UC Berkeley
UC Berkeley Previously Published Works

Title

Implicit reward-based motor learning

Permalink

https://escholarship.org/uc/item/9rm3s9pk

Journal

Experimental Brain Research, 241(9)

ISSN

0014-4819

Authors

van Mastrigt, Nina M

Tsay, Jonathan S

Wang, Tianhe

et al.

Publication Date

2023-09-01

DOI

10.1007/s00221-023-06683-w

Copyright Information

Peer reviewed

RESEARCH ARTICLE

Implicit reward-based motor learning

Nina M. van Mastrigt[1] · Jonathan S. Tsay[2] · Tianhe Wang[2] · Guy Avraham[2] · Sabrina J. Abram[2] ·
Katinka van der Kooij[1] · Jeroen B. J. Smeets[1] · Richard B. Ivry[2]

Abstract
Binary feedback, providing information solely about task success or failure, can be sufficient to drive motor learning. While binary feedback can induce explicit adjustments in movement strategy, it remains unclear if this type of feedback also induces implicit learning. We examined this question in a center-out reaching task by gradually moving an invisible reward zone away from a visual target to a final rotation of 7.5° or 25° in a between-group design. Participants received binary feedback, indicating if the movement intersected the reward zone. By the end of the training, both groups modified their reach angle by about 95% of the rotation. We quantified implicit learning by measuring performance in a subsequent no-feedback aftereffect phase, in which participants were told to forgo any adopted movement strategies and reach directly to the visual target. The results showed a small, but robust (2–3°) aftereffect in both groups, highlighting that binary feedback elicits implicit learning. Notably, for both groups, reaches to two flanking generalization targets were biased in the same direction as the aftereffect. This pattern is at odds with the hypothesis that implicit learning is a form of use-dependent learning. Rather, the results suggest that binary feedback can be sufficient to recalibrate a sensorimotor map.

Keywords Reward-based motor learning · Reinforcement learning · Implicit learning · Visuomotor rotation · Use-dependent learning · Reward

Introduction

The execution of accurate movements relies on sensory feedback. Variants of sensorimotor adaptation experiments have been used to study the role of different forms of feedback on motor learning. In a typical visuomotor adaptation experiment, participants perform target-directed center-out reaching movements with feedback of the unseen hand limited to a visual cursor. To study learning, the position of the cursor is altered, resulting in a sensory prediction error, defined by the difference between the predicted and actual cursor position (Izawa and Shadmehr 2011; Kim et al. 2018; Morehead et al. 2017; Shadmehr et al. 2010; Synofzik et al. 2008; Tseng et al. 2007). This directional error can drive

different forms of learning. It can produce recalibration of a so-called sensorimotor map, such that a subsequent movement to that target will be shifted in the direction opposite to the perturbed feedback, a process known as sensorimotor adaptation (Kim et al. 2021; Krakauer 2009; Krakauer et al. 2019). It can also elicit explicit strategies to reduce the error; for example, the participant might aim away from the target (Bond and Taylor 2015; Taylor et al. 2014).

Feedback can also be limited to binary information conveying success or failure. In reaching tasks, success can be defined by the hand intersecting an invisible reward zone. To elicit learning, the reward zone is displaced from the target. This might be done in an abrupt manner. For example, success suddenly requires reaches into a reward zone that is centered 30° from the target. Alternatively, the reward zone can be shifted in a gradual manner, for example in 5° increments eventually reaching a maximum displacement of 30°. Following the introduction of the perturbation, success requires a movement that is off-target. While participants can find it challenging to learn when the shift is large or introduced abruptly (Brudner et al. 2016; Holland et al. 2018), many studies have shown that binary feedback is sufficient to

Communicated by Winston D Byblow.

✉ Nina M. van Mastrigt
  n.m.van.mastrigt@vu.nl

1 Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

2 CognAc Lab, UC Berkeley, Berkeley, CA, USA

produce learning when the shift is introduced in a gradual manner (Cashaback et al. 2019; Izawa and Shadmehr 2011; Therrien et al. 2016, 2018; van der Kooij et al. 2021; van der Kooij and Smeets 2018).

While sensory prediction errors and binary reward feedback can produce similar adjustments in behavior, there are marked differences in the representational changes associated with these two forms of learning (Morehead and Orban de Xivry 2021; Therrien and Wong 2022). For example, adaptation from sensory prediction errors is greatly attenuated when any delay is introduced between the movement and feedback, whereas adaptation from binary reward feedback is minimally impacted by delays up to a few seconds (Brudner et al. 2016; Schween and Hegele 2017). In addition, the acquired behavior is more persistent following reward-based feedback compared to error-based feedback (Bao and Lei 2022; Galea et al. 2015; Shmuelof et al. 2012; Therrien et al. 2016).

Learning processes can also be evaluated in terms of the degree to which they result in implicit and explicit changes in behavior. A large body of literature has shown that adaptation from sensory prediction errors occurs in an automatic and implicit manner (Kim et al. 2018; Mazzoni and Krakauer 2006; Morehead et al. 2017). Adaptation can also result from re-aiming, which is explicit and under volitional control. To date, less is known about implicit changes in behavior in response to binary feedback. Following the convention in the adaptation literature, a strong probe of implicit learning is to focus on behavioral changes that persist when the feedback is eliminated and participants are reminded to reach directly to the target (Maresch et al. 2021a, b). When probed in this manner following reward feedback, a small aftereffect is observed. For example, following a shift of the reward zone of 25°, the average heading angle at the start of the aftereffect phase was around 5° (Holland et al. 2018). This suggests that reward-based learning is largely the result of a volitional change in strategy. Consistent with this hypothesis, disrupting explicit processes by introducing a secondary task attenuates learning from binary feedback (Codol et al. 2018; Holland et al. 2018). Nonetheless, the fact that there is an aftereffect, even if small, indicates binary feedback can induce implicit learning (Codol et al. 2018; Holland et al. 2018, 2019).

What might be the source of this implicit component? We can consider two, non-mutually exclusive hypotheses. The first hypothesis centers on the idea that the behavioral change resulting from binary feedback includes a contribution from implicit, use-dependent learning. As implied by the name, use-dependent learning refers to a movement bias toward frequently repeated movements (Diedrichsen et al. 2010; Huang et al. 2011; Marinovic et al. 2017; Mawase et al. 2017; Tsay et al. 2022; Verstynen and Sabes 2011). Tracking the reward zone will result in movements that are

shifted in a consistent direction relative to the visual target. In an aftereffect phase, a use-dependent bias would produce a residual implicit bias in this direction. Interestingly, the 3–4° aftereffect following training with binary feedback is similar in magnitude to that observed in the studies of use-dependent learning that exclude errors in action selection (Tsay et al. 2022).

A second hypothesis is that binary feedback induces implicit recalibration of a sensorimotor map. Mechanistically, implicit recalibration could occur, because the binary feedback alters the contingency between action plans and their associated movements. Feedback that indicates task success would strengthen the association between the goal to reach to a visual target and movements linked to that target, even if these are toward a reward zone that is displaced relative to the visual target. Feedback that indicates task failure would weaken this association. Compared to error-based learning, recalibration from reward feedback would appear to be much more limited given that the aftereffect following binary feedback is much smaller than that following cursor feedback for similar perturbation sizes (Bond and Taylor 2015; Codol et al. 2018; Holland et al. 2018; Leow et al. 2018; Taylor and Ivry 2014).

Here, we report the results of an experiment designed to assess these use-dependent learning and implicit recalibration hypotheses. Providing binary feedback only, we examined how participants learned to respond to either a small (7.5°) or large perturbation (25°) of the reward zone. For both groups, the perturbation was introduced in a gradual manner. Assuming that participants in the small perturbation condition will have little awareness of the perturbation, this condition provides a strong test of the role of implicit processes in reward-based learning. In contrast, we assumed that participants in the large condition would eventually adopt a strategy.

To assess implicit learning in both conditions, we measured reaching in an aftereffect phase in which all feedback was eliminated and participants were instructed to reach directly to the target. The implicit recalibration and use-dependent hypothesis both predict aftereffects in the Small and Large conditions. To compare the two hypotheses, we included two probe targets in the aftereffect phase, displaced by 15° from the training target location (Fig. 1c). The inclusion of the probe targets allowed us to ask how implicit learning, if observed, generalized. By the implicit recalibration hypothesis, we would expect that reaches to the probe targets would be biased to a similar extent and in the same direction as reaches to the trained target. By the use-dependent hypothesis, we should observe that reaches to the probe targets would be attracted toward the trained movement. For the Small perturbation condition, the biases to the two probe targets should be in the opposite direction, since the
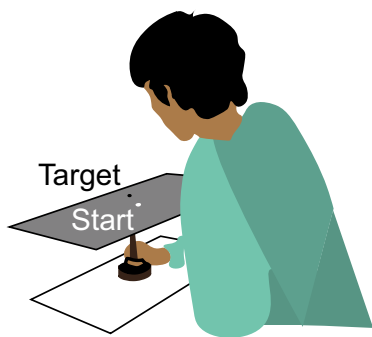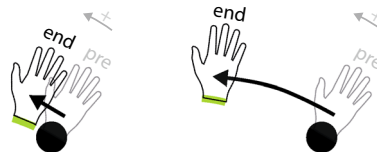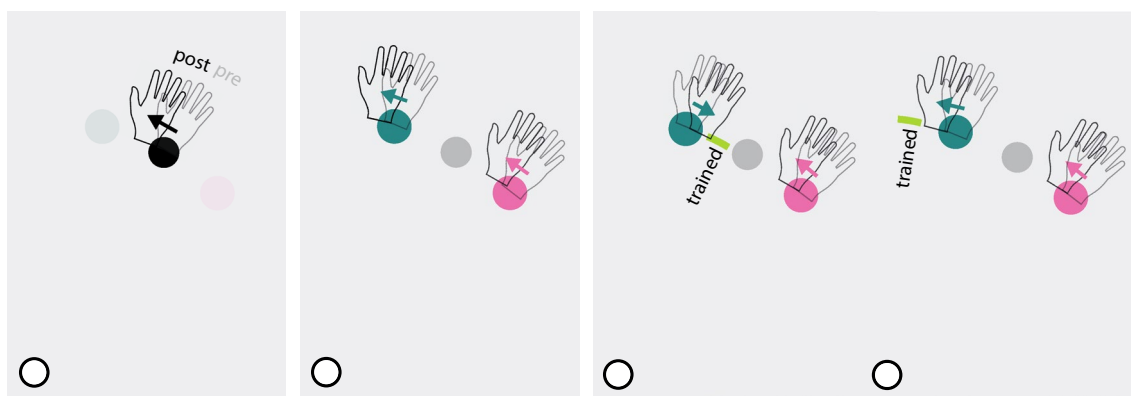
Fig. 1 Schematic outline of key hypotheses on implicit reward-based motor learning. **a** Schematic of a participant in the experimental apparatus. **b** Training phase. Participants made center-out reaching movements from a white starting circle to a black training target. A pleasant auditory "ding" was provided when the movement passed within the reward zone (green arch); otherwise, an unpleasant "buzz" was played. **c** No-feedback phase. Participants were instructed to reach directly to a visual target. The target appeared at the training location or one of two probe locations (± 5°). Participants were instructed to forgo any strategy adopted during the training phase. Left panel shows implicit learning as measured by an aftereffect, defined as a change in hand angle for reaches to the training target from pre-training (translucent hand) to post-training (solid hand). Middle panel shows probe target reaching predictions for the implicit recalibration hypothesis. Reaches will be biased in the same direction for both probe targets independent of size of the perturbation. Right panel shows probe target reaching predictions for the use-dependent learning hypothesis. For the Small perturbation condition, the biases will be in opposite directions, since the reaches during training fall between the two probe locations. For the Large perturbation condition, the direction of the bias for the probe target nearest the reward zone will depend on the degree of learning. In the example depicted here, the participant showed full learning (beyond both probe targets, > 15°), and thus, the direction of the biases for the use-dependent prediction is the same for both probe targets. For participants who show reduced learning (in between both probe targets, < 15°), the predictions for use-dependent learning are as shown for the Small perturbation

trained movement falls between the two probe locations. The predictions are less clear for the large perturbation condition and will depend on the magnitude of learning. Biases to the two probe targets will be in the same direction if participants fully track the 25° shift of the reward zone. However, if the trained movement falls short of the reward zone, the biases will become less symmetric and even have opposite signs once the trained movement is less than 15°.

Methods

Participants

Sixty-eight right-handed young adults were recruited from the research participant pool of the Department of Psychology at the University of California, Berkeley. Twenty eight (22 females, 6 males; reported age: mean 20.5, SD 2.3 years) were assigned to the "Small" perturbation group and 40 (27 females, 13 males; reported age: mean 21.5, SD 5.7 years) were assigned to the "Large" perturbation group. Participants received either course credit or financial compensation for their participation, along with a $5 completion bonus paid to all participants. Based on self-reports, participants had normal or corrected-to-normal vision and hearing. The protocol was approved by the institutional review board at UC Berkeley.

Of the original 68 participants, 20 were excluded from all analyses. 16 of these (8 per group) were excluded based on their responses to a post-experiment questionnaire (see "Experimental design") that indicated they failed to follow the instructions. Four other participants in the large group were excluded for idiosyncratic reasons: one fell asleep during the task, one reported, after the experiment having performed in a similar experiment, one did not use the apparatus correctly, and one experienced an equipment failure. Thus, the analyses reported below are based on data obtained from 20 participants (16 females; 10 for credit; reported age: mean 20.9, SD 2.4 years) in the Small perturbation group and of 28 participants (16 females; 16 for credit; reported age: mean 21.8, SD 6.0 years) in the Large perturbation group.

Experimental setup

The participant sat in front of a table in a small, darkened room. A horizontally oriented computer screen (24″, ASUS, Taipei, Taiwan) constituted the upper surface of the table, with a 17″ digitizing tablet (Wacom Co., Kazo, Japan) positioned 27 cm below the screen (Fig. 1a). Stimuli were presented on the computer (refresh rate = 60 Hz) and the participant's movements along the digitizing tablet were recorded from a digitizing pen (sampling rate = 200 Hz) that was embedded in a custom-made paddle, ensuring the pen maintained a vertical position. Vision of the hand was obscured by the screen. A computer (Dell OptiPlex 7040, Round Rock, Texas) with a Windows 7 operating system (Microsoft Co., Redmond, Washington) was used to run the custom experimental software in Matlab (The MathWorks, Natick, Massachusetts), using Psychtoolbox extensions (Brainard 1997; Kleiner et al. 2007).

Trial structure

Each trial started with the appearance of a white "start" circle (radius = 0.42 cm), presented near the center of the screen. The participant was required to move the paddle to position the digitizing pen within the start circle. To guide the participant to the start location, a white ring was presented, with the radius of the ring indicating the distance from the pen to the start position. Movement toward the start position reduced the size of the ring. When the pen was within 0.84 cm of the start circle, the ring was replaced by a white circle (radius = 0.17 cm) that indicated the position of the pen, allowing the participant to move the pen into the start circle.

When the paddle had been in the start circle for 300 ms, a visual target (circle with radius = 0.28 cm) appeared 7 cm from the start circle at either 45°, 60°, or 75° (Fig. 1b, c). The participant was instructed to move in rapid manner, attempting to slice through the target. Auditory feedback was presented when the movement amplitude exceeded 7 cm. On trials with performance feedback (see below), a pleasant "bing" indicated that the movement was successful (e.g., passed through the target when feedback was veridical) and an aversive "buzz" indicated that the movement was unsuccessful. On no-feedback trials (in the baseline and aftereffect phases), a "knock" sound was played. This indicated that the required reach amplitude had been exceeded, but it did not provide feedback on whether the movement was within the reward zone or not.

To make participants move at similar, and relatively rapid speeds, an auditory message "Too slow" was played 800 ms after the performance feedback if the movement time was longer than 600 ms. This was the case on 3% of the trials. Note that these trials were included in the analyses given that the participants were provided reward feedback on these trials and thus would be expected to contribute to learning.

The feedback ring appeared directly after the feedback was given. Note that using a ring during the return movement, the participant received feedback indicating only the radial position of the hand. Angular position was only provided when the hand was very close to the start position: then, the ring turned into a cursor. This method was used, so that any effect of adaptation to the rotated feedback (see below) would be minimally visible to the participant during the return movement.

Experimental design

The experimenter instructed the participant that the purpose of the experiment was to study how well people can control arm movements in the absence of visual feedback. The participant was told that they would control an invisible cursor, and they were asked to make reaching movements

that would make the invisible cursor intersect a visual target (Fig. 1a). The experimenter described how a "bing" and "buzz" would indicate if the reach had intersected or missed the target, respectively. The experimenter then completed ten demonstration trials to demonstrate how the hand controlled the cursor movement. The target was always presented at the 60° location, and during these trials, the auditory feedback was accompanied by veridical cursor feedback.

After the ten demonstration trials, the participant was told that the cursor would no longer be visible during the reach, but that auditory feedback would be presented on most trials to indicate task outcome. However, on some trials, the participant would hear a "knock" sound, and this sound was uninformative concerning task outcome. To motivate the participant for all trials, the participant was informed that the computer kept track of all successful reaches and that a score in the top-third of high scores across participants would result in a $5 bonus (which was actually paid to all participants).

The main experiment consisted of three phases: baseline, training, and aftereffect, with the experimenter providing instructions at the beginning of each phase. The baseline phase was composed of 150 trials with feedback limited to the uninformative "knock" sound. The target appeared at each of the three possible locations on 50 of the baseline trials, with the order determined randomly. These trials allowed the participant to become familiar with the apparatus, learn to move at the appropriate speed, and provided a measure of natural biases for each of the three target locations (Kuling et al. 2019; van der Kooij et al. 2013).

The training phase was composed of 700 trials, with the target always appearing at the middle location (60°) and auditory feedback provided to indicate target hits or misses. For the first 100 trials, the reward zone was centered around the participant's individual bias while reaching to the trained target and extended 2° in both directions; if, for example, the individual's mean reach to the central target was rotated by 3° in the clockwise direction (at 57°), the initial reward zone spanned from 55° to 59°. Unbeknownst to the participant, the reward zone was gradually shifted over the next 500 trials. This was achieved by rotating the reward zone by 1.5° every 100 trials for the Small perturbation group and by 2.5° every 50 trials for the Large perturbation group. The rotation was either clockwise or counterclockwise, counterbalanced between participants. For the last 100 trials of the training phase, the reward zone remained fixed, displaced by 7.5° or 25° from its starting position for the Small and Large perturbation groups, respectively. A 2-min break was provided halfway through the 700-trial training phase.

Note that we expected that the participants in the Small group would likely remain unaware of the perturbation, since the shift was introduced gradually and the total displacement fell within 1–2 standard deviations of normal reach variability (Gaffin-Cahn et al. 2019). In contrast, we expected that participants in the Large group would likely become aware of the perturbation at some point during the training phase as the discrepancy between the visual target and hand movement would likely fall outside the individuals' normal reach variability.

Following the training phase, the participant completed an aftereffect phase of 150 trials. Prior to the start of the phase, the participant was instructed that the feedback might have been altered over the course of the training phase. To equally inform and instruct participants with different levels of awareness of the perturbation, the participant was informed that there were two groups of participants, an aligned group and a misaligned group. For the aligned group, the invisible cursor had always moved exactly with the position of the hand; for the misaligned group, the invisible cursor was slightly displaced from the position of the hand. To ensure that the participant understood the difference, they were asked to explain the difference between the two groups in their own words. If the explanation failed to capture the difference, the experimenter repeated the explanation. The experimenter then stated that for the final phase of the experiment, the cursor would be aligned with the hand for everyone, irrespective of initial group assignment, and thus, they should reach straight to the target to make the cursor hit the target. As in the baseline phase, reaches during this phase were performed with only the uninformative feedback, with the phase composed of 50 reaches to each of the three targets. Participants were again instructed that accuracy would be recorded during this phase to determine a final performance bonus.

At the end of the experiment, the participant completed a questionnaire consisting of five questions (Online Resource 1). Question 1 asked if they believed the feedback had been veridical or perturbed and Question 2 asked for their confidence concerning their response to Question 1, using a 7-point rating scale (1 = not confident, 7 = very confident). For Questions 3 and 4, the participants were asked to report (forced choice) where they had aimed during the training and aftereffect phases, respectively. Note that Question 4 was used to determine if the participant had followed the instructions. Those who answered that they had aimed to the left or right of the target during the aftereffect phase were excluded from all of the analyses (n = 16). For Question 5, the participant was informed that they had been in the Misaligned feedback group and were asked to indicate (forced choice) if the feedback had been perturbed: to the left or to the right. As the answers to this question were below chance level in the Small perturbation group, for the Large perturbation group, the illustrations for the two choices were slightly changed to match the hand movements of the participants better.

The total duration of the experiment was approximately 1 h.

Data analysis

Based on the data reported in Holland et al. (2018), a sample size of 21 would be required to detect implicit learning in our task with power of 0.80. We had recruited 40 and 28 participants for the Large and Small perturbation group, respectively, to put us safely above this number. However, as noted above, the final sample sizes were 28 and 20 for the Large and Small groups due to various exclusionary criteria.

Reach angle was determined by the line from the start position to where the digitizing pen crossed the 7 cm radius around the start position. The mean reach angle during the baseline trials was used to characterize individual biases for each of the three target locations separately (50 reaches/target). All analyses were based on the reach angles during the training and aftereffect phases, with these angles expressed relative to that participant's baseline bias for the corresponding target. Positive values correspond to reach angles shifted in the direction of the rotated reward zone.

We calculated the final learning as the mean reach angle in the last 100 trials of the training phase. To test for implicit learning, we calculated the mean reach angle to the training target in the aftereffect phase. For generalization, we calculated the mean reach angle for each of the two probe targets in the aftereffect phase.

Statistics

A preliminary analysis indicated that the final learning and aftereffect scores were not normally distributed (see Fig. 2). Therefore, we employed non-parametric tests in the statistical evaluation of the results. To test whether the final learning and aftereffect were larger than zero, we performed a one-tailed Wilcoxon signed-rank test on these variables for each group (Small and Large). To test whether implicit learning was different for the two perturbation sizes, we performed a two-tailed Wilcoxon rank-sum test on the aftereffect scores for the two groups. Since each group's implicit learning values were used in two statistical tests, we corrected for multiple comparisons, using a significance criterion of 0.025.

For the generalization data, we defined the percentage generalization as the mean of the two probe target biases, divided by the aftereffect at the training target. We used a one-tailed Wilcoxon signed-rank test to test whether the percentage generalization values were significantly larger than zero. To evaluate the form of generalization, we defined generalization asymmetry as the difference between the reaching bias to the probe target opposite the reward zone and the probe target in the direction of the reward zone. The

use-dependent learning hypothesis predicts that this value will be positive for the Small perturbation condition. The implicit recalibration hypothesis predicts that this value will be zero (if generalization is exactly the same for both targets, but see (Nikooyan and Ahmed 2015)). To evaluate the two hypotheses, we used a Wilcoxon signed-rank test to test whether the generalization asymmetry values were significantly greater than zero.

No statistics were performed on the questionnaire data.

Results

Learning

To evaluate how people modified their movements given the gradual change in the reward zone, we analyzed the reach angle at the end of learning in both the Small (maximum shift of 7.5°) and Large (maximum shift of 25°) perturbation groups. Both groups learned to compensate for the feedback perturbation (Fig. 2a, b). Participants in the Small perturbation group showed a median final learning of 7.1° (IQR [5.8°, 7.8°], $p < 0.001$, $z = 3.9$, Ws = 210, $r = 0.20$) (Fig. 2c, horizontal axis). Participants in the Large perturbation group showed a median final learning of 23.7° (IQR [9.6°, 25.2°], $p < 0.001$, $z = 4.2$, Ws = 390, $r = 0.16$ (Fig. 2d, horizontal axis). For both groups, this corresponds to a mean perceptual change of 95% of the perturbation size (Small: IQR = 77%–104%; Large: IQR = 38%–101%).

As can be seen in Fig. 2c, d (horizontal axes), learning was more variable in the Large perturbation group than in the Small perturbation group. For the latter, all of the participants changed their reaches in the direction of the perturbation and 86% ended up with a mean heading angle over the final 100 trials that was within the final reward zone. In contrast, only 70% of the participants in the Large perturbation group reached the final reward zone (Online Resource 2). Four participants in this group exhibited a mean hand angle over the final 100 trials that was in the opposite direction of the reward zone.

Aftereffect

The central aim of our study was to examine whether binary feedback regarding success or failure induces implicit motor learning. To this end, we focused on the reach direction during the aftereffect phase when the feedback was removed and participants were instructed to reach directly to the target.

Both groups showed a significant aftereffect (Fig. 2c, d vertical axes). Participants in the Small perturbation group had a median aftereffect of 3.4° (IQR [2.2°, 7.8°]; $p < 0.001$, $z = 3.90$, Ws = 210, $r = 0.20$). Participants in the Large perturbation group had a median aftereffect of 2.2°
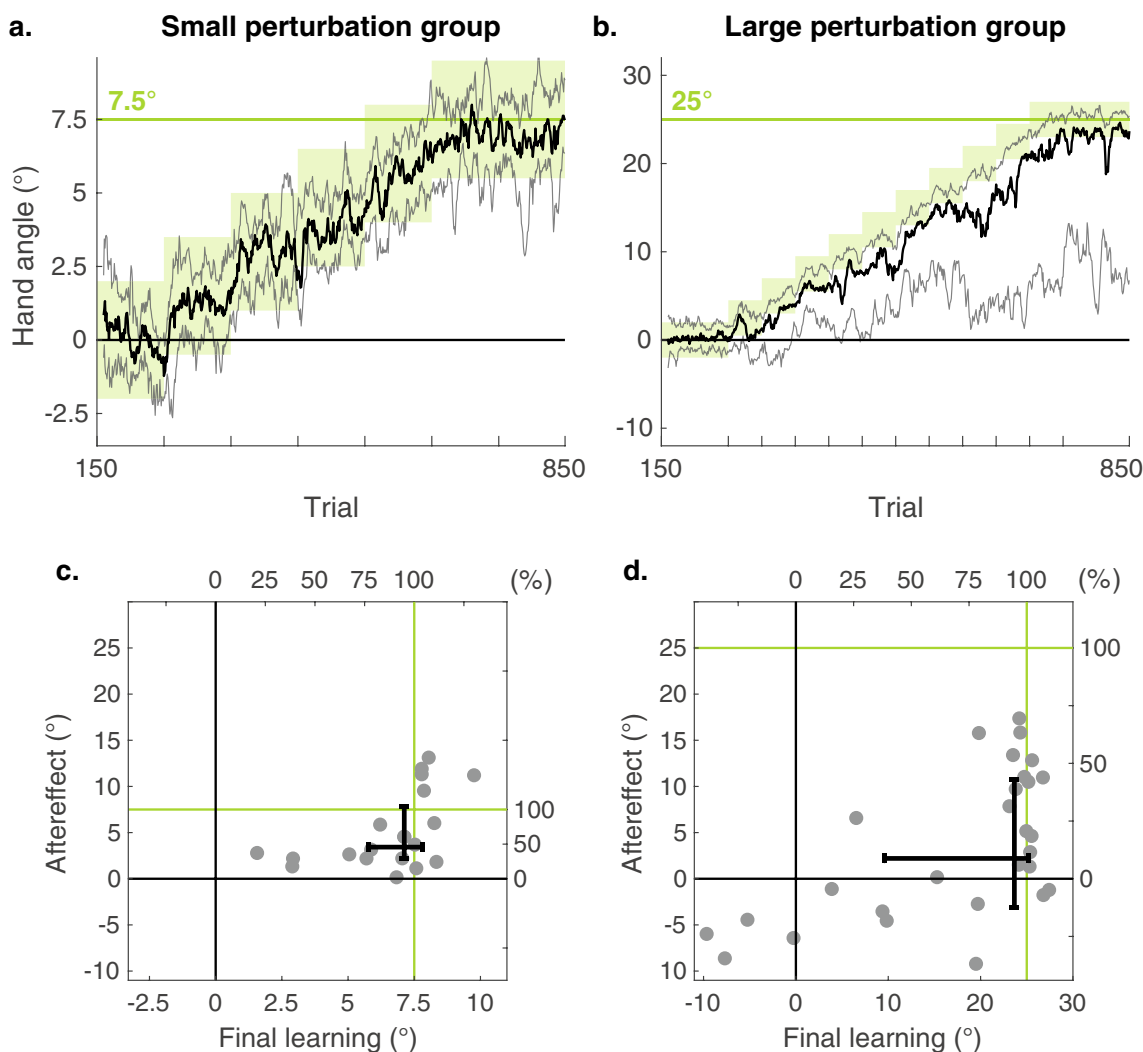
Fig. 2 The effect of binary reward feedback on reaching. **a**, **b** The training phase. Gradually changing the rewarded hand angles (green zone) leads to learning, as indicated by the change in reach angle. We plot the median (solid thick lines) over all participants with the interquartile range (opaque lines) for the Small perturbation group (**a**) and Large perturbation group (**b**). Note that the vertical axes are scaled to the perturbation size. For display purposes, the curves are smoothed with a running average with a window size of ten trials. **c**, **d** Aftereffect as a function of the final learning for both groups. Each gray dot corresponds to a participant and error bars indicate the interquartile range per group. The green lines indicate the perturbation size

(IQR [− 3.1°, 10.7°], $p = 0.02$, $z = 2.02$, Ws = 292, $r = 0.07$). Importantly, we found no difference between the magnitude of the aftereffect for the Small and Large perturbation groups ($p = 0.24$, $z = -1.2$, U = 434).

As can be seen in the figure, the four participants in the Large group who had negative final learning scores also showed negative aftereffects. When these participants are excluded, the median aftereffect for this group increases to 4.90°. As with the original analysis, there remains no difference in the magnitude of the aftereffect for the Small and Large groups in this secondary analysis ($p = 0.89$, $z = -0.13$, U = 456).

In summary, the aftereffect data indicate that there is an implicit component to learning that occurs in response to binary feedback. The magnitude of the aftereffect in both the Small and Large perturbation groups was of similar size and quite small.

Generalization

We included reaches to two probe targets in the aftereffect phase, asking how learning generalized to regions of the workspace neighboring the training target. Both groups exhibited generalization in that the reaches to the probe locations were significantly shifted from the baseline phase. In terms of the direction of the shift, the mean values were all positive, meaning that the change in reach direction for the probe targets was in the same direction as the change in

reach direction to the training target (Fig. 3a). Participants in the Small perturbation group had a median reaching bias of 3.5° to the probe target in the direction of the learning and of 3.6° to the other probe target. The corresponding biases were 1.6° and 0.7° for the Large perturbation group. The latter values increase to 2.3° and 4.3° if the four negative final learners are excluded.

The generalization data are not consistent with the use-dependent learning hypothesis. The use-dependent learning hypothesis had predicted biases in opposite directions for the two probes in the Small perturbation group, since the trained movement was between the two probe targets. This would predict positive generalization asymmetry scores. In the Large perturbation group, the predictions are less clear, since they depend on the location of the trained movement relative to the probe targets. For participants for whom the final trained movement fell between the probe targets (i.e., < 15°), the use-dependent hypothesis would predict positive generalization asymmetry scores, similar to the prediction for the Small group. However, for participants who fully followed the reward zone, the trained movement was beyond both the probe targets. As such, the use-dependent hypothesis would predict biases for the

two probes in the same direction, although the magnitudes would differ (see Fig. 1c). For both groups, the analyses showed that the asymmetry scores were not significantly larger than zero (Fig. 3b; Small: median = − 1.0°, IQR [− 3.0°, 3.5°], $p = 0.55$, $z = -0.06$, Ws = 89; Large: median = 0.0°, IQR [− 2.7°, 2.0°], IQR [53.3%, 100.5%], $p = 0.96$, $z = -0.05$, Ws = 201). Moreover, we observed no relationship between final learning and the generalization asymmetry score (Fig. 3b).

In contrast, the generalization data are consistent with the implicit recalibration hypothesis. When reaching to the two probe targets, the direction of the probe biases was the same as that observed for the training target, namely in the direction of the perturbation (Fig. 3a, b). We calculated the magnitude of generalization as the mean of the two probe target biases, as a percentage of the aftereffect (Fig. 3c). These values were significantly greater than zero for the Small ($p < 0.001$, $z = 3.7$, Ws = 205, $r = 0.19$) and Large ($p < 0.001$, $z = 4.0$, Ws = 379, $r = 0.14$) groups. In both groups, the amount of generalization was 83% of the bias observed for the training target (Small: IQR [53.3%, 100.5%]; Large: IQR [54.3%, 100.6%]). In summary, while the interpretation of the generalization results is problematic
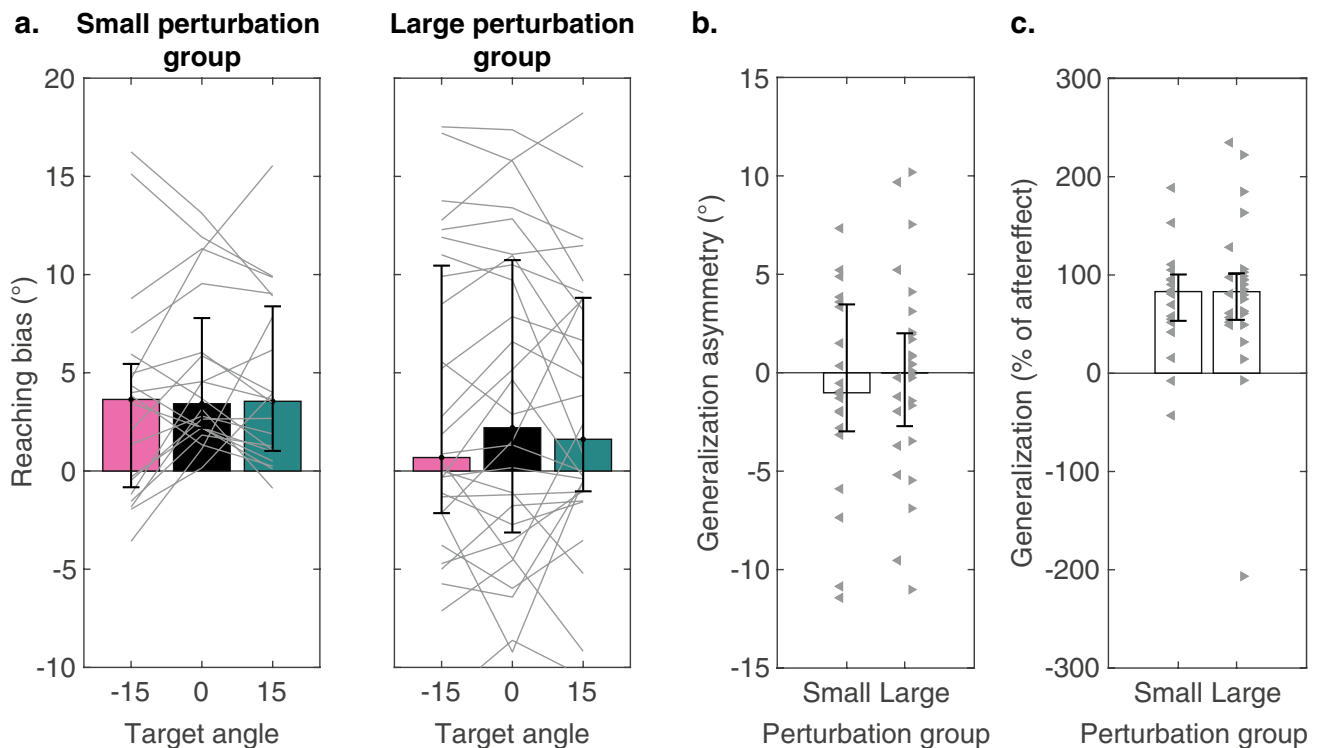

Fig. 3 Aftereffects and generalization of learning. Bars and error bars indicate medians and interquartile ranges. **a** Reaching biases for the training target (black) and two probe targets (see Fig. 1c). Thin lines indicate data from individual participants. **b** Asymmetry in reaching biases to probe targets. Dots indicate the individual participants. For the Large perturbation group, the left-pointing triangles indicate participants with final learning < 15°. The right-pointing triangles indicate participants with final learning > 15°. **c** Generalization quantified as a percentage of the aftereffect

for the Large group, the results for the Small group provide compelling support for the implicit recalibration hypothesis.

Awareness of the feedback perturbation

As expected, participants in the Small perturbation group were generally unaware that the reward zone had shifted over the course of the experiment. When asked to judge if they had been in the group with veridical feedback or shifted feedback, 60% reported that the feedback was not perturbed with an average confidence of 3.3 on a 7-point scale (Online Resource 3). When forced to choose between saying if they aimed left, right, or straight to the target during the training phase, 50% reported having aimed straight to the target and 50% reported aiming away from the target. However, of the latter, half reported aiming in the direction of the shifted reward zone and the other half reported aiming in the opposite direction. These survey data, in combination with the fact that all participants in the Small perturbation group showed a shift in reaching in the direction of the perturbation, provide compelling evidence that there was little if any awareness of the experimental manipulation nor use of a compensatory strategy.

A very different picture emerged from the survey data for the Large perturbation group. The majority (82%) reported that the feedback was perturbed with an average confidence of 4.8 on the 7-point scale. When asked whether they aimed left of, right of, or straight to the target during the training phase, 75% of the participants reported having aimed off target in the direction of the shifted reward zone, whereas 21% reported having aimed straight to the target. In summary, the survey data indicate that the participants in the Large perturbation group were aware of the experimental manipulation and adopted a re-aiming strategy to compensate for the shift in the reward zone. There was no clear relation between the questionnaire reports and aftereffects (Online Resource 3).

Discussion

In the present study, we examined whether binary feedback can induce implicit learning in response to shifts in a hidden reward zone. Based on the previous work (Codol et al. 2018; Holland et al. 2018, 2019), we expected that the learning would include an implicit component. Participants performed a center-out reaching task and were only provided binary feedback to indicate if the movement ended in a reward zone that gradually shifted to be centered 7.5° or 25° from the visual target, with the expectation that awareness of the perturbation would be minimal in the former and that the latter would entail some explicit component. During training, participants in both groups learned to compensate for the rotated feedback. When the feedback was removed

after training and participants were instructed to move to the target, their reaches were biased in the direction of learning, with an aftereffect of 2–3° in both groups. To test generalization, the no-feedback phase also included reaches to probe targets that flanked the training target. On these probe target trials, participants exhibited a shift in reach angle that was in the same direction as the shift associated with the training target. These results suggest that binary feedback can induce implicit reward-based motor learning and that this learning reflects implicit recalibration of a sensorimotor map.

Small and saturated implicit learning in response to binary feedback

Our study employed multiple approaches to prevent explicit processes from contaminating our assessment of implicit learning. First, we focused on the aftereffect in a phase without feedback and in which we provided explicit instructions to stop using any strategy that might have been used during the training. Second, we introduced the perturbation in a gradual manner, and most importantly, included a small perturbation group in which the displacement per step was within 1.5 standard deviations of baseline reach variability (Online resource 4) (Gaffin-Cahn et al. 2019). Thus, for this group, it is likely that behavioral changes during the training phase occurred implicitly. Third, we used questionnaires to directly assess awareness of the perturbation. The responses to the survey confirmed that, during the perturbation phase, awareness and strategy use were minimal in the Small perturbation group but high in the Large perturbation group.

We observed a small, but consistent aftereffect of around 2–3° in the Small and Large perturbation groups, evidence of implicit learning in response to binary feedback. The magnitude of the aftereffect for the Large group is smaller than that previously reported in the other studies using a perturbation of comparable size; for example, in Holland et al. (2018, 2019), the aftereffects in response to a perturbation of 25° were around 5° when including all participants (learners and non-learners). However, during an initial no-feedback aftereffect phase, Holland et al. instructed their participants to keep reaching as they had done during training. Subsequently, the participants were instructed to stop using a strategy. This protocol may have contaminated the final aftereffect measure by adding extra strategy trials and the challenge to switch between tasks.

The inclusion of the Small perturbation group not only provided a condition in which awareness should be minimized during the training phase, but also allowed us to directly compare how perturbation size impacted the magnitude of implicit learning from binary feedback. Interestingly, the size of the aftereffect did not scale with perturbation size. Indeed, in terms of mean value, the size was larger in the Small condition (3.4°) compared to the Large condition

(2.2°), although this difference was not significant. This null result was also observed in a secondary analysis in which we excluded the four participants in the Large condition who showed a negative final learning score.

While future testing is required to sample a broader range of perturbation sizes, the present results suggest that the magnitude of implicit learning from binary feedback is relatively small and saturates, at least for perturbations larger than 7.5°. A similar saturation is also observed for implicit learning from sensory prediction errors in response to perturbations ranging from 15° up to 90° (Bond and Taylor 2015; Morehead et al. 2017). However, the upper bound for implicit learning in response to sensory prediction errors is in the range 15° to 25° (Bond and Taylor 2015; Morehead et al. 2017).

The large variability observed in the performance of the Large perturbation group does limit what can be inferred about how implicit reward-based learning scales with the size of the perturbation. A substantial number of participants in this group failed to track the reward zone and, in general, these individuals had negative aftereffect scores that approximated their final learning score. Negative aftereffects were also observed in some participants in the Large group who successfully tracked the reward zone. As indicated by the survey data (see Online Resource 3), these participants were generally aware of the perturbation and invoked some form of strategy to aid performance. Assuming implicit reward-based learning has a small upper bound, success in fully tracking the perturbation would require discovery and maintenance of an aiming strategy. This is likely to be quite variable given the binary nature of the feedback and the absence of visual feedback of the hand. "Turning off" this strategy during the aftereffect phase would likely add additional variability.

Mechanisms of implicit learning in response to binary feedback

In the following section, we will consider the mechanisms underlying implicit learning in response to binary feedback. Similar to what has been reported in the studies of error-based learning (Bond and Taylor 2015; Morehead et al. 2017) and use-dependent learning (Tsay et al. 2022), implicit learning in response to binary feedback seems to saturate. However, there are notable differences between these three implicit forms of learning. While the magnitude of implicit use-dependent biases is similar to the magnitude of the aftereffect observed in the present study, the generalization pattern did not show any evidence of attraction toward the training location. As such, the current results fail to support the idea that implicit learning from binary feedback is a manifestation of use-dependent learning. On the other hand, while the generalization pattern is similar for binary

and cursor feedback, the magnitude of the binary feedback effect is much smaller than that observed in response to cursor feedback (Bond and Taylor 2015; Morehead et al. 2017). This size discrepancy makes it unlikely that binary feedback operates on similar mechanisms in inducing implicit recalibration of the sensorimotor map.

How, then, does binary feedback result in implicit learning? We outline three implicit recalibration hypotheses. First, implicit learning in response to binary feedback could be the result of motor recalibration, retuning the mapping between a visual target location and its associated movement. The contingency between action and reward outcome will lead to that action being associated with a new movement plan (Avraham et al. 2022). This hypothesis predicts that there is no sensory recalibration: training would not influence reports of where the visual target is perceived and perceived locations of the hand, so that they are similar before and after training. Second, implicit learning could be the result of visual recalibration of the target, i.e., a bias in the perceived location of the visual target. This hypothesis predicts visual sensory remapping: for example, if asked to report the perceived target location by reaching with the non-trained hand, we would observe a bias toward the reward zone (Simani et al. 2007). Third, implicit learning could be the result of proprioceptive recalibration, i.e., a bias in perceived hand position. This hypothesis predicts proprioceptive sensory remapping. For example, static reports of perceived hand position would be biased in the opposite direction of the perturbation (Tsay and Ivry 2022).

Future studies employing fine-grained psychophysical tests can evaluate the merits of these different hypotheses, asking if implicit learning in response to binary feedback originates from implicit recalibration of a sensory and/or motor mapping, and how this evolves over the course of learning.

Conclusion

Our data add to a growing body of evidence, indicating that motor learning encompasses multiple processes where both explicit and implicit processes drive behavioral changes (Kim et al. 2021; Morehead and Orban de Xivry 2021; Therrien and Wong 2022). The results provide compelling evidence of implicit learning in response to binary feedback and rule out that this effect is a form of use-dependent learning. Less clear is whether this implicit learning entails the same mechanisms, albeit in attenuated form, as occur during learning from sensory prediction errors, or reflects the operation of distinct, reward-based mechanisms.

Declarations

References

Avraham G, Taylor JA, Breska A, Ivry RB, McDougle SD (2022) Contextual effects in sensorimotor adaptation adhere to associative learning rules. Elife 11:e75801. https://doi.org/10.7554/elife.75801

Bao S, Lei Y (2022) Memory decay and generalization following distinct motor learning mechanisms. J Neurophysiol 128:1534–1545. https://doi.org/10.1152/jn.00105.2022

Bond KM, Taylor JA (2015) Flexible explicit but rigid implicit learning in a visuomotor adaptation task. J Neurophysiol 113(10):3836–3849. https://doi.org/10.1152/jn.00009.2015

Brainard DH (1997) The Psychophyics Toolbox. Spat vis 10(4):433–436

Brudner SN, Kethidi N, Graeupner D, Ivry RB, Taylor JA (2016) Delayed feedback during sensorimotor learning selectively disrupts adaptation but not strategy use. J Neurophysiol 115(3):1499–1511. https://doi.org/10.1152/jn.00066.2015

Cashaback JGA, Lao CK, Palidis DJ, Coltman SK, McGregor HR, Gribble PL (2019) The gradient of the reinforcement landscape influences sensorimotor learning. PLoS Comput Biol 15(3):e1006839. https://doi.org/10.1371/journal.pcbi.1006839

Codol O, Holland PJ, Galea JM (2018) The relationship between reinforcement and explicit control during visuomotor adaptation. Sci Rep 8(9121):1–11. https://doi.org/10.1038/s41598-018-27378-1

Diedrichsen J, White O, Newman D, Lally N (2010) Use-dependent and error-based learning of motor behaviors. J Neurosci 30(15):5159–5166. https://doi.org/10.1523/JNEUROSCI.5406-09.2010

Gaffin-Cahn E, Hudson TE, Landy MS (2019) Did I do that? Detecting a perturbation to visual feedback in a reaching task. J vis 19(1):5. https://doi.org/10.1167/19.1.5

Galea JM, Mallia E, Rothwell J, Diedrichsen J (2015) The effects of reward and punishment on motor skill learning. Nat Neurosci 18(4):597–604. https://doi.org/10.1016/j.cobeha.2017.11.011

Holland P, Codol O, Galea JM (2018) Contribution of explicit processes to reinforcement-based motor learning. J Neurophysiol 119(6):2241–2255. https://doi.org/10.1152/jn.00901.2017

Holland P, Codol O, Oxley E, Taylor M, Hamshere E, Joseph S, Huffer L, Galea J (2019) Domain-specific working memory, but not dopamine-related genetic variability, shapes reward-based motor learning. J Neurosci 49(47):9383–9396. https://doi.org/10.1523/JNEUROSCI.0583-19.2019

Huang VS, Haith A, Mazzoni P, Krakauer JW (2011) Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. Neuron 70(4):787–801. https://doi.org/10.1016/j.neuron.2011.04.012

Izawa J, Shadmehr R (2011) Learning from sensory and reward prediction errors during motor adaptation. PLoS Comput Biol 7(3):e1002012. https://doi.org/10.1371/journal.pcbi.1002012

Kim HE, Avraham G, Ivry RB (2021) The psychology of reaching: action selection, movement implementation, and sensorimotor learning. Annu Rev Psychol 72:61–95. https://doi.org/10.1146/annurev-psych-010419-051053

Kim HE, Morehead JR, Parvin DE, Moazzezi R, Ivry RB (2018) Invariant errors reveal limitations in motor correction rather than constraints on error sensitivity. Commun Biol. https://doi.org/10.1038/s42003-018-0021-y

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in psychtoolbox-3. Perception 36(14):1–16

Krakauer JW (2009) Motor learning and consolidation: the case of visuomotor rotation. Prog Motor Control 957:405–421. https://doi.org/10.1007/978-3-319-47313-0

Krakauer JW, Hadjiosif AM, Xu J, Wong AL, Haith AM (2019) Motor learning. Compr Physiol 9(2):613–663. https://doi.org/10.1002/cphy.c170043

Kuling IA, de Brouwer AJ, Smeets JBJ, Flanagan JR (2019) Correcting for natural visuo-proprioceptive matching errors based on reward as opposed to error feedback does not lead to higher retention. Exp Brain Res 237(3):735–741. https://doi.org/10.1007/s00221-018-5456-3

Leow LA, Marinovic W, de Rugy A, Carroll TJ (2018) Task errors contribute to implicit aftereffects in sensorimotor adaptation. Eur J Neurosci 48(11):3397–3409. https://doi.org/10.1111/ejn.14213

Marinovic W, Poh E, De Rugy A, Carroll TJ (2017) Action history influences subsequent movement via two distinct processes. Elife 6:1–23. https://doi.org/10.7554/eLife.26713

Mawase F, Uehara S, Bastian AJ, Celnik P (2017) Motor learning enhances use-dependent plasticity. J Neurosci 37(10):2673–2685. https://doi.org/10.1523/JNEUROSCI.3303-16.2017

Mazzoni P, Krakauer JW (2006) An implicit plan overrides an explicit strategy during visuomotor adaptation. J Neurosci 26(14):3642–3645. https://doi.org/10.1523/JNEUROSCI.5317-05.2006

Morehead JR, Orban de Xivry J-J (2021) A synthesis of the many errors and learning processes of visuomotor adaptation. BioRxiv, James 1891 1–50

Morehead JR, Taylor JA, Parvin DE, Ivry RB (2017) Characteristics of implicit sensorimotor adaptation revealed by task-irrelevant

clamped feedback. J Cogn Neurosci 26(6):1–10. https://doi.org/10.1162/jocn

Nikooyan AA, Ahmed AA (2015) Reward feedback accelerates motor learning. J Neurophysiol 113(2):633–646. https://doi.org/10.1152/jn.00032.2014

Schween R, Hegele M (2017) Feedback delay attenuates implicit but facilitates explicit adjustments to a visuomotor rotation. Neurobiol Learn Mem 140:124–133. https://doi.org/10.1016/j.nlm.2017.02.015

Shadmehr R, Smith MA, Krakauer JW (2010) Error correction, sensory prediction, and adaptation in motor control. Annu Rev Neurosci 33:89–108. https://doi.org/10.1146/annurev-neuro-060909-153135

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. https://doi.org/10.1523/JNEUROSCI.2184-12.2012

Simani MC, McGuire LMM, Sabes PN (2007) Visual-shift adaptation is composed of separable sensory and task-dependent effects. J Neurophysiol 98(5):2827–2841. https://doi.org/10.1152/jn.00290.2007

Synofzik M, Lindner A, Thier P (2008) The cerebellum updates predictions about the visual consequences of one's behavior. Curr Biol 18(11):814–818. https://doi.org/10.1016/j.cub.2008.04.071

Taylor JA, Krakauer JW, Ivry RB (2014) Explicit and implicit contributions to learning in a sensorimotor adaptation task. J Neurosci 34(8):3023–3032. https://doi.org/10.1523/JNEUROSCI.3619-13.2014

Taylor JA, Ivry RB (2014) Cerebellar and prefrontal cortex contributions to adaptation, strategies, and reinforcement learning. In: Progress in brain research, 1st edn, vol 210. Elsevier B.V. https://doi.org/10.1016/B978-0-444-63356-9.00009-1

Therrien AS, Wong AL (2022) Mechanisms of human motor learning do not function independently. Front Hum Neurosci 15(January):1–9. https://doi.org/10.3389/fnhum.2021.785992

Therrien AS, Wolpert DM, Bastian AJ (2016) Effective Reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. Brain 139(1):101–114. https://doi.org/10.1093/brain/awv329

Therrien AS, Wolpert DM, Bastian AJ (2018) Increasing motor noise impairs reinforcement learning in healthy individuals. Eneuro 5(June):ENEURO.0050-18.2018. https://doi.org/10.1523/ENEURO.0050-18.2018

Tsay JS, Ivry RB (2022) Understanding implicit sensorimotor adaptation as a process of proprioceptive re-alignment. ELife 1–45.

Tsay JS, Kim HE, Saxena A, Parvin DE, Verstynen T, Ivry RB (2022) Dissociable use-dependent processes for volitional goal-directed reaching. Proc R Soc B 289:20220415. https://doi.org/10.21856/j-pep.2021.4.08

Tseng YW, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ (2007) Sensory prediction errors drive cerebellum-dependent adaptation of reaching. J Neurophysiol 98(1):54–62. https://doi.org/10.1152/jn.00266.2007

van der Kooij K, Brenner E, Van Beers RJ, Schot WD, Smeets JBJ (2013) Alignment to natural and imposed mismatches between the senses. J Neurophysiol 109(7):1890–1899. https://doi.org/10.1152/jn.00845.2012

van der Kooij K, Smeets JBJ (2018) Reward-based motor adaptation can generalize across actions. J Exp Psychol: Learn Mem Cogn 45(1):71–81. https://doi.org/10.1037/xlm0000573

van der Kooij K, van Mastrigt NM, Crowe EM, Smeets JBJ (2021) Learning a reach trajectory based on binary reward feedback. Sci Rep 11(1):1–15. https://doi.org/10.1038/s41598-020-80155-x

Verstynen T, Sabes PN (2011) How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. J Neurosci 31(27):10050–10059. https://doi.org/10.1523/JNEUROSCI.6525-10.2011