol Bull 111(2):361.

Nagengast AJ, Braun DA, Wolpert DM (2011) Risk sensitivity in a motor task with speed-accuracy trade-off. J Neurophysiol 105(6):2668–2674.

Nioche A, Bourgeois-Gironde S, Boraud T (2019) An asymmetry of treatment between lotteries involving gains and losses in rhesus monkeys. Scient Rep 9(1):10441.

Niv Y, Edlund JA, Dayan P, O'Doherty JP (2012) Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J Neurosci 32(2):551–562.

Olveczky B, Andalman A, Fee M (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol 3 (5):153.

Panigrahi B, Martin KA, Li Y, Graves AR, Vollmer A, Olson L, Mensh BD, Karpova AY, Dudman JT (2015) Dopamine is required for the neural representation and control of movement vigor. Cell 162 (6):1418–1430.

Pekny S, Izawa J, Shadmehr R (2015) Reward-dependent modulation of movement variability. J Neurosci 35(9):4015–4024.

Reinkensmeyer DJ, Guigon E, Maier MA (2012) A computational model of use-dependent motor recovery following a stroke: optimizing corticospinal activations via reinforcement learning can explain residual capacity and other strength recovery dynamics. Neural Networks: Off J Int Neural Network Soc 29-30:60–69.

Reinkensmeyer DJ, Burdet E, Casadio M, Krakauer JW, Kwakkel G, Lang CE, Swinnen SP, Ward NS, Schweighofer N (2016) Computational neurorehabilitation: modeling plasticity and learning to predict recovery. J Neuro-Eng Rehabil 13(1):42.

Robinson OJ, Frank MJ, Sahakian BJ, Cools R (2010) Dissociable responses to punishment in distinct striatal regions during reversal learning. NeuroImage 51(4):1459–1467.

Roth AM, Calalo JA, Lokesh R, Sullivan SR, Grill S, Jeka JJ, et al. Reinforcement-based processes actively regulate motor exploration along redundant solution manifolds. bioRxiv; 2023, 2023–02.

Scholz J, Schoner G (1999) The uncontrolled manifold concept: identifying control variables for a functional task. Exp Brain Res 126(3):289–306.

Shadmehr R, Reppert TR, Summerside EM, Yoon T, Ahmed AA (2019) Movement vigor as a reflection of subjective economic utility. Trends Neurosci 42(5):323–336.

Shmuelof L, Huang VS, Haith AM, Delnicki RJ, Mazzoni P, Krakauer JW (2012) Overcoming motor forgetting through reinforcement of learned actions. J Neurosci 32(42):14617–14621a.

Sidarta A, van Vugt FT, Ostry DJ (2018) Somatosensory working memory in human reinforcement-based motor learning. J Neurophysiol 120(6):3275–3286.

Sidarta A, Komar J, Ostry DJ (2022) Clustering analysis of movement kinematics in reinforcement learning. J Neurophysiol 127 (2):341–353.

Song Y, Smiley-Oyen AL (2017) Probability differently modulating the effects of reward and punishment on visuomotor adaptation. Exp Brain Res 235(12):3605–3618.

Song Y, Lu S, Smiley-Oyen AL (2020) Differential motor learning via reward and punishment. Quart J Exp Psychol 73(2):249–259.

Sukumar S, Shadmehr R, Ahmed A. Effects of reward history on decision-making and movement vigor; 2021.

Summerside E, Shadmehr R, Ahmed A (2018) Vigor of reaching movements: reward discounts the cost of effort. Journal of neurophysiology 119(6):2347–2357.

Sutter K, Oostwoud Wijdenes L, van Beers RJ, Medendorp WP (2021) Movement preparation time determines movement variability. J Neurophysiol 125(6):2375–2383.

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

Therrien A, Wolpert D, Bastian A. Increasing motor noise impairs reinforcement learning in healthy individuals; 2018.

Therrien AS, Statton MA, Bastian AJ (2021) Reinforcement signaling can be used to reduce elements of cerebellar reaching ataxia. Cerebellum (London, England) 20(1):62–73.

Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211(4481):453–458.

van Beers R (2009) Motor learning is optimally tuned to the properties of motor noise. Neuron 63(3):406–417.

van Beers R, Haggard P, Wolpert D (2004) The role of execution noise in movement variability. J Neurophysiol 91(2):1050–1063.

van Beers R, Brenner E, Smeets J (2013) Random walk of motor planning in task-irrelevant dimensions. J Neurophysiol 109 (4):969–977.

van der Kooij K, Smeets JBJ (2019) Reward-based motor adaptation can generalize across actions. J Exp Psychol: Learn, Memory, Cognition 45(1):71.

van der Kooij K, van Mastrigt NM, Cashaback JGA (2023) Failure induces task-irrelevant exploration during a stencil task. Exp Brain Res 241(2):677–686.

Van Der Vliet R, Frens MA, De Vreede L, Jonker ZD, Ribbers GM, Selles RW, et al. Individual differences in motor noise and adaptation rate are optimally related. eneuro, 2018;5 (4).

Van Mastrigt N, Smeets J, Van Der Kooij K (2020) Quantifying exploration in reward-based motor learning. Plos One 15 (4):0226789.

van Mastrigt NM, van der Kooij K, Smeets JB (2021) Pitfalls in quantifying exploration in reward-based motor learning and how to avoid them. Biol Cybern 115(4):365–382.

Vassiliadis P, Derosiere G, Dubuc C, Lete A, Crevecoeur F, Hummel FC, Duque J (2021). Reward boosts reinforcement-based motor learning. iScience 24(7):102821.

Wächter T, Lungu OV, Liu T, Willingham DT, Ashe J (2009) Differential effect of reward and punishment on procedural learning. J Neurosci 29(2):436–443.

Worthy DA, Hawthorne MJ, Otto AR (2013) Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. Psychon Bull Rev 20 (2):364–371.

Wu H, Miyamoto Y, Castro L, Olveczky B, Smith M (2014) Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nat Neurosci 17(2):312–321.

APPENDIX A. SUPPLEMENTARY MATERIAL