The contribution of explicit processes to reinforcement-based motor learning

Holland, Peter James; Codol, Olivier; Galea, Joseph M.

DOI:
10.1152/jn.00901.2017

Document Version
Peer reviewed version

Citation for published version (Harvard):
Holland, PJ, Codol, O & Galea, JM 2018, 'The contribution of explicit processes to reinforcement-based motor learning', *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00901.2017

The contribution of explicit processes to reinforcement-based motor learning

Peter Holland[1*], Olivier Codol[1] & Joseph M Galea[1]

[1]School of Psychology, University of Birmingham, UK

*Corresponding Author

Author Contributions

PH and JM designed the experiments. PH collected the data. PH and OC analyzed the data. PH, OC and JM drafted the manuscript

Running Head

Explicit contribution to reinforcement-based motor learning

Correspondence

Peter Holland

School of Psychology

University of Birmingham, UK

Email: p.j.holland@bham.ac.uk

Phone: +44 (0) 121 414 2861

Abstract

Despite increasing interest in the role of reward in motor learning, the underlying mechanisms remain ill-defined. In particular, the contribution of explicit processes to reward-based motor learning is unclear. To address this, we examined subject's (n=30) ability to learn to compensate for a gradually introduced 25^0 visuomotor rotation with only reward-based feedback (binary success/failure). Only two-thirds of subjects (n=20) were successful at the maximum angle. The remaining subjects initially followed the rotation but after a variable number of trials began to reach at an insufficiently large angle and subsequently returned to near baseline performance (n=10). Furthermore, those that were successful accomplished this largely via a large explicit component, evidenced by a reduction in reach angle when asked to remove any strategy they employed. However, both groups displayed a small degree of remaining retention even after the removal of this explicit component. All subjects made greater and more variable changes in reach angle following incorrect (unrewarded) trials. However, subjects who failed to learn showed decreased sensitivity to errors, even in the initial period in which they followed the rotation, a pattern previously found in Parkinsonian patients. In a second experiment, the addition of a secondary mental rotation task completely abolished learning (n=10), whilst a control group replicated the results of the first experiment (n=10). These results emphasize a pivotal role of explicit processes during reinforcement-based motor learning, and the susceptibility of this form of learning to disruption has important implications for its potential therapeutic benefits.

Keywords

Motor Learning, Reward, Strategies, Visuomotor Adaptation

New & Noteworthy

We demonstrate that learning a visuomotor rotation with only reward-based feedback is principally accomplished via the development of a large explicit component. Furthermore, this form of learning is susceptible to disruption with a secondary task. The results suggest that future experiments utilizing reward-based feedback should aim to dissect the roles of implicit and explicit reinforcement learning systems. Therapeutic motor learning approaches based on reward should be aware of the sensitivity to disruption.

Introduction

The motor system's ability to adapt to changes in the environment is essential for maintaining accurate movements (Tseng et al., 2007). Such adaptive behavior is thought to involve several distinct learning systems (Haith and Krakauer, 2013; Izawa and Shadmehr, 2011; Smith et al., 2006). For example, the two-state model proposed by Smith et al. (2006) has been able to explain a range of results in force-field adaptation paradigms in which a force is applied to perturb a reaching movement. The model states that learning is accomplished via both 'fast' and 'slow' processes, the 'fast' process learns rapidly but has poor retention, whereas the 'slow' process learns more slowly but retains this information over a longer timescale. Subsequently using a visuomotor rotation paradigm, in which the visible direction of a cursor is rotated from the actual direction of hand movement, it has been suggested that the 'fast' process resembles explicit re-aiming whereas the 'slow' process is implicit (McDougle et al., 2015). The implicit aspect may be composed of several different processes (McDougle et al., 2015), the first and most widely researched being cerebellar adaptation (Izawa et al., 2012). However, additional processes such as use-dependent plasticity and reinforcement of actions that lead to task success are required to fully explain experimental findings (Huang et al., 2014). Haith and Krakauer (2013) have proposed a scheme based on these four processes that attempts a synthesis between the principles of motor learning and the distinction between model-based and model-free mechanisms proposed for reinforcement learning and decision-making (Doll et al., 2016).

The addition of rewarding feedback has proven beneficial in increasing retention of adaptation (Galea et al., 2015; Shmuelof et al., 2012; Therrien et al., 2016) and motor skills (Abe et al., 2011; Chen et al., 2018; Dayan et al., 2014). Findings such as these have generated interest in the possibility that the addition of reward to rehabilitation regimes may improve the length of time that adaptations are maintained after training (Goodman et al., 2014; Quattrocchi et al., 2017; Shmuelof et al., 2012). However, it is still unclear which of the multiple systems mediating motor learning reward may be acting on. Motor learning via purely reward based feedback is also possible and has been applied in two separate forms: binary and graded. Graded point based reward is often based on the distance of the reaching movement from the target and provides information about the magnitude but not the direction of the error (Manley et al., 2014; Nikooyan and Ahmed, 2015). Graded feedback has proved sufficient for learning abrupt rotations (Nikooyan and Ahmed, 2015), however, in certain conditions explicit awareness is required for successful learning (Manley et al., 2014). An alternative method is to only provide binary feedback in which the reward signals task success, such as hitting a target (Izawa and Shadmehr, 2011; Pekny et al., 2015; Therrien et al., 2016). In contrast to graded feedback, only gradually introduced perturbations have successfully been learnt via binary feedback alone (van der Kooij and Overvliet, 2016) and the contribution of explicit processes has yet to be examined.

In classical visuomotor adaptation, in which full visual feedback of the cursor is available, gradual adaptation is considered to be largely implicit (Galea et al., 2010). However, this may not be the case when only end-point feedback is provided (Saijo and Gomi, 2010). The question remains as to whether learning a gradually introduced visuomotor rotation based on binary feedback also mainly involves implicit processes. Various methods (Huberdeau et al., 2015) have been used to separate the implicit and explicit components of learning such as asking subjects to verbally report aiming directions (McDougle et al., 2015; Taylor et al., 2014) and forcing subjects to move at reduced reaction times (Haith et al., 2015; Leow et al., 2017). In the current paradigm, we assessed the contribution of explicit processes at the end of the learning period by removing all feedback but asking subjects to maintain their performance. Subsequently, we asked subjects to remove any explicit strategy they may have been using. Such an approach has previously been used to measure the relative implicit and explicit components of adaptation to different sizes of visuomotor rotations (Werner et al., 2015). It is important to note that here we define the explicit component to learning as the amount that participants could remove on request. Such a definition maybe more akin to awareness (Werner et al., 2015) or a form of cognitive control (Cavanagh et al., 2009), rather than an explicit strategy which is often defined as a subject's ability to verbalize the strategy they have employed. Therefore, we do not believe subjects had to be able to verbalize a strategy in order for learning to be defined as explicit.

Our second approach to investigating the explicit contribution to learning based on binary feedback was the introduction of a dual task in order to divide cognitive load and suppress the use of explicit processes. Dual task designs have previously successfully been employed to disrupt explicit processes in adaptation (Galea et al., 2010; Taylor and Thoroughman, 2007, 2008), sequence learning (Brown and Robertson, 2007) and motor skill learning (Liao and Masters, 2001). Various forms of dual task have been used such as counting auditory stimuli (Maxwell et al., 2001), repeating an auditory stimulus (Galea et al., 2010) or recalling words from a memorized list (Keisler and Shadmehr, 2010). We selected a mental rotation task based on using an electronic library of three-dimensional shapes (Peters and Battista, 2008; Shepard and Metzler, 1971). This particular task was selected in order to maximize the likelihood of interfering with the explicit re-aiming process. Indeed, it has previously been shown that both spatial working memory and mental rotation ability correlate with performance in the early 'fast' phase of adaptation (Anguera et al., 2009; Christou et al., 2016). Additionally, depletion of spatial working memory resources prior to visuomotor adaptation is detrimental to performance in the early phase (Anguera et al., 2012). Furthermore, the same prefrontal regions are activated during the early phase of adaptation and during the performance of a mental rotation task (Anguera et al., 2009). It has also been suggested that the explicit process of re-aiming in response to visuomotor rotations may involve a mental rotation of the required movement direction (Georgopoulos and Massey, 1987)

If the learning of a gradually introduced rotation via binary feedback is dominated by explicit processes, this should be evidenced by a large change in performance when subjects are asked to remove any strategy. Furthermore, the dual task should severely disrupt learning and could possibly unmask any implicit process.

Materials and Methods

Subjects

Sixty healthy volunteers aged between 18 and 35 participated in the study. Forty subjects (thirty-seven females, mean age = 19.9 years) completed experiment 1 and twenty (fifteen females, mean age = 21.6 years) in experiment 2. The number of subjects was selected to match the group size that is commonly employed within the field of motor learning (Morehead et al., 2017; Shmuelof et al., 2012; Therrien et al., 2016) and was not based on *a priori* power analysis. All subjects were right-handed with no history of neurological or motor impairment and had normal or corrected-normal vision. Volunteers were recruited from the undergraduate pool in the School of Psychology and wider student population at the University of Birmingham and all gave written informed consent. Subjects were remunerated with their choice of either course credits or money (£7.50/hour). The study was approved by the local ethics committee of the University of Birmingham and performed in accordance with those guidelines.

Experimental Protocol

A similar paradigm has previously been employed and the current protocol was designed to replicate this as closely as possible (Therrien et al., 2016). In addition to the rotation of 15°, we extended this paradigm to a 25° rotation. Subjects performed reaching movements with their right arm using a KINARM (B-KIN Technologies), Figure 1A. Subjects were seated in front of a horizontally placed mirror that reflected the visual stimuli presented on a screen above (60 Hz refresh rate). Reaching movements were performed in the horizontal plane whilst subjects held the handle of a robotic manipulandum, with the arm hidden from view by the mirror.

Experiment 1

Two different paradigms were employed in Experiment 1, both consisted of a gradually introduced rotation of the required angle of reach for a trial to be considered successful. The maximal extent of the rotation was either 15° (n=10) or 25° (n=30). The motivation for the use of the two different magnitudes of rotation was first to replicate the results of Therrien et al. (2016) and subsequently to investigate if subjects could successfully adapt to a larger angle (25°) than previously employed in binary feedback based motor learning. Subjects were required to learn the rotation on the basis of only binary feedback indicating if they had successfully hit the target region. After the rotation had reached the maximal extent, all feedback was extinguished and two further blocks of trials were performed to assay the level of retention and to what extent this was explicit in nature.

A total of 470 or 670 trials were performed for the 15° and 25° paradigms, respectively. Each trial followed an identical sequence. Initially a starting position was displayed on screen (red colored circle, 1cm radius), after subjects had moved the position of the cursor (white circle, 0.5cm radius) into the starting position, the starting position changed color from red to green. After a small delay (randomly generated, 500-700ms), in which subjects had to maintain the position of the cursor within the starting circle, a target (red circle, 1cm radius) appeared directly in front of the starting circle at a distance of 10cm. Subjects were instructed to make rapid 'shooting' movements that intercepted a visual target, they were instructed that they did not have to attempt to terminate their movement in the target but pass directly through it (Figure 1B). If the cursor intercepted a 'reward region' (±5.67°), initially centered on the visible target, the movement was considered successful and the target changed color from red to green and a large (8x8cm) green 'tick' was displayed at a distance of 20cm directly in front of the starting position (Figure 1C). However, if the cursor did not intercept the reward region the trial was considered unsuccessful and the visible target disappeared from view. Movement times, defined as the time from leaving the starting circle to reaching a radial distance of 10cm, were constrained to a range of 200-1000ms. Movements outside of this range but at the correct angle were counted as incorrect trials and no tick was displayed. As a visual cue, movements outside of the acceptable duration were signaled with a change of the target color, blue for too slow and yellow for too fast. After the completion of a reaching movement the robot returned the handle to the start position and subjects were instructed to passively allow this whilst maintaining their grip on the handle, during the passive movement subjects continued to receive no visual feedback of hand position. Reaction times, defined as the difference in time between the appearance of the target and the time at which the cursor left the starting circle, were limited to a maximum 600ms. If a movement was not initiated before this time, the target disappeared and the next trial began after a small delay and these trials were excluded from further analysis.

After an initial period of ten trials, in which the cursor position was constantly visible, for the remainder of the experiment it was extinguished. The only feedback subjects received was a binary (success/fail) signal indicating if the angle of reach was correct, in the form of a change of target color and the appearance of the tick. For an initial period of forty trials, the reward region remained centered on the position of the visual target, after this it was shifted in steps of 1° every twenty trials. The number of trials within the initial period and the rate of introduction of the rotation were identical in the 15° and 25° paradigms, only the total number of trials required to reach the maximum angle differed. This manipulation ensured that for a reaching movement to be considered correct it must be made at an increasingly rotated angle from the visual target (Figure 1C). Subjects were pseudo-randomly assigned to groups that received either a clockwise or counter-clockwise rotation. Once the reward region had reached the maximal angle, either 15° or 25°, it was held constant for an additional twenty trials. Subsequently, subjects were informed that they would no longer receive any feedback about their performance but that they should continue to perform in the same manner as before; this 'Maintain' block consisted of fifty trials. Following this, subjects were asked a series of simple questions to assay their awareness of the rotation; answers were noted by the experimenter. Firstly, subjects were asked 'Did you notice anything change during the course of the experiment?'. Secondly, 'Did you deliberately change anything about how you were performing the task?'. If the answer to the second question was affirmative they were asked a follow-up question 'What did you do?'. Subsequently all subjects were told 'During the task we secretly moved the position of the target that you had to hit. You will still not receive information on whether you hit the target or not but please try to move as you did at the start of the experiment'. Throughout the text we refer to this instruction as being asked to remove any strategy. Crucially subjects were not informed of the direction or magnitude of the rotation they had experienced. The final 'Remove' block consisted of fifty trials.

In order to test for any effects on retention due to the passage of time it took subjects to respond to the questions we performed a control experiment. The first 570 trials of the experiment were identical to the 25° paradigm previously described. However, at the end of the first block of fifty trials of no visual feedback (Maintain 1 block) subjects were asked to respond verbally to two questions from the BAS reward responsiveness section of the BIS/BAS questionnaire. These questions were selected on the basis of pilot experiments which demonstrated that they took approximately the same length of time to complete as the awareness related questions described previously. After subjects had responded to these questions they performed another block of fifty trials in which they received no feedback but were instructed to continue reaching in the same manner as before (Maintain 2 block). Subsequently, subjects were asked the task awareness questions, those that occurred in between Maintain and Remove blocks in the main experiment. The answers were noted down by the experimenter and subjects were then instructed to remove any strategy they had employed and then completed another fifty trials without visual or binary feedback (Remove block). For this experiment, we recruited an additional ten subjects who were successful in compensating for the final angle of rotation (fifteen in total recruited), the direction of the rotation was counterbalanced between subjects.

The position of the handle throughout the task was recorded at a sampling rate of 1 kHz and saved for offline analysis.

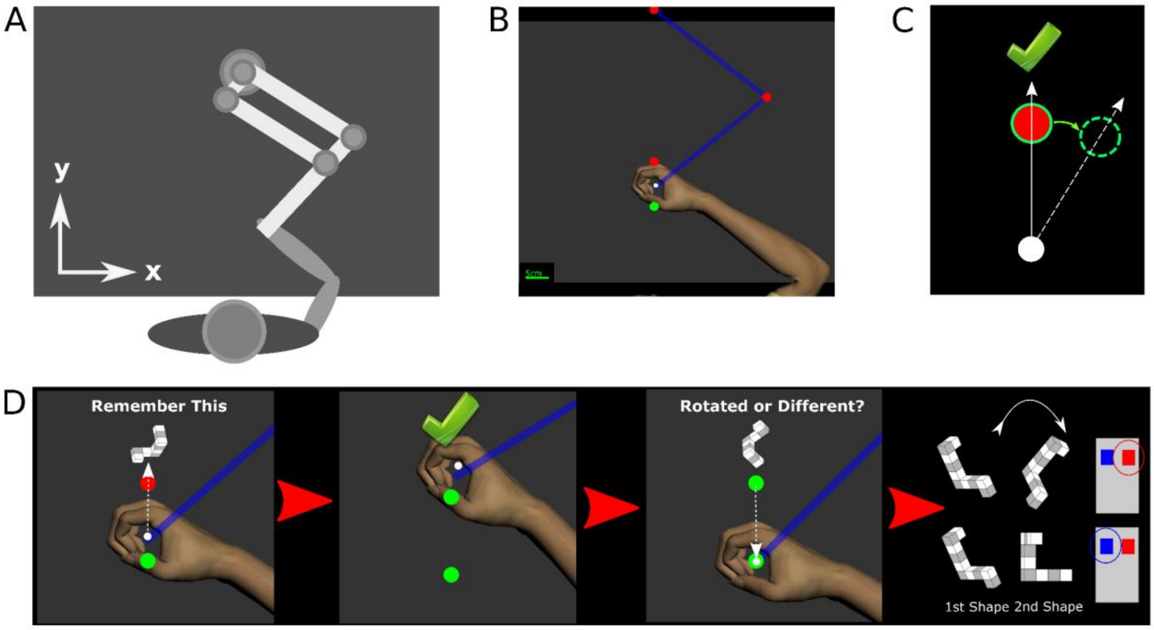


Figure 1. Experimental design. ***A,*** *Subjects held the handle of robotic manipulandum with their right hand, the position of the arm and handle was hidden from sight and feedback was provided on a horizontal screen.* ***B,*** *Subjects made 'shooting' movements from a starting position (green circle) towards a target (red circle), after the initial practice trials the position of the cursor (white circle) was no longer visible at any point.* ***C,*** *Successful trials were indicated to the subject with the display of a green tick after the cursor had passed through a region centered on the target, over the course of the paradigm the position of the reward region gradually moved (solid green circle to dashed green circle) whilst the visible target (red circle) remained in the central location. By the end of the learning period a successful reach (dotted white line) was rotated by a maximum of either 15° or 25°.* ***D,*** *Time-course of Experiment 2, at the same time as the target appeared on screen a 'shape' was also displayed slightly above it, the subject was asked to memorize this shape. After the reach was completed and the hand returned to the starting position subjects used their left hand to respond with a button press as to whether they believed the new shape shown on screen was a rotated version of the shape or an entirely different shape.*

Experiment 2

Experiment 2 comprised of the same reaching task as Experiment 1 but with the addition of a mental rotation dual task. The dual task required subjects to hold a three-dimensional shape in working memory for the duration of the reaching movement (Figure 1D). Subjects had to respond with a button press using their left hand to indicate if a shape displayed at the end of the reaching movement was a rotated version of a shape displayed at the time of target presentation or a different shape.

Shapes had the form of a series of connected cubes, alternately colored grey and white, they were selected from an electronic library designed on the basis of the Shepard and Metzler type stimuli (Peters and Battista, 2008; Shepard and Metzler, 1971). All rotations were performed within the plane of the screen, i.e. although the stimuli represented three-dimensional shapes all rotations were in two-dimensions. A subset of 26 shapes were selected from the library for use in this experiment and are available on https://osf.io/vwr7c/. The trial protocol was the same as that employed in Experiment 1 but at the time when the target circle appeared, a randomly selected shape from the subset was displayed in an 8x8cm region at a position 20cm away from the starting position. Subjects were instructed to commit this shape to memory. The shape remained visible on screen until the end of the reaching movement, the point at which the radial amplitude of the cursor exceeded 10cm. The shape was then extinguished and the same binary feedback as employed in Experiment 1 was displayed. After the robot had guided the handle back to the starting position a second shape was displayed in the same position as the first. In half of the trials this was an identical shape to the first one but had undergone a rotation selected at random from a uniform distribution of 0-360°, in the other half of trials it was a different shape selected at random from the library. The order of trials in which the shape was either rotated or different was randomized and subjects had a maximum of 2s to respond. Subjects in the Dual Task group (n=10) were instructed to press the right-sided button of two buttons on a button box held in their left hand if they believed the second shape to be a rotated version of the first one and the left-sided button if they believed it was a different shape. Importantly subjects were given no feedback on their performance in the dual task but were informed prior to the experiment that this would be monitored, the responses were recorded and analyzed offline. This design was selected in order to avoid any interfering effects of rewarding feedback from the dual task with the binary feedback in the reaching task. As a control, another group of subjects received identical visual stimuli but were instructed to press a random button of the two on each trial. Subjects were pseudo-randomly assigned to either the Control or Dual Task groups.

For Experiment 2, the familiarization period at the start of the experiment, in which the position of the cursor was visible, was extended to twenty trials in order for subjects to have sufficient time to acclimatize to the additional timing requirements of the button press. The paradigm subsequently followed that of Experiment 1 with a maximal angular rotation of 25°.

Data Analysis

All data analysis was performed with custom written routines in MATLAB (The Mathworks) and extracted data and all code required to reproduce the analysis and figures in this paper are freely available on (https://osf.io/vwr7c/).

The end point angle of each reaching movement was calculated either at the time that the cursor intercepted the reward region or in the case of incorrect trials when the cursor reached a radial amplitude of 10cm. An angle of zero degrees was defined as a movement directly ahead, i.e. toward the visible target position. A positive angle of rotation was defined as a clockwise shift of the reward region, and reach angles and target positions for the counter-clockwise rotation were sign-transformed to positive values for comparability. The 'Baseline' period was defined as the first forty trials without visual feedback of the cursor, during which the reward region was centered on the visual target. Subjects were considered to have successfully learnt the rotation if the mean end point angle of the reaching movements fell within the reward region during the last twenty trials before the 'Maintain' period, a time at which the rotation was held constant at its maximal value.

During the retention phase of the experiment (last one hundred trials), we calculated the amount of retention that could be accounted for by explicit and implicit processes. A subject's implicit retention was defined as the difference between the mean reach angle in the final fifty trials ('Remove' blocks), after subjects had been instructed to remove any strategy they had been using, and mean reach angle during the 'Baseline' blocks. A subject's explicit retention was defined as the difference between the mean reach angle during the 'Maintain' blocks, the first fifty trials after removal of binary feedback in which subjects were instructed to continue reaching as before, and the implicit retention.

In order to analyze the effect of reward on subjects behavior, we conducted trial-by-trial analysis in a manner similar to one that has previously been employed for analysis of reaching performance in response to binary feedback (Pekny et al., 2015). The change in reach angle following trial *n*, $\Delta u^{(n)}$, was defined as the difference between consecutive trials:

$$\Delta u^{(n)} = u^{(n+1)} - u^{n}$$

Subsequently we examined the distributions of Δu following only rewarded (correct) or unrewarded (wrong) trials. The resulting distributions of Δu were non-normal and therefore we analyzed and report the median and median absolute deviation (MAD) of each subject's distributions. We also examined the absolute change in reach angle $|\Delta u|$, i.e. the magnitude of change regardless of direction.

In order to investigate the effects of a reward history spanning multiple trials we examined the $|\Delta u|$ following all possible combinations of success in the previous three trials. We first searched each subject's responses for the occurrence of all eight possible sequences of reward and calculated the mean change in reach angle following each. We then quantified this behavior using a model in which $|\Delta u|$ was a function of the outcome of the previous three trials as well as variability (ε) that could not be accounted for by the recent outcomes (Pekny et al., 2015):

$$|u(n)| = \alpha_0\big(1 - R(n)\big) + \alpha_1\big(1 - R(n-1)\big) + \alpha_2\big(1 - R(n-2)\big) + \varepsilon$$

In the above equation, R represents the presence of reward on a given trial with a value of 1 for a correct trial, $R(n)$ therefore represents the presence of reward on the previous trial with $R(n-1)$ and $R(n-2)$ the preceding two trials. The components α_0, α_1 and α_2 represent the sensitivity to the outcomes of these trials with higher values indicating subjects made larger changes in response to the outcome of that trial. The values of these components were estimated using the least squares error solution to the equation using the mean value of $|\Delta u|$ recorded for each sequence on a subject-by-subject basis. We repeated this analysis using $|\Delta u|$ of every occurrence of a sequence (i.e. trial-by-trial analysis rather than using a mean value), and obtained similar estimates for the components. The model fits for both methods are reported as R^2 values in the results section.

The verbal responses to the questions asked before the start of the 'Remove' block was noted down by the experimenter and analyzed offline. A subject's awareness of the perturbation and efforts to deliberately counter it were rated on a scale of 0, 0.5 and 1, with 0 indicating no awareness and 1 indicating full awareness, including deliberately aiming at a rotated angle. A score of 0.5 was given when subjects were aware of some change throughout the course of the experiment but could not accurately state the nature of the perturbation or what they changed about their movement to counter it.

Statistical Analysis

Statistical analysis was performed in MATLAB. In order to test for initial effects mixed design ANOVAs were used, with Group (25RotSucces, 25RotFail etc.) as the between-subjects factor and time-point (Baseline, 15° Block, Maintain etc.) or MeasuredVariable (Median Δu, Reward Component etc.) as the within-subjects factor. The Greenhouse-Geiser correction was applied in cases of violation of sphericity and corrected p-values and degrees of freedom are reported in the text. In cases in which a significant interaction was found in the ANOVA, post-hoc tests were performed to test for differences between groups at each TimePoint or MeasuredVariable. As data was often found to be non-normally distributed using Kolmogorov-Smirnov tests, the non-parametric Kruskal-Wallis test was applied throughout. In cases of a significant effect of group on an individual outcome measure, further pairwise comparisons of mean group ranks were employed and Bonferroni corrected p-values are reported in the text. For tests of a difference of a single group from zero, such as in testing for implicit learning, Wilcoxon-Signed Rank tests were employed and Bonferroni corrected p-values are reported in the text. A critical significance level of $\alpha=0.05$ was used to determine statistical significance. The probability density estimates displayed as shaded regions in distribution plot figures were estimated using a Gaussian kernel.

Results

Experiment 1: Successfully learning to compensate for a 25^{o} rotation includes a large explicit component

We first sought to investigate the size of a gradual introduced visuomotor rotation that subjects can learn based on binary feedback. All subjects who experienced the 15^0 rotation (15Rot group) learnt to fully compensate (Figure 2A). Successful compensation was defined as having a mean reach angle within the reward region in the final twenty trials before the retention phase. However, for the 25^0 group (25Rot, magenta group, Figure 2B), the average reach direction fell outside the reward region, indicating incomplete learning. Underlying the mean performance was a split in behavior: some subjects successfully learnt the full rotation, whereas one third of subjects did not. On the basis of this behavior, they were categorized into two subgroups: 25RotSuccess (red group, N=20) and 25RotFail (blue group, N=10), respectively.

Next, we compared reach angle for the three groups (15Rot, 25RotSuccess and 25RotFail) at specific time points in order to gain an understanding at which stage the difference emerged (Figure 2C, D). Despite no difference between groups at baseline ($H(2) = 4.03$, $p = 0.13$, Kruskal Wallis), a difference had emerged at 15° ($H(2) = 9.63$, $p = 0.008$; Figure 2C). Specifically, reach angle for the 25RotFail group was lower than both the 15Rot ($p = 0.022$) and the 25RotSuccess groups ($p = 0.014$). During the 'Maintain' phase, when binary feedback had been removed but subjects were instructed to continue reaching as before, there was a significant effect of group ($H(2) = 20.08$, $p < 0.001$; Figure 2B, C). Unsurprisingly, the 25RotSuccess group was greater than the 15Rot ($p = 0.002$) and the 25RotFail groups ($p < 0.001$). Crucially, after subjects were instructed to remove any strategy and reach as they did at the beginning of the experiment, there was no difference between the groups ($H(2) = 0.78$, $p = 0.68$; Figure 2B, C). Analysis of the reach angles during the paradigm revealed that even at a rotation of 15° there was divergence between the 25RotFail and 25RotSuccess groups. Furthermore, the instruction to remove any strategy resulted in a return to a similar level of performance across all three groups.

We probed the nature of learning by calculating the implicit and explicit components of retention (Figure 2D). Implicit retention reflected the retention after removal of any strategies, whereas Explicit retention represented the change in behavior accounted for by the removal of strategies. The Explicit component of the 25RotSuccess group was greater than both 15Rot ($p = 0.006$) and 25RotFail ($p = 0.006$). Furthermore, only the 25RotSuccess ($Z = 210$, $p < 0.001$) group had a significant Explicit component to their retention. Whilst there was no effect of Group on the Implicit component ($H(2) = 1.84$, $p = 0.40$), both groups in the 25° paradigm showed a significant difference from 0 (25RotSuccess, $Z = 193$, $p = 0.001$; 25RotFail, $Z = 48$, $p = 0.014$), however, the 15Rot group was no longer significant after correction for multiple comparisons ($Z = 48$, uncorrected $p = 0.037$, corrected $p = 0.111$). Therefore, whilst all three groups showed a similar small level of implicit retention, only the subjects who successfully learnt the 25° rotation showed evidence for explicit learning. Whilst at a group level there was no evidence for an explicit component to retention in either the 15Rot or 25RotFail groups, there was variability within the groups with 2 subjects in each group displaying Explicit components greater than 10°.

It is possible that the reduction in reach angle observed between the 'Maintain' and 'Remove' blocks in the 25RotSuccess group could be accounted for by the decay of a labile memory during the time in which the awareness questions were asked (Smith et al., 2006). In the 25Rot paradigm, the time between the end of the 'Maintain' block and the start of the 'Remove' block was 37.16±8.49s. The time taken for the two control questions between the 'Maintain 1' and 'Maintain 2' blocks for the ten subjects in 25RotControl group was 49.48±8.63s, and for the awareness questions and instruction to remove strategy between 'Maintain 2' and 'Remove' was 45.80±13.38s. There was no significant difference between the length of time taken for either set of questions in the 25RotControl group and those in the 25Rot group ($H(2) = 5.47$, $p = 0.065$; Figure 2E). Crucially, we observed no difference in reach angle between 'Maintain 1' and 'Maintain 2' ($Z=36$, $p=0.432$). However, there was a clear reduction in reach angle following the instruction to remove any strategy between 'Maintain 2' and 'Remove' ($Z=52$, $p=0.010$). These results indicate that the passage of time is not the critical factor causing the reduction in reach angle observed, but rather it is the instruction to remove any strategy subjects had employed.

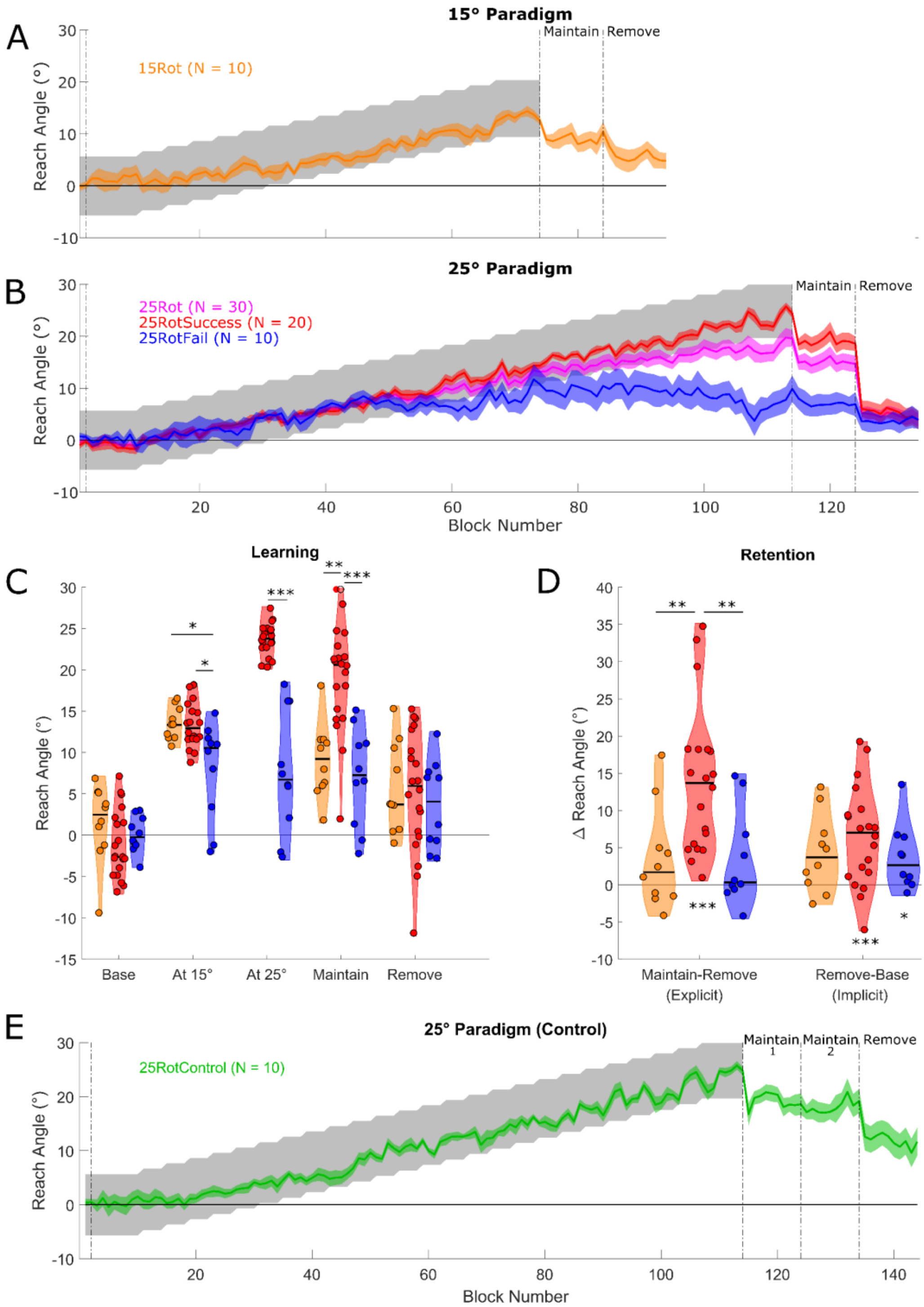


Figure 2. Experiment 1: group performance. ***A***, *Reach angle averaged over blocks of 5 trials, solid colored lines represent the mean of each group and the shaded region represents SEM. The average behavior of subjects in the 15Rot paradigm (Orange) fell consistently within the rewarded region (grey shaded region) indicating successful learning.* ***B***, *Average reach angle over blocks for all subjects in the 25Rot paradigm (magenta) and also the same subjects split into two groups based on*

success at the final angle (25RotSuccess – red, 25RotFail – blue). ***C****, Distribution plots displaying the reach angles for subjects in the three groups at various timepoints throughout the experiment with individual data points overlaid on an estimate of the distribution. Horizontal black line in the distribution represents the group median****. D****, Distribution plots of the computed variables of Implicit ('Remove-Baseline') and Explicit ('Maintain-Implicit') retention. Significance stars above horizontal black bars indicate differences between the groups (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Significance stars below the distributions represent a significant difference from zero.* ***E,*** *Reach angle averaged over blocks of 5 trials for subjects in the 25RotControl group. There was no reduction in reach angle during the time taken for the control questions between Maintain 1 and Maintain 2 blocks. However, when subjects were subsequently asked to remove their strategy, the period between Maintain 2 and Remove blocks, a significant reduction in reach angle was observed.*

In order to understand the mechanism of learning, and how this might differ between the 25RotSuccess and 25RotFail groups, we examined trial-by-trial behavior. Two distinct types of behavior were apparent (Figure 3). Behavior in those that failed (Figure 3B) was initially similar to successful subjects (Figure 3A), but at some point subjects began to fail to reach at a sufficient angle. Subsequently the angle of reach began to decline over further trials, despite a continued lack of reward. However, given the length of the paradigm it unclear if this reduction was limited to the angle of the last successful trial they experienced or would have continued to baseline levels given more trials. The angles at which subjects in the 25RotFail group failed varied (mean=13.0 ± 5.1^0), but all displayed the same pattern of return to baseline (Figure 3C). Given the apparently similar behavior in the initial learning stage, it is important to know whether there are differences even at this early stage. To this end, we only included trials in the initial successful period for the 25RotFail group in all subsequent analysis of trial-by-trial behavior, i.e. trials on the left-hand side of the vertical colored line for each subject (Figure 3C). For the 25RotSuccess and 15Rot groups all trials during the learning period were analyzed. Crucially, there was no difference in the percentage of correct trials within this period between the groups ($H(2) = 2.19$, $p = 0.33$).

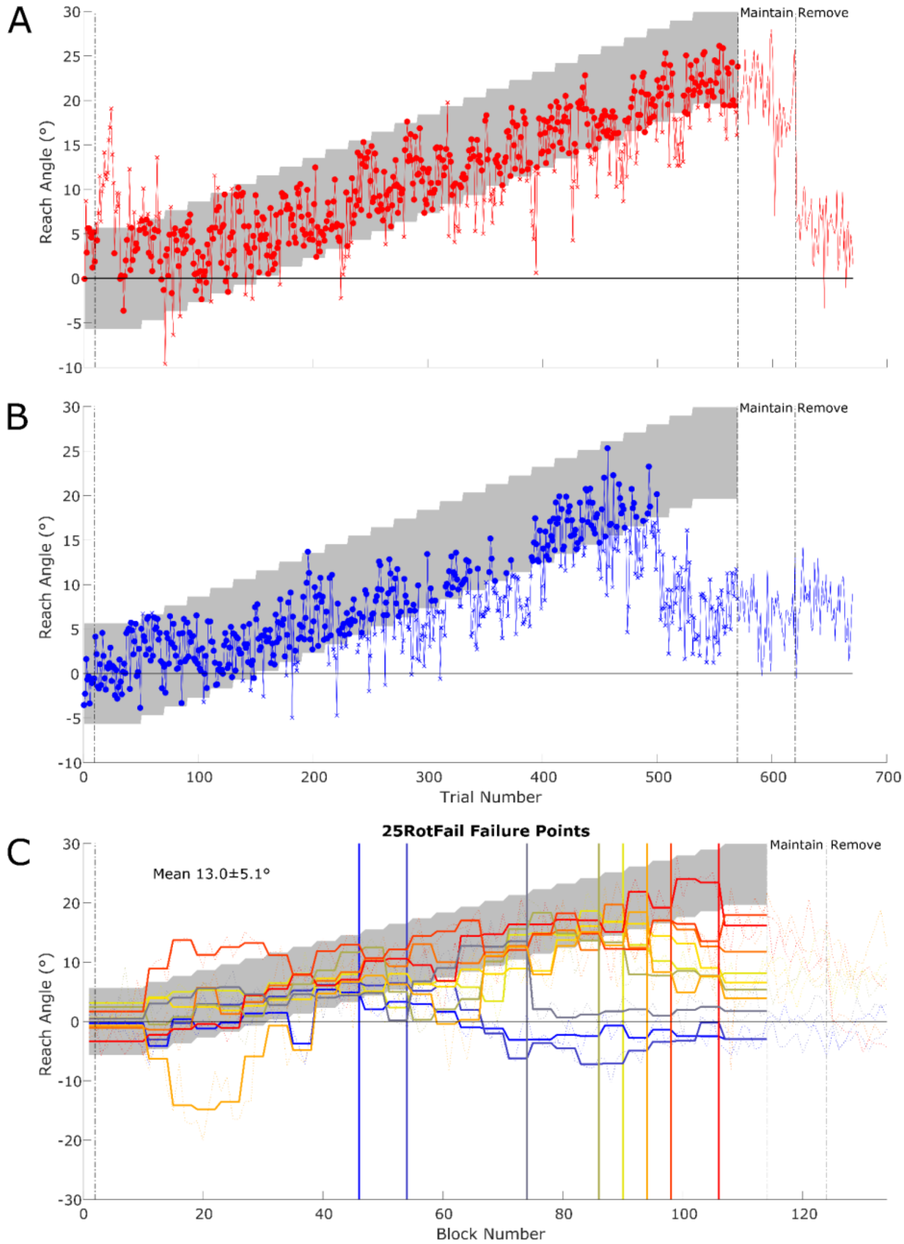


Figure 3. Experiment 1: trial-by-trial behavior. *Example of trial by trial reach angles from a subject who was successful at the final angle (**A**) and one who was unsuccessful (**B**). In each case rewarded trials are indicated with a circular marker and non-rewarded trials with a 'x'. The grey shaded region indicates the reward region. **C,** Failure points for subjects in the 25RotFail group, thick lines are the mean reach angle for each subject at each rotation angle, thin lines represent mean of each block (average of 5 trials), colors go from hot to cold matching failure angles ranging from high to low. Vertical lines represent the last angle at which mean reach fell within rewarded region for each subject. The mean and standard deviation of all angles of failure is displayed as text.*

Next, we examined if changes in reach angle were affected by the outcome of the previous trial. A similar analysis has been employed previously (Pekny et al., 2015). We examined the distributions of Δu following only rewarded (Correct) or unrewarded (Wrong) trials. The resulting distributions of Δu were non-normal and therefore we report the median and median absolute deviation from the median (MAD). Whilst the median Δu was greater following unrewarded trials ($F(1,37) = 119.80$, $p < 0.001$; Figure 4A), this effect was similar across groups ($F(2,37) = 1.18$, $p = 0.64$). Similarly, the MAD of Δu was also greater following Wrong trials, indicating that not only did all groups make larger changes in reach angle but also that there was greater variability in these changes (Figure 4B). Despite a significant interaction with Group ($F(2,37) = 5.32$, $p = 0.019$), the trend for a higher MAD of Δu following Wrong trials for the 25RotSuccess group (Figure 4B) did not reach significance after correction for multiple comparisons ($H(2) = 5.63$, $p = 0.06$). Subsequently we repeated the analysis but considered the absolute change in reach angle ($|\Delta u|$, Figure 4C, D). Here there was a significant interaction with Group for both median $|\Delta u|$ ($F(2,37) = 7.89$, $p = 0.003$) and MAD of $|\Delta u|$ ($F(2,37) = 7.39$, $p = 0.004$) following Wrong trials. Post-hoc tests revealed that the 25RotSuccess group displayed a significantly greater median $|\Delta u|$ ($p = 0.024$) and MAD of $|\Delta u|$ ($p = 0.035$) than the 25RotFail group. There was no difference between the groups in the magnitude or variability of the change in reach angle after correct trials. The analysis of the absolute changes in reach angle reveal that even during the period in which they are successful, the 25RotFail group made smaller and less variable changes following unrewarded trials.

In addition to the effect of the previous trial it is possible that subjects were sensitive to a history of outcomes spanning multiple previous trials (Pekny et al., 2015). In order to investigate the effects of reward history, we examined the $|\Delta u|$ following all possible combinations of success in the previous three trials (Figure 4E). We quantified this behavior using a model in which $|\Delta u|$ was a function of the outcome of the previous three trials. The components α_0, α_1 and α_2 represent the sensitivity to the outcome of the last three trials with α_0 being the most recent (Figure 4F), ε represents variability that could not be accounted for by the recent outcomes. There was an interaction between component and group (F(3.49,64.51) = 4.49, p = 0.004). All groups were most sensitive to the most recent trial outcome (α_0) with the 25RotSuccess group displaying significantly greater change than 25RotFail (p = 0.001). There was no difference between groups for other components indicating that differences in behavior were driven by the sensitivity to the outcome of the most recent trial. R^2 values for model fits based on the mean $|\Delta u|$ of each sequence had a mean of 0.90 and a range of 0.67 to 0.99, model fits based on a trial by trial basis had a mean R^2 of 0.39 and a range of 0.15 to 0.57. From these results it becomes apparent that, even in the initial period of success, subjects who will go on to fail to learn the full rotation show a decreased sensitivity to errors.

There was no difference between groups for either movement time (H(2) = 4.82, p = 0.090) or reaction time (H(2) = 4.01, p = 0.13). The mean of the median movement times across subjects was 0.38±0.08s. Additionally, within the 25RotFail group reaction and movement times did not differ before and after the point of failure (Z = 28, p = 1 and Z = 40, p = 0.23 respectively). In response to the questions asked to probe awareness, we found no significant difference between the groups ($\chi^2(2)$ = 3.75, p = 0.15). However, within the 25RotSuccess group there was a significantly non-uniform distribution of answers (χ2(2) = 9.1, p = 0.005) with 60% of participants reporting a specific strategy to counter the rotation and only one reporting not to notice any change. The remainder of subjects reported some awareness of a change (categorized as 0.5 on our scale), or an explicit effort to counter it, but were often not confident in describing the change or could not easily verbalize their strategy. There was no difference between the subjects reporting full or partial awareness in terms of the quantified Explicit component to retention (Z = 123, p = 0.837).

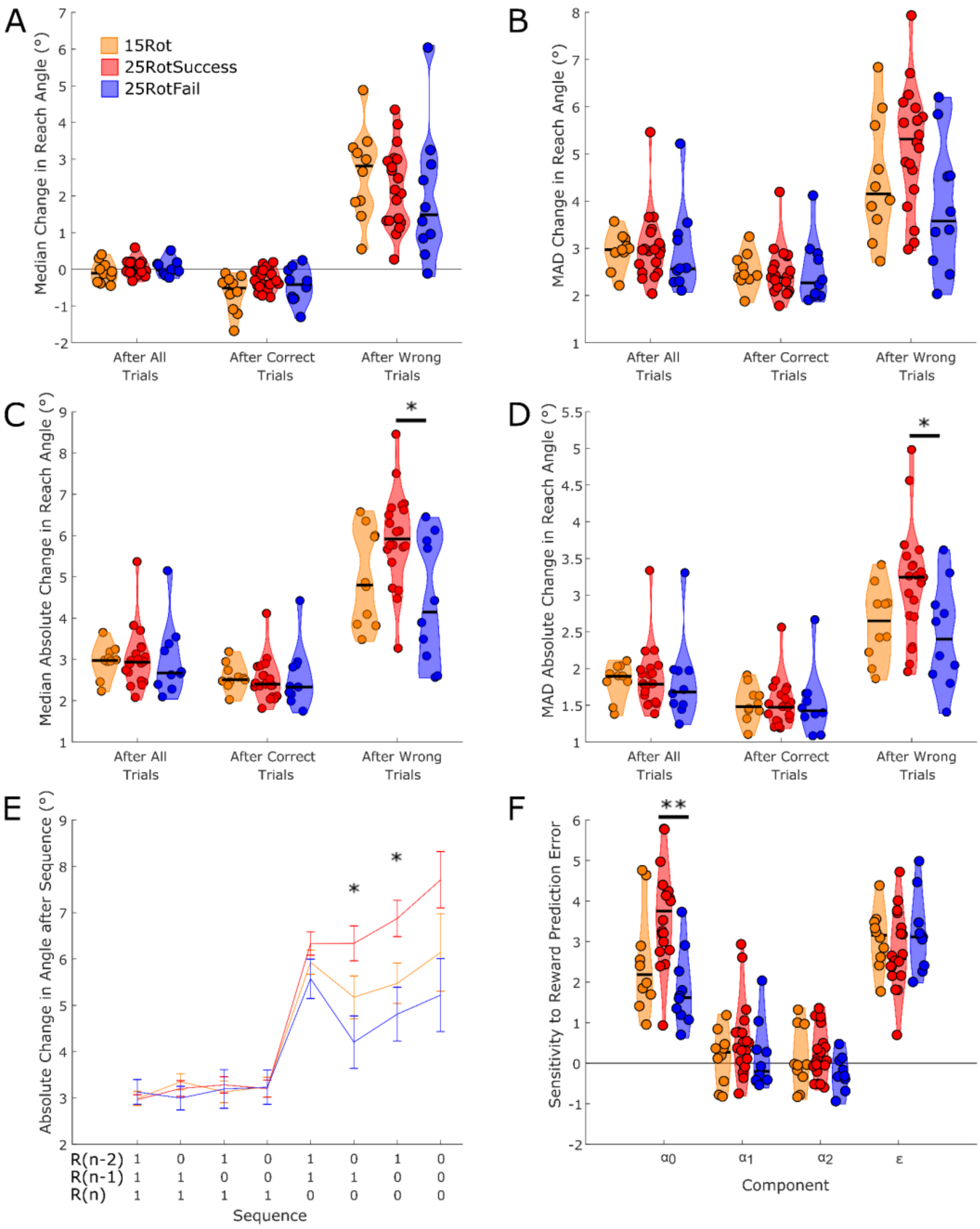


Figure 4. Experiment 1: performance after correct and incorrect trials. *Analysis of the effects of the success of the previous trial and reward history on trial by trial changes in reach angle for the three groups in Experiment 1 (15Rot – Orange, 25RotSuccess – Red, 25RotFail – Blue). Median (**A**) and MAD (**B**) of change in reach angle separated by the success of the previous trial. Median (**C**) and MAD (**D**) of the absolute change in reach angle separated by the success of the previous trial.* ***E****, The absolute change in reach angle following all combinations of trial success over the previous three trials.* ***F****, Sensitivity to the outcomes of each of the previous trials. Significance stars above horizontal black bars indicate differences between the groups (* P < 0.05, ** P < 0.01).*

Experiment 2: Addition of a dual task prevents learning

Following the finding of Experiment 1 that successful reinforcement-based motor learning involves a strong explicit component, we sought to investigate if it was possible to disrupt learning by dividing cognitive load. To this end, we required subjects to hold a shape in memory during the period of movement (Figure 1D).

The DualTask (N=10) group displayed little learning and none successfully compensated for the maximum rotation (Green group, Figure 5A). As in Experiment 1, the Control (N=10) group on average fell short of complete learning (Purple group, Figure 5A, B), indicated by the mean reach direction falling outside the reward region in the final learning blocks. However, the average of the group obscures a similar split in behavior with only six subjects successfully learning the full rotation and four failing to do so, which we will label (ControlSuccess and ControlFail, respectively; Figure 5B).

Examining performance in the same time periods as Experiment 1 (Figure 5C) revealed no difference between the three groups at baseline ($H(2) = 0.38$, $p = 0.83$). However, by the time the angle of rotation had increased to 15° a significant difference had already emerged ($H(2) = 6.88$, $p = 0.03$), with the DualTask group displaying lower reach angle than ControlSuccess ($p = 0.011$).

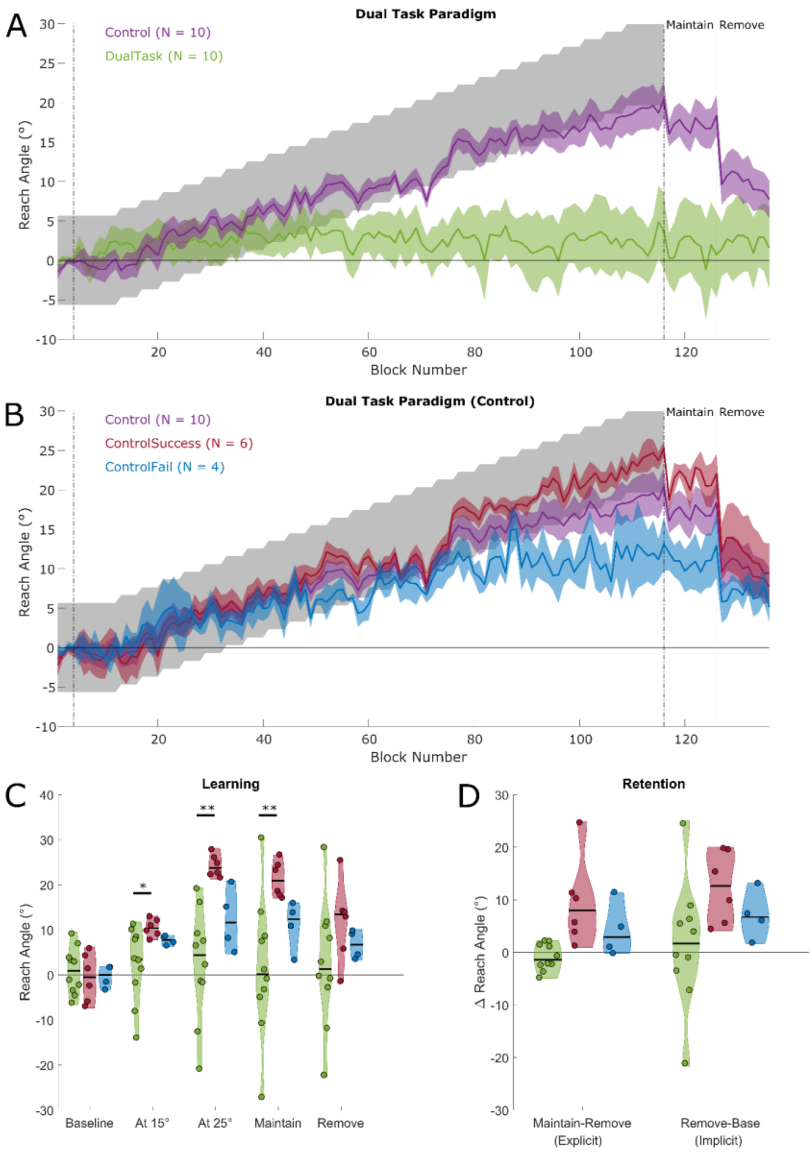


Figure 5. Experiment 2: group performance. *Change in reach angle over blocks (average of 5 trials) during the dual task experiment.* ***A****, Group performance for the DualTask (Green) and Control (Purple) task groups, the line indicates the mean and shaded region the SEM. The grey shaded region represents the reward region.* ***B****, the split of the control task group into ControlSuccess (Dark Red) and ControlFail (Blue).* ***C****, Distribution plots displaying the performance at different time points for the dual task, and split control groups. The shaded region represents an estimation of the distribution and is overlaid with data for each individual subject.* ***D****, Distribution plots of the difference in reach angle during retention phases indicating the implicit and explicit components of retention. Significance stars above horizontal black bars indicate differences between the groups (* $P < 0.05$, ** $P < 0.01$).*

As can be seen from the performance of individuals in the DualTask group (Figure 6), there were very few correct trials (mean angle of failure 6.0°) rendering the analysis of trials within the successful period employed for Experiment 1 invalid. Despite this limitation for the DualTask group, the analysis could still elucidate differences between the ControlSuccess and ControlFail groups and reassuringly the mean angle of failure in ControlFail group is 13°, similar to Experiment 1. However, the small group numbers preclude statistical comparison between the ControlSuccess and ControlFail groups but the pattern of behavior was visually similar to that in Experiment 1 (Figure 7). Overall the analysis of sensitivity to reward history produced remarkably similar results to Experiment 1 with the primary difference between those who learn and those who fail to do so being the sensitivity to the outcome of the most recent trial (Figure 7F).

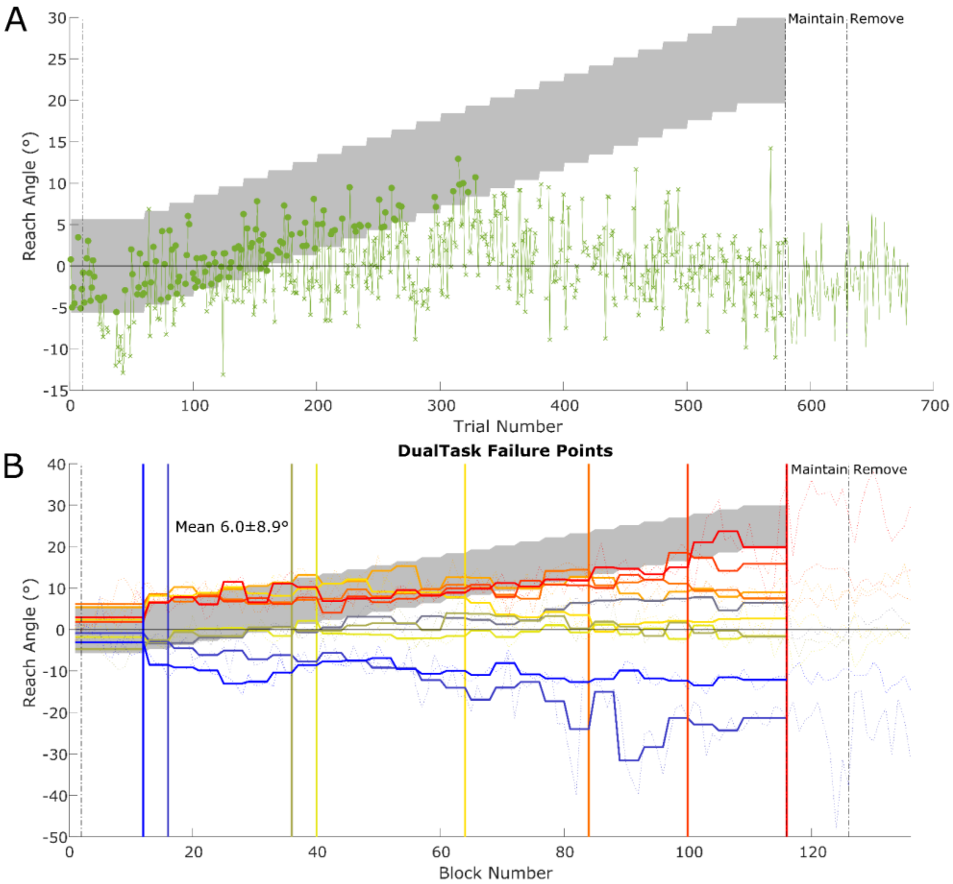


Figure 6. Experiment 2: trial-by-trial behavior. *Example of trial by trial reach angles from a subject performing the dual task (**A**) rewarded trials are indicated with a circular marker and non-rewarded trials with a 'x'. The grey shaded region represents the reward region.* ***B,*** *Failure points for subjects in the DualTask group, thick lines are the mean reach angle for each subject at each rotation angle, thin lines represent mean of each block, colors go from hot to cold matching failure angles ranging from high to low. Vertical lines represent the last angle at which mean reach fell within rewarded region for each subject. Th mean and standard deviation of the angle of failure is reported as text in the figure.*

Finally, the DualTask subjects successfully engaged in the task mental rotation task as evidenced by a significant difference in percentage of correct button presses ($H(2) = 15.30$, $p < 0.001$). The DualTask group responded correctly (67.21 ± 3.60%) more than either the ControlSuccess ($p = 0.014$) and the ControlFail ($p = 0.002$) groups. Engagement in the DualTask increased reaction time when compared to ControlSuccess ($p = 0.007$). There was no effect of Group on movement time ($H(2) = 0.33$, $p = 0.84$).

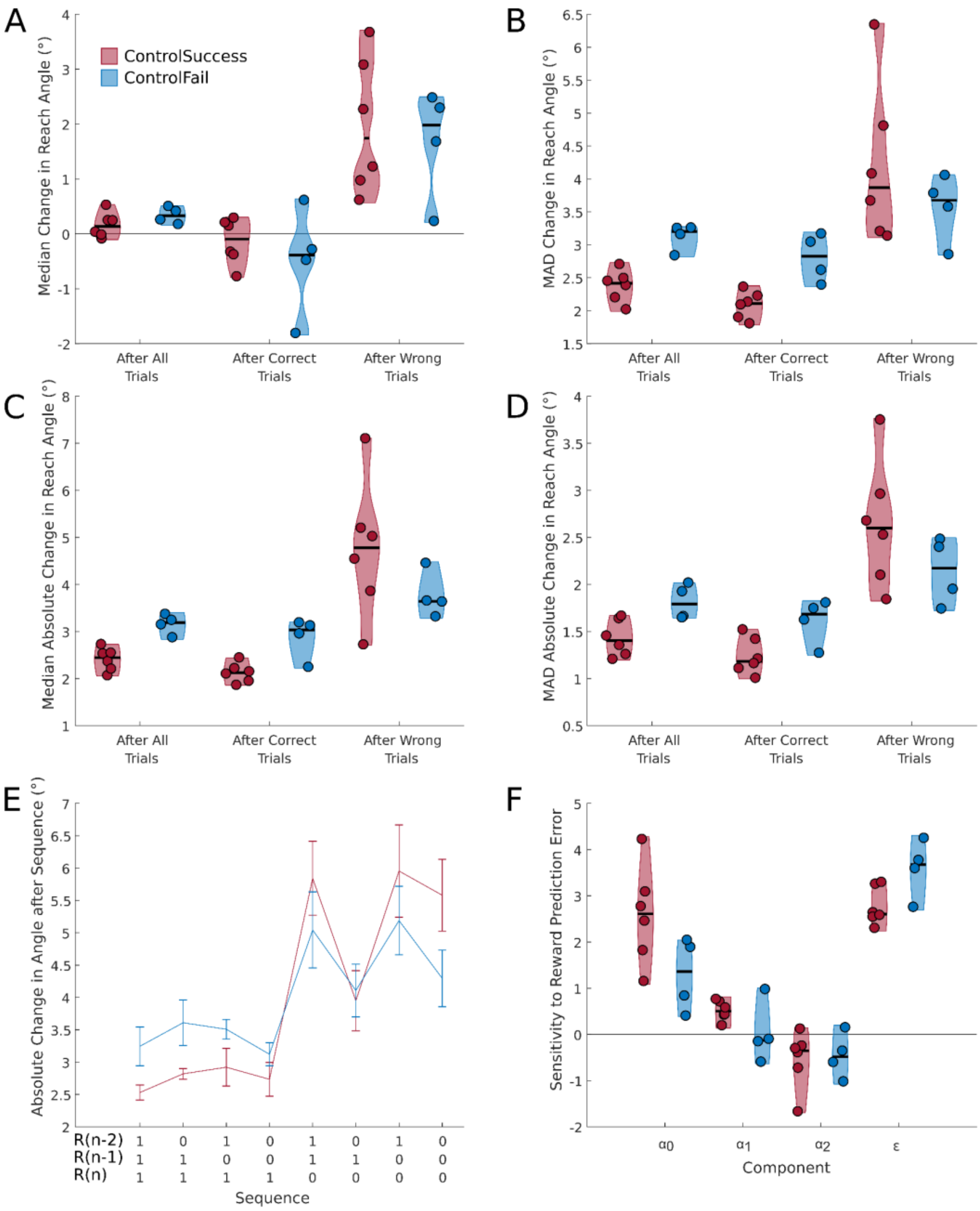


Figure 7. Experiment 2: performance after correct and incorrect trials*. Analysis of the effects of the success of the previous trial and reward history on trial by trial changes in reach angle for the two groups performing the control task in Experiment 2. Distribution plots for median (**A**) and MAD (**B**) of change in reach angle separated by the success of the previous trial. Median (**C**) and MAD (**D**) of the absolute change in reach angle separated by the success of the previous trial. **E**, the absolute*

change in reach angle following all combinations of trial success over the previous three trials. ***F****, sensitivity to the outcomes of each of the previous trials.*

Discussion

The role of explicit processes during reinforcement-based motor learning was previously unclear. Here, we reveal that successfully learning to compensate for large, gradually introduced, rotations based on binary (reinforcement-based) feedback involves the development of a strong explicit component, and that not all subjects are able to do so. In both Experiment 1 and the Control group of Experiment 2 only two thirds of subjects were able to successfully learn a large perturbation, and those that did accomplished this principally via explicit processes. Analysis of the trial-by-trial behavior indicated that subjects adjusted their motor commands mainly in response to incorrect trials, and that they were most sensitive to errors made in the most recent trial. Subjects who would go on to fail to learn the full rotation exhibited reduced sensitivity to errors, even in the initial period in which they successfully followed the rotation. Further evidence for the explicit nature of the learning in this task was provided by Experiment 2, where increasing cognitive load via the addition of a dual task prevented learning.

Previous experiments investigating the learning of rotations based on binary feedback have employed relatively small angles (Izawa and Shadmehr, 2011; Pekny et al., 2015; Therrien et al., 2016), with the 15° rotation used by Therrien et al. (2016) the largest reported to date. Indeed, when a rotation of 15° was used in Experiment 1 all subjects were successful in fully compensating for the visual rotation. Furthermore, there was no evidence for an explicit component to retention in the subjects who learnt the 15° rotation. In contrast, successful subjects in both experiments with a 25° rotation demonstrated a large explicit component to the learning, evidenced by a large reduction in the reach angle when asked to remove any strategy. It could therefore be speculated that multiple mechanisms might be available when learning from binary feedback, but that if the size of the perturbation exceeds a certain magnitude an explicit process is required to compensate for it. Previously it has been suggested that additional learning mechanisms are recruited in response to gradually introduced visuomotor rotations when only end-point feedback is available, (Izawa and Shadmehr, 2011; Saijo and Gomi, 2010). Indeed Saijo and Gomi (2010) suggest, on the basis of an increase in reaction times, that explicit changes in motor planning occur in this paradigm. Furthermore, similarly to the results presented here, the authors also find that not all subjects are able to accomplish this. However, none of the previous studies investigating learning of rotations based on binary feedback (Izawa and Shadmehr, 2011; Pekny et al., 2015; Therrien et al., 2016) have attempted to dissect the role of implicit and explicit processes. However, learning a rotation based on binary feedback was not accompanied by a change in perceived hand position, as was found when learning was based on full visual feedback of the cursor (Izawa and Shadmehr, 2011). This could be taken as evidence that the learning described by the authors was also explicit in nature in contrast to the implicit, cerebellar-driven, adaptation.

There is increasing appreciation of the role of explicit processes in traditional visuomotor adaptation paradigms, in which visibility of the cursor ensures that both direction and magnitude of the error are available (Bond and Taylor, 2015, 2017). The use of an 'error-clamp' technique has estimated the limit of implicit adaptation based on sensory prediction errors to be at around 15° (Morehead et al., 2017). Such an estimate is roughly in accordance with other estimates obtained either by the use of forcibly reduced movement preparation times (Haith et al., 2015; Leow et al., 2017), self-reporting of aiming directions (Bond and Taylor, 2015) or the difference between trials with and without an explicit component (Werner et al., 2015). It is important to note in our data that all groups, with the exception of those performing the dual task, display a small amount of retention even after the removal of the explicit component suggesting that there is some implicit aspect to the learning. Presumably the implicit learning process triggered in the current study is distinct from the sensory prediction error driven processes as here the error signal is binary in nature and provides no information about direction or magnitude of error. However, it is interesting that such implicit processes appear to be unable to compensate for rotations greater than 15-20°, with explicit mechanisms required for greater angles. Haith and Krakauer (2013) have proposed a theoretical framework in which model-based (strategic/explicit) and implicit model-free (reinforcement/use-dependent) learning processes contribute to motor learning. Our findings suggest that in the current paradigm these processes might be engaged but that implicit processes are limited in the size of rotation they can learn. It remains to be seen if this is a limitation of magnitude, as with learning from sensory prediction errors, or a limitation of speed. In other words, if the rotation was introduced more gradually or held constant for a longer period, could this implicit process account for all learning? It is unclear whether the implicit retention observed here reflects use-dependent learning, implicit reinforcement learning or a combination of both (Diedrichsen et al., 2010). However, the current experimental design does not allow us to dissociate between these possibilities. Interestingly, the greatest amount of implicit retention was observed in the 25RotControl group who had received an additional fifty no feedback trials. Given the lack of reward in these trials, this suggests that use-dependent learning at least contributes to the implicit retention observed.

We measured the explicit contribution to learning via the use of an include/exclude design similar to Werner et al. (2015), which probes the contribution at the end of learning. Other approaches such as asking subjects to verbally report the aiming direction (Taylor et al., 2014) have the advantage of probing the relative contributions of implicit and explicit processes throughout learning. However, it has been suggested that this method may increase the explicit component by priming subjects that re-aiming is beneficial (Leow et al., 2017; Taylor et al., 2014). Such priming may be particular powerful in paradigms like the current one as it has been shown that explicit awareness of the dimensions over which to explore is required for motor learning based on binary feedback (Manley et al., 2014). Alternatively, forcing subjects to respond at reduced reaction times can also suppresses the explicit component of adapting to a rotation (Haith et al., 2015; Leow et al., 2017). However, Leow et al. (2017) report that even at extremely short reaction times re-aiming to a single target, as used here, is still possible. In future, approaches such as measuring eye movement (Rand and Rentsch, 2016) may be beneficial to measure the explicit component during learning without priming subjects.

There is ongoing debate about the precise definition of the terms implicit and explicit when applied in a motor learning context (Kleynen et al., 2014). As the authors note implicit and explicit learning may not represent a dichotomy but instead ends of a continuum. The results of this experiment suggest that indeed a binary distinction may not be possible as successful participants here demonstrate awareness but mixed levels of verbalizable strategies, even when they are able to return to reaching at baseline angles on request. Distinction of these possibilities is further complicated by relying on questionnaires (Shanks and John, 1994). Moreover, responses are not always easy to classify into categories and some subjects hold their views in low-confidence. Here we define the explicit component to learning as the amount that participants could remove on request. Such a definition of explicit motor control (Mazzoni and Wexler, 2009) could be more akin to awareness (Werner et al., 2015) or a form of cognitive control (Cavanagh et al., 2009), rather than an explicit strategy which is often defined as a subject's ability to verbalize the strategy they have employed.

In order to investigate the mechanism through which subjects learnt to counter the rotation we employed the same analysis as Pekny et al., (2015). However, their study didn't involve learning as such, as the rotation was immediately washed out. Despite this, our results are remarkably similar, in that subjects in both studies made larger and more variable changes in actions following trials in which they made an error. Sidarta et al. (2016) have also described a similar pattern of behavior when subjects attempt to find a hidden target zone based on binary feedback, with greater reductions in error following incorrect trials. Our results indicate that subjects who were unable to learn the full rotation made smaller and less variable changes in response to errors and this was primarily driven by their sensitivity to the outcome of the previous trial. Learning from errors has been suggested to be a signature of explicit reinforcement learning, in contrast to learning from success in implicit learning (Loonis et al., 2017). Therefore, the finding that the difference between successful and unsuccessful subjects in the current experiments was in response to errors further supports the idea that it is the sensitivity of the explicit system that is important for this task. However, from the data presented here it is impossible to determine if the corrections following errors are explicit in nature or due to implicit motor variability (He et al., 2016; Wu et al., 2014). In future, similar experiments investigating the presence of neural signatures of explicit learning in tasks such as this may be able to shed light on which process underlie trial-by-trial changes (Loonis et al., 2017). Interestingly, the pattern of reduced sensitivity to errors found for unsuccessful subjects in the current experiment was similar to that described for parkinsonian patients (Pekny et al., 2015). Genetic variability in various aspects of the dopaminergic system has previously been linked to differential performance in reinforcement learning (Frank et al., 2007, 2009), and the balance of model-free and model-based decision-making systems (Doll et al., 2016). Future experiments assessing if the same genetic principles apply to motor learning based on reward may be useful in not only explaining the variation in response but also cementing the links between the principles of reinforcement learning and motor learning (Chen et al., 2017, 2018). Interestingly, the magnitude of changes made in response to errors in a binary feedback based motor learning task was correlated with connectivity changes between motor areas, prefrontal cortex and the intraparietal sulcus (Sidarta et al., 2016). The prefrontal cortex and intraparietal sulcus have been associated with the model-based decision making system (Gläscher et al., 2010), adding further evidence for a pivotal role of explicit systems in reward-based motor learning. However, it should be noted that effects of attention and motivation cannot be ruled out in the current paradigm. Therefore, accompanying neurophysiological measures of these variables may be useful in elucidating their possible contribution.

The efficacy of the dual task paradigm employed here in preventing learning is remarkable. Dual tasks have previously been employed in conjunction with motor adaptation to visuomotor rotations (Galea et al., 2010), force-fields (Keisler and Shadmehr, 2010; Taylor and Thoroughman, 2007, 2008), as well as during the learning of motor skills (Maxwell et al., 2001) and sequence learning (Brown and Robertson, 2007). Galea et al. (2010) demonstrated that a secondary task can slow the rate of adaptation to both a gradually and abruptly introduced visuomotor rotation. Keisler and Shadmehr (2010) found that a declarative memory task could interfere with the ‘fast’ adaptation system but that a demanding cognitive task without the memory component did not. Furthermore, inhibition of the ‘fast’ process led to an increase in the ‘slow’, non-declarative process. Similarly in a sequence learning task a dual task with a declarative element increased the procedural learning suggesting that these two aspects of learning may be in competition (Brown and Robertson, 2007). It could therefore be hypothesized that the use of a dual task in the current paradigm would shift learning from explicit to the implicit system. However, the current data suggest that this did not occur and for this paradigm the explicit system is necessary to compensate for large rotations, and cannot be substituted for by an increase in the use of the implicit learning system. Alternatively, if the implicit system is not engaged by the nature of this task then it would be impossible for it to compensate for the disruption of the explicit system. Arguing against this possibility is the fact that implicit retention was observed in this paradigm, suggesting that the implicit system is indeed engaged, at least to some degree. Whereas previous experiments have employed secondary tasks that involve more verbal systems (Galea et al., 2010; Keisler and Shadmehr, 2010; Taylor and Thoroughman, 2007), we selected the dual task which would have the maximum likelihood of disrupting the explicit system (Anguera et al., 2009; Georgopoulos and Massey, 1987). As the difficulty of the secondary task has been linked with the amount of disruption (Taylor and Thoroughman, 2008), it is also possible that the specific nature of the task may also be important and this is an interesting area for future study. One other possibility is that constant impairment of performance due to the secondary task may reduce intrinsic motivation of subjects (Liao and Masters, 2001).

The distinction between implicit and explicit reinforcement systems engaging in learning motor tasks is not merely academic. At least part of the increased interest in the addition of reward to motor adaptation and learning is due to the finding that it increases retention (Abe et al., 2011; Dayan et al., 2014, 2014; Galea et al., 2015; Shmuelof et al., 2012; Therrien et al., 2016), along with the promise this may have in a rehabilitation setting (Goodman et al., 2014; Quattrocchi et al., 2017). However, if the benefits are primarily due to explicit or strategic processes, they may be poorly transferred to other environments and be susceptible to disruption. In line with this, it has been demonstrated that motor skills, such as golf putting or playing table tennis, are less disrupted by manipulations such as dividing cognitive load, reducing reaction times or performing in stressful situations when learnt implicitly (Liao and Masters, 2001; Maxwell et al., 2001). If the final goal of the addition of reward to motor learning tasks is to increase retention for practical rehabilitation then it may be that methods that increase the implicit contribution are required such as employing learning by analogy, reducing errors during learning or the addition of dual tasks (Liao and Masters, 2001). However, the choice and difficulty of the dual task should be made with caution as from the data presented here it may be too disruptive and ultimately prevent learning.

Acknowledgements

The authors would like to thank Dr. Xiuli Chen for advice concerning the analysis of the data.

Grants

PH, OC, and JM were supported by the European Research Council grant MotMotLearn 637488.

Disclosures

The authors declare no competing financial interests.

References

Abe, M., Schambra, H., Wassermann, E.M., Luckenbaugh, D., Schweighofer, N., and Cohen, L.G. (2011). Reward Improves Long-Term Retention of a Motor Memory through Induction of Offline Memory Gains. Curr. Biol. 21, 557–562.

Anguera, J.A., Reuter-Lorenz, P.A., Willingham, D.T., and Seidler, R.D. (2009). Contributions of Spatial Working Memory to Visuomotor Learning. J. Cogn. Neurosci. 22, 1917–1930.

Anguera, J.A., Bernard, J.A., Jaeggi, S.M., Buschkuehl, M., Benson, B.L., Jennett, S., Humfleet, J., Reuter-Lorenz, P.A., Jonides, J., and Seidler, R.D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. Behav. Brain Res. 228, 107–115.

Bond, K.M., and Taylor, J.A. (2015). Flexible explicit but rigid implicit learning in a visuomotor adaptation task. J. Neurophysiol. 113, 3836–3849.

Bond, K.M., and Taylor, J.A. (2017). Structural Learning in a Visuomotor Adaptation Task Is Explicitly Accessible. ENeuro 4.

Brown, R.M., and Robertson, E.M. (2007). Inducing motor skill improvements with a declarative task. Nat. Neurosci. 10, nn1836.

Cavanagh, J.F., Cohen, M.X., and Allen, J.J.B. (2009). Prelude to and Resolution of an Error: EEG Phase Synchrony Reveals Cognitive Control Dynamics during Action Monitoring. J. Neurosci. 29, 98–105.

Chen, X., Mohr, K., and Galea, J.M. (2017). Predicting explorative motor learning using decision-making and motor noise. PLOS Comput. Biol. 13, e1005503.

Chen, X., Holland, P., and Galea, J.M. (2018). The effects of reward and punishment on motor skill learning. Curr. Opin. Behav. Sci. 20, 83–88.

Christou, A.I., Miall, R.C., McNab, F., and Galea, J.M. (2016). Individual differences in explicit and implicit visuomotor learning and working memory capacity. Sci. Rep. 6.

Dayan, E., Averbeck, B.B., Richmond, B.J., and Cohen, L.G. (2014). Stochastic reinforcement benefits skill acquisition. Learn. Mem. 21, 140–142.

Diedrichsen, J., White, O., Newman, D., and Lally, N. (2010). Use-Dependent and Error-Based Learning of Motor Behaviors. J. Neurosci. 30, 5159–5166.

Doll, B.B., Bath, K.G., Daw, N.D., and Frank, M.J. (2016). Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. J. Neurosci. 36, 1211–1222.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., and Hutchison, K.E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc. Natl. Acad. Sci. U. S. A. 104, 16311–16316.

Frank, M.J., Doll, B.B., Oas-Terpstra, J., and Moreno, F. (2009). The neurogenetics of exploration and exploitation: Prefrontal and striatal dopaminergic components. Nat. Neurosci. 12, 1062–1068.

Galea, J.M., Sami, S.A., Albert, N.B., and Miall, R.C. (2010). Secondary tasks impair adaptation to step- and gradual-visual displacements. Exp. Brain Res. 202, 473–484.

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

Georgopoulos, A.P., and Massey, J.T. (1987). Cognitive spatial-motor processes. Exp. Brain Res. 65, 361–370.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. Neuron 66, 585–595.

Goodman, R.N., Rietschel, J.C., Roy, A., Jung, B.C., Diaz, J., Macko, R.F., and Forrester, L.W. (2014). Increased reward in ankle robotics training enhances motor control and cortical efficiency in stroke. J. Rehabil. Res. Dev. 51, 213–228.

Haith, A.M., and Krakauer, J.W. (2013). Model-Based and Model-Free Mechanisms of Human Motor Learning. Adv. Exp. Med. Biol. 782, 1–21.

Haith, A.M., Huberdeau, D.M., and Krakauer, J.W. (2015). The Influence of Movement Preparation Time on the Expression of Visuomotor Learning and Savings. J. Neurosci. 35, 5109–5117.

He, K., Liang, Y., Abdollahi, F., Bittmann, M.F., Kording, K., and Wei, K. (2016). The Statistical Determinants of the Speed of Motor Learning. PLOS Comput Biol 12, e1005023.

Huang, J.-J., Yen, C.-T., Tsao, H.-W., Tsai, M.-L., and Huang, C. (2014). Neuronal Oscillations in Golgi Cells and Purkinje Cells are Accompanied by Decreases in Shannon Information Entropy. The Cerebellum 13, 97–108.

Huberdeau, D.M., Krakauer, J.W., and Haith, A.M. (2015). Dual-process decomposition in human sensorimotor adaptation. Curr. Opin. Neurobiol. 33, 71–77.

Izawa, J., and Shadmehr, R. (2011). Learning from Sensory and Reward Prediction Errors during Motor Adaptation. PLOS Comput Biol 7, e1002012.

Izawa, J., Criscimagna-Hemminger, S.E., and Shadmehr, R. (2012). Cerebellar Contributions to Reach Adaptation and Learning Sensory Consequences of Action. J. Neurosci. 32, 4230–4239.

Keisler, A., and Shadmehr, R. (2010). A Shared Resource between Declarative Memory and Motor Memory. J. Neurosci. 30, 14817–14823.

Kleynen, M., Braun, S.M., Bleijlevens, M.H., Lexis, M.A., Rasquin, S.M., Halfens, J., Wilson, M.R., Beurskens, A.J., and Masters, R.S.W. (2014). Using a Delphi Technique to Seek Consensus Regarding Definitions, Descriptions and Classification of Terms Related to Implicit and Explicit Forms of Motor Learning. PLOS ONE 9, e100227.

van der Kooij, K., and Overvliet, K.E. (2016). Rewarding imperfect motor performance reduces adaptive changes. Exp. Brain Res. 234, 1441–1450.

Leow, L.-A., Gunn, R., Marinovic, W., and Carroll, T.J. (2017). Estimating the implicit component of visuomotor rotation learning by constraining movement preparation time. J. Neurophysiol. jn.00834.2016.

Liao, C.-M., and Masters, R.S.W. (2001). Analogy learning: A means to implicit motor learning. J. Sports Sci. 19, 307–319.

Loonis, R.F., Brincat, S.L., Antzoulatos, E.G., and Miller, E.K. (2017). A Meta-Analysis Suggests Different Neural Correlates for Implicit and Explicit Learning. Neuron 96, 521–534.e7.

Manley, H., Dayan, P., and Diedrichsen, J. (2014). When Money Is Not Enough: Awareness, Success, and Variability in Motor Learning. PLOS ONE 9, e86580.

Maxwell, J.P., Masters, R.S.W., Kerr, E., and Weedon, E. (2001). The implicit benefit of learning without errors. Q. J. Exp. Psychol. Sect. A 54, 1049–1068.

Mazzoni, P., and Wexler, N.S. (2009). Parallel Explicit and Implicit Control of Reaching. PLOS ONE 4, e7557.

McDougle, S.D., Bond, K.M., and Taylor, J.A. (2015). Explicit and Implicit Processes Constitute the Fast and Slow Processes of Sensorimotor Learning. J. Neurosci. 35, 9568–9579.

Morehead, J.R., Taylor, J.A., Parvin, D.E., and Ivry, R.B. (2017). Characteristics of Implicit Sensorimotor Adaptation Revealed by Task-irrelevant Clamped Feedback. J. Cogn. Neurosci. 29, 1061–1074.

Nikooyan, A.A., and Ahmed, A.A. (2015). Reward feedback accelerates motor learning. J. Neurophysiol. 113, 633–646.

Pekny, S.E., Izawa, J., and Shadmehr, R. (2015). Reward-Dependent Modulation of Movement Variability. J. Neurosci. 35, 4015–4024.

Peters, M., and Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. Brain Cogn. 66, 260–264.

Quattrocchi, G., Greenwood, R., Rothwell, J.C., Galea, J.M., and Bestmann, S. (2017). Reward and punishment enhance motor adaptation in stroke. BioRxiv 106377.

Rand, M.K., and Rentsch, S. (2016). Eye-Hand Coordination during Visuomotor Adaptation with Different Rotation Angles: Effects of Terminal Visual Feedback. PLOS ONE 11, e0164602.

Saijo, N., and Gomi, H. (2010). Multiple Motor Learning Strategies in Visuomotor Rotation. PLOS ONE 5, e9399.

Shanks, D.R., and John, M.F.S. (1994). Characteristics of dissociable human learning systems. Behav. Brain Sci. 17, 367–395.

Shepard, R.N., and Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. Science 171, 701–703.

Shmuelof, L., Huang, V.S., Haith, A.M., Delnicki, R.J., Mazzoni, P., and Krakauer, J.W. (2012). Overcoming Motor "Forgetting" Through Reinforcement Of Learned Actions. J. Neurosci. 32, 14617–14621.

Sidarta, A., Vahdat, S., Bernardi, N.F., and Ostry, D.J. (2016). Somatic and Reinforcement-Based Plasticity in the Initial Stages of Human Motor Learning. J. Neurosci. 36, 11682–11692.

Smith, M.A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting Adaptive Processes with Different Timescales Underlie Short-Term Motor Learning. PLOS Biol. 4, e179.

Taylor, J.A., and Thoroughman, K.A. (2007). Divided Attention Impairs Human Motor Adaptation But Not Feedback Control. J. Neurophysiol. 98, 317–326.

Taylor, J.A., and Thoroughman, K.A. (2008). Motor Adaptation Scaled by the Difficulty of a Secondary Cognitive Task. PLoS ONE 3.

Taylor, J.A., Krakauer, J.W., and Ivry, R.B. (2014). Explicit and Implicit Contributions to Learning in a Sensorimotor Adaptation Task. J. Neurosci. 34, 3023–3032.

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

Tseng, Y., Diedrichsen, J., Krakauer, J.W., Shadmehr, R., and Bastian, A.J. (2007). Sensory Prediction Errors Drive Cerebellum-Dependent Adaptation of Reaching. J. Neurophysiol. 98, 54–62.

Werner, S., Aken, B.C. van, Hulst, T., Frens, M.A., Geest, J.N. van der, Strüder, H.K., and Donchin, O. (2015). Awareness of Sensorimotor Adaptation to Visual Rotations of Different Size. PLOS ONE 10, e0123321.

Wu, H.G., Miyamoto, Y.R., Castro, L.N.G., Ölveczky, B.P., and Smith, M.A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nat. Neurosci. 17, 312–321.

Figure Captions

***Figure 1. Experimental design**. **A,** Subjects held the handle of robotic manipulandum with their right hand, the position of the arm and handle was hidden from sight and feedback was provided on a horizontal screen. **B,** Subjects made 'shooting' movements from a starting position (green circle) towards a target (red circle), after the initial practice trials the position of the cursor (white circle) was no longer visible at any point. **C,** Successful trials were indicated to the subject with the display of a green tick after the cursor had passed through a region centered on the target, over the course of the paradigm the position of the reward region gradually moved (solid green circle to dashed green circle) whilst the visible target (red circle) remained in the central location. By the end of the learning period a successful reach (dotted white line) was rotated by a maximum of either 15° or 25°. **D,** Time-course of Experiment 2, at the same time as the target appeared on screen a 'shape' was also displayed slightly above it, the subject was asked to memorize this shape. After the reach was completed and the hand returned to the starting position subjects used their left hand to respond with a button press as to whether they believed the new shape shown on screen was a rotated version of the shape or an entirely different shape.*

***Figure 2. Experiment 1: group performance.** **A**, Reach angle averaged over blocks of 5 trials, solid colored lines represent the mean of each group and the shaded region represents SEM. The average behavior of subjects in the 15Rot paradigm (Orange) fell consistently within the rewarded region (grey shaded region) indicating successful learning. **B**, Average reach angle over blocks for all subjects in the 25Rot paradigm (magenta) and also the same subjects split into two groups based on success at the final angle (25RotSuccess – red, 25RotFail – blue). **C**, Distribution plots displaying the reach angles for subjects in the three groups at various timepoints throughout the experiment with individual data points overlaid on an estimate of the distribution. Horizontal black line in the distribution represents the group median**.** **D**, Distribution plots of the computed variables of Implicit ('Remove-Baseline') and Explicit ('Maintain-Implicit') retention. Significance stars above horizontal black bars indicate differences between the groups (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Significance stars below the distributions represent a significant difference from zero. **E,** Reach angle averaged over blocks of 5 trials for subjects in the 25RotControl group. There was no reduction in*

reach angle during the time taken for the control questions between Maintain 1 and Maintain 2 blocks. However, when subjects were subsequently asked to remove their strategy, the period between Maintain 2 and Remove blocks, a significant reduction in reach angle was observed.

Figure 3. Experiment 1: trial-by-trial behavior. *Example of trial by trial reach angles from a subject who was successful at the final angle (**A**) and one who was unsuccessful (**B**). In each case rewarded trials are indicated with a circular marker and non-rewarded trials with a 'x'. The grey shaded region indicates the reward region. **C,** Failure points for subjects in the 25RotFail group, thick lines are the mean reach angle for each subject at each rotation angle, thin lines represent mean of each block (average of 5 trials), colors go from hot to cold matching failure angles ranging from high to low. Vertical lines represent the last angle at which mean reach fell within rewarded region for each subject. The mean and standard deviation of all angles of failure is displayed as text.*

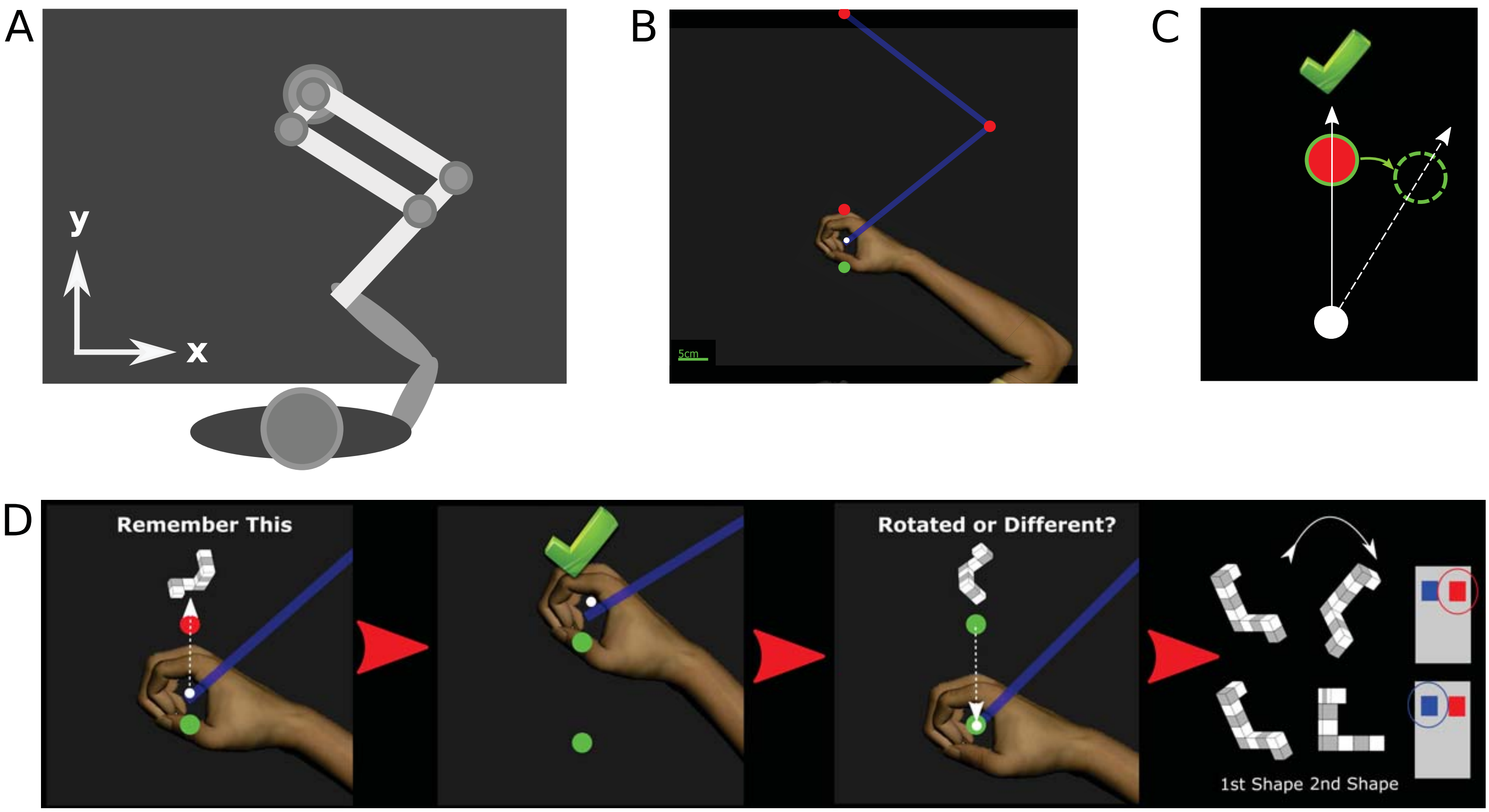
Figure 4. Experiment 1: performance after correct and incorrect trials. *Analysis of the effects of the success of the previous trial and reward history on trial by trial changes in reach angle for the three groups in Experiment 1 (15Rot – Orange, 25RotSuccess – Red, 25RotFail – Blue). Median (**A**) and MAD (**B**) of change in reach angle separated by the success of the previous trial. Median (**C**) and MAD (**D**) of the absolute change in reach angle separated by the success of the previous trial. **E**, The absolute change in reach angle following all combinations of trial success over the previous three trials. **F**, Sensitivity to the outcomes of each of the previous trials. Significance stars above horizontal black bars indicate differences between the groups (* $P < 0.05$, ** $P < 0.01$).*

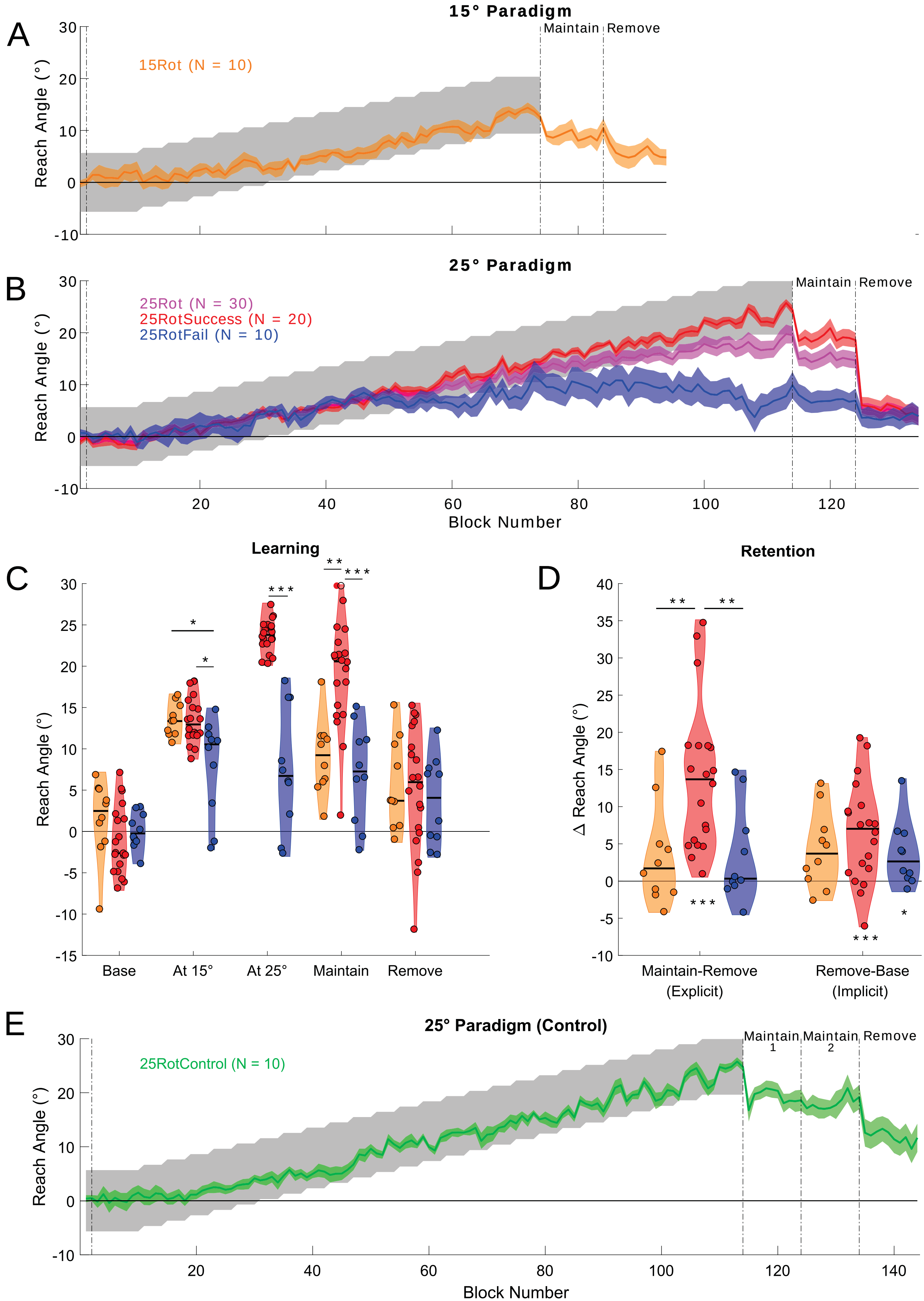
Figure 5. Experiment 2: group performance. *Change in reach angle over blocks (average of 5 trials) during the dual task experiment. **A**, Group performance for the DualTask (Green) and Control (Purple) task groups, the line indicates the mean and shaded region the SEM. The grey shaded region represents the reward region. **B**, the split of the control task group into ControlSuccess (Dark Red) and ControlFail (Blue). **C**, Distribution plots displaying the performance at different time points for the dual task, and split control groups. The shaded region represents an estimation of the distribution*

and is overlaid with data for each individual subject. ***D****, Distribution plots of the difference in reach angle during retention phases indicating the implicit and explicit components of retention. Significance stars above horizontal black bars indicate differences between the groups (* P < 0.05, ** P < 0.01).*

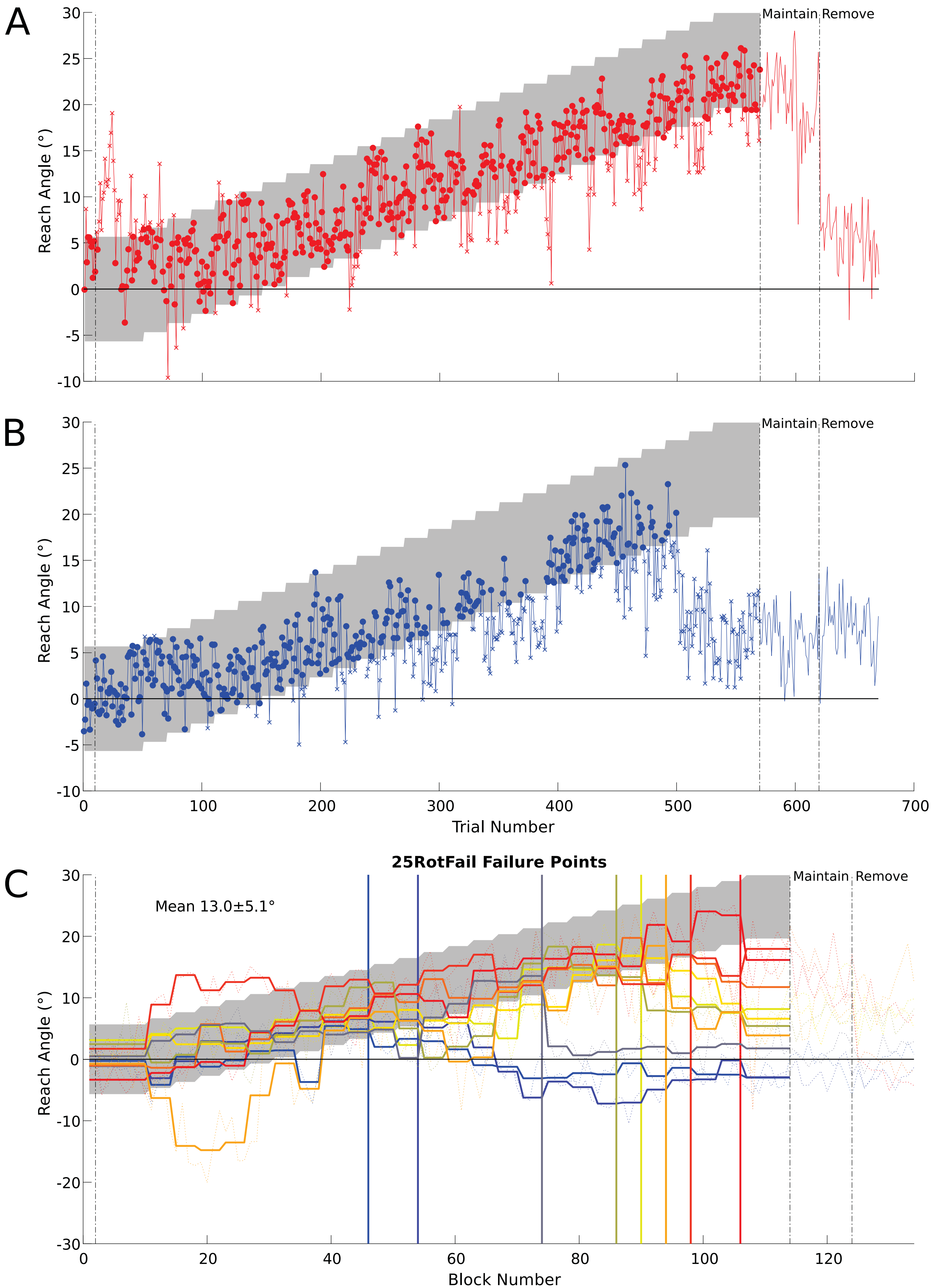
Figure 6. Experiment 2: trial-by-trial behavior. *Example of trial by trial reach angles from a subject performing the dual task (****A****) rewarded trials are indicated with a circular marker and non-rewarded trials with a 'x'. The grey shaded region represents the reward region.* ***B,*** *Failure points for subjects in the DualTask group, thick lines are the mean reach angle for each subject at each rotation angle, thin lines represent mean of each block, colors go from hot to cold matching failure angles ranging from high to low. Vertical lines represent the last angle at which mean reach fell within rewarded region for each subject. Th mean and standard deviation of the angle of failure is reported as text in the figure.*

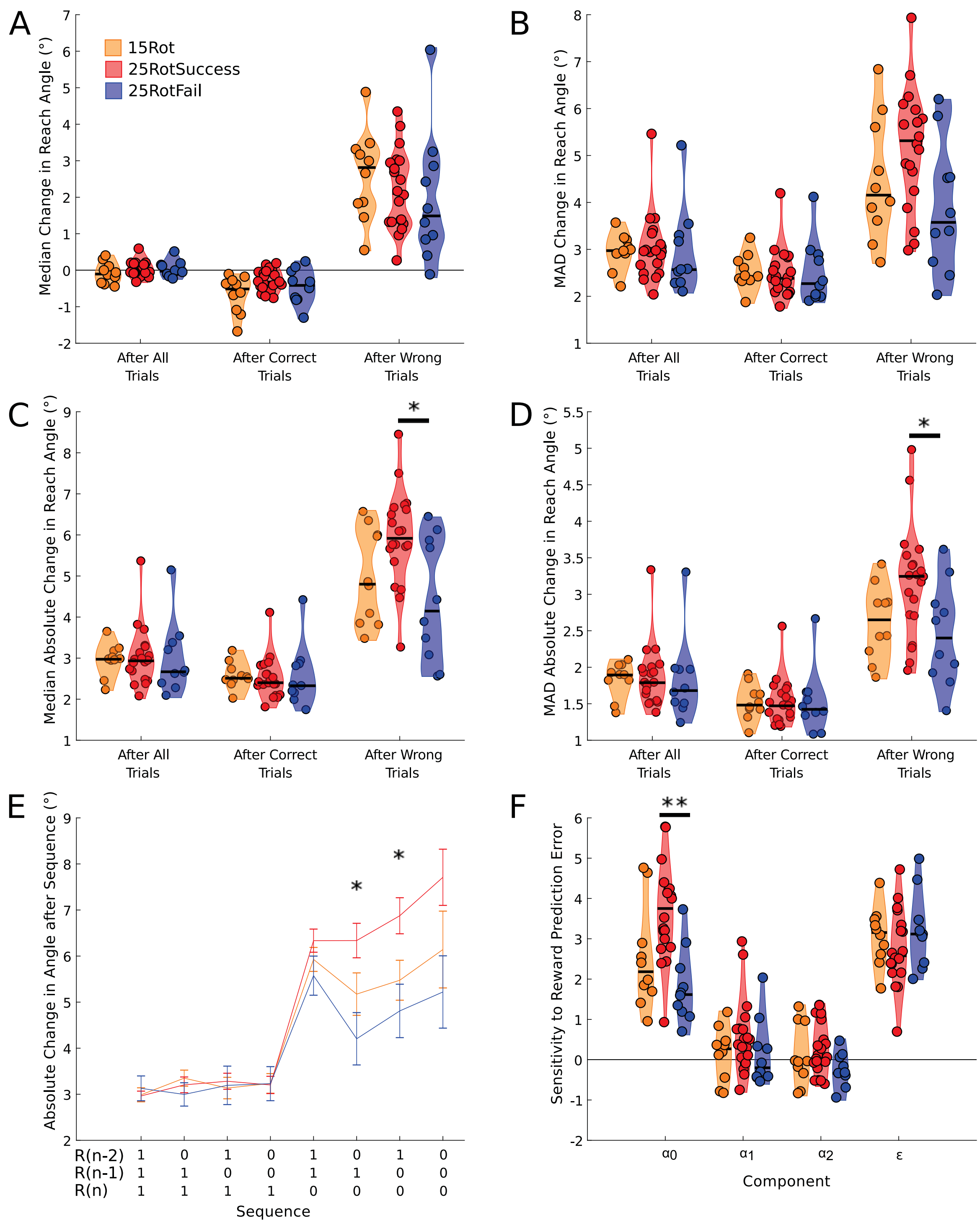
Figure 7. Experiment 2: performance after correct and incorrect trials*. Analysis of the effects of the success of the previous trial and reward history on trial by trial changes in reach angle for the two groups performing the control task in Experiment 2. Distribution plots for median (****A****) and MAD (****B****) of change in reach angle separated by the success of the previous trial. Median (****C****) and MAD (****D****) of the absolute change in reach angle separated by the success of the previous trial.* ***E****, the absolute change in reach angle following all combinations of trial success over the previous three trials.* ***F****, sensitivity to the outcomes of each of the previous trials.*

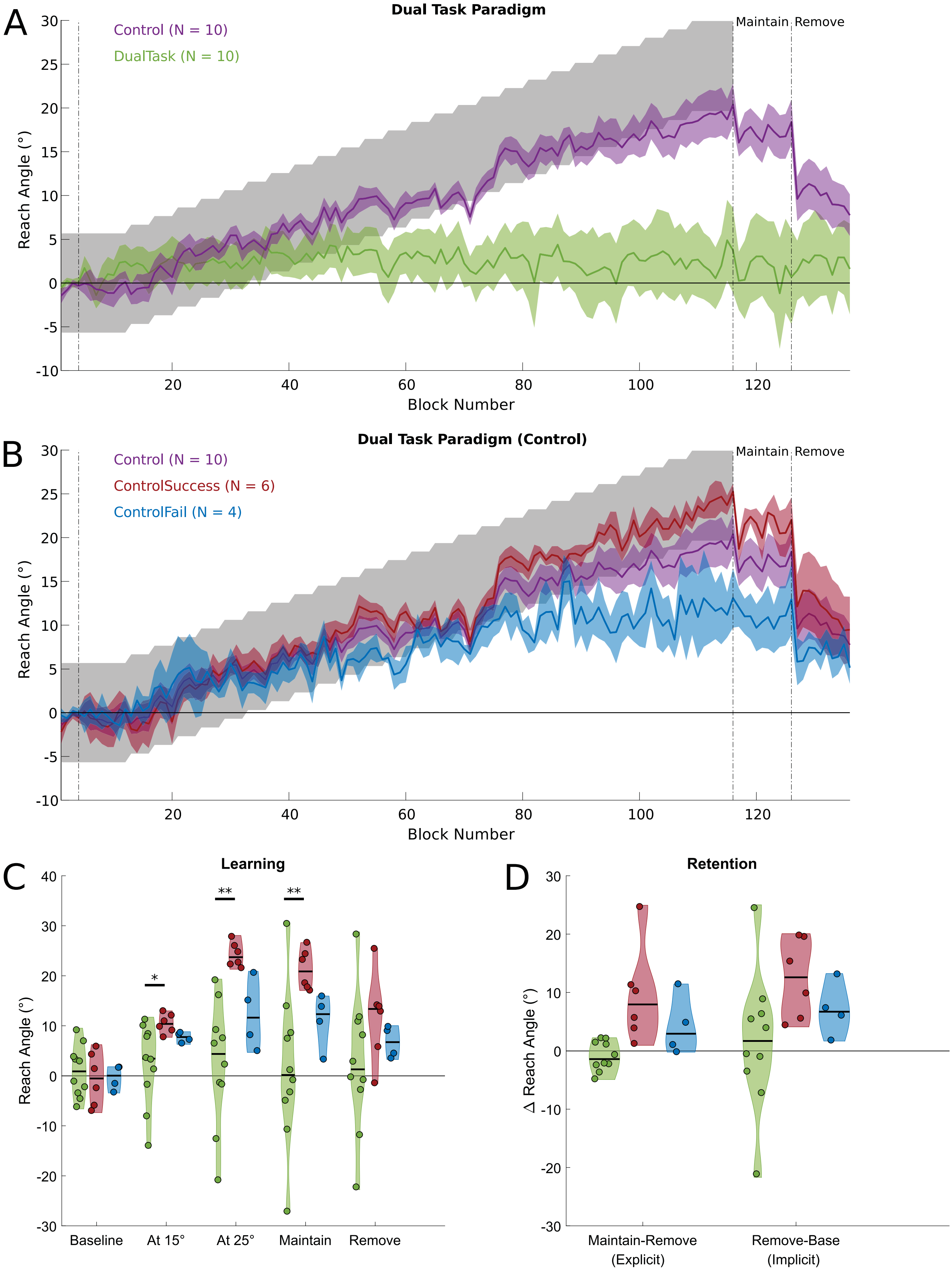

A
y
x
B
5cm
C
D
Remember This
Rotated or Different?
1st Shape 2nd Shape

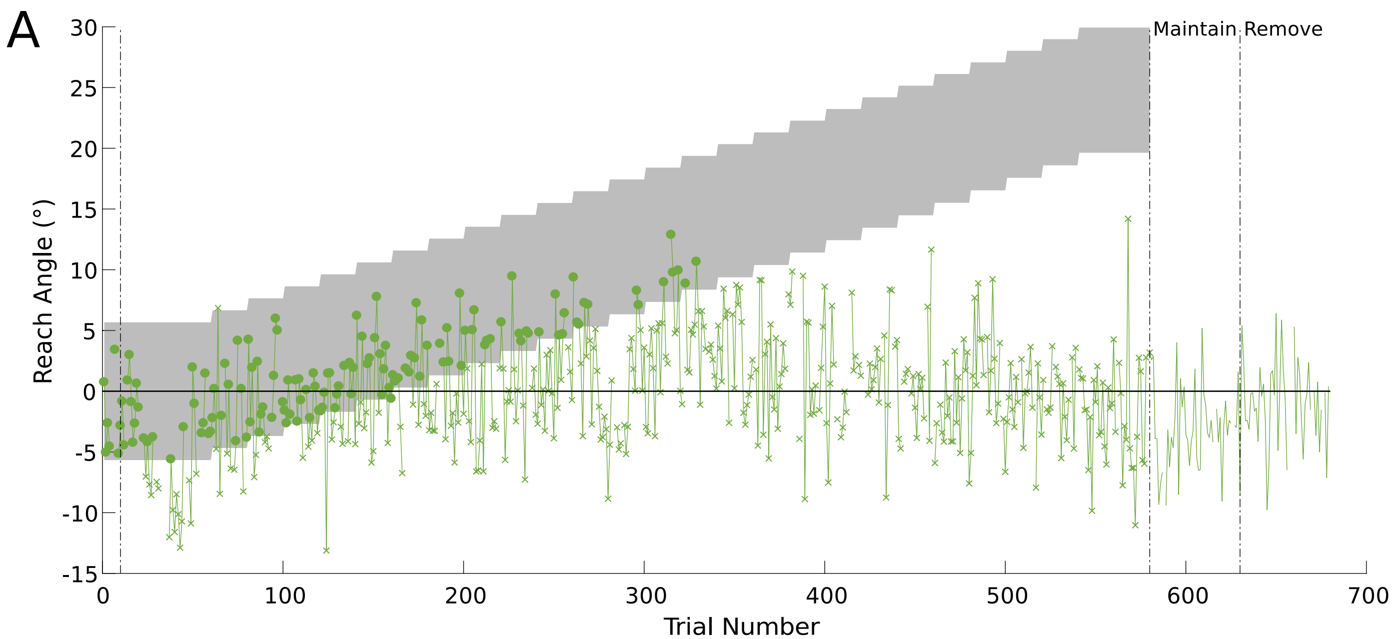


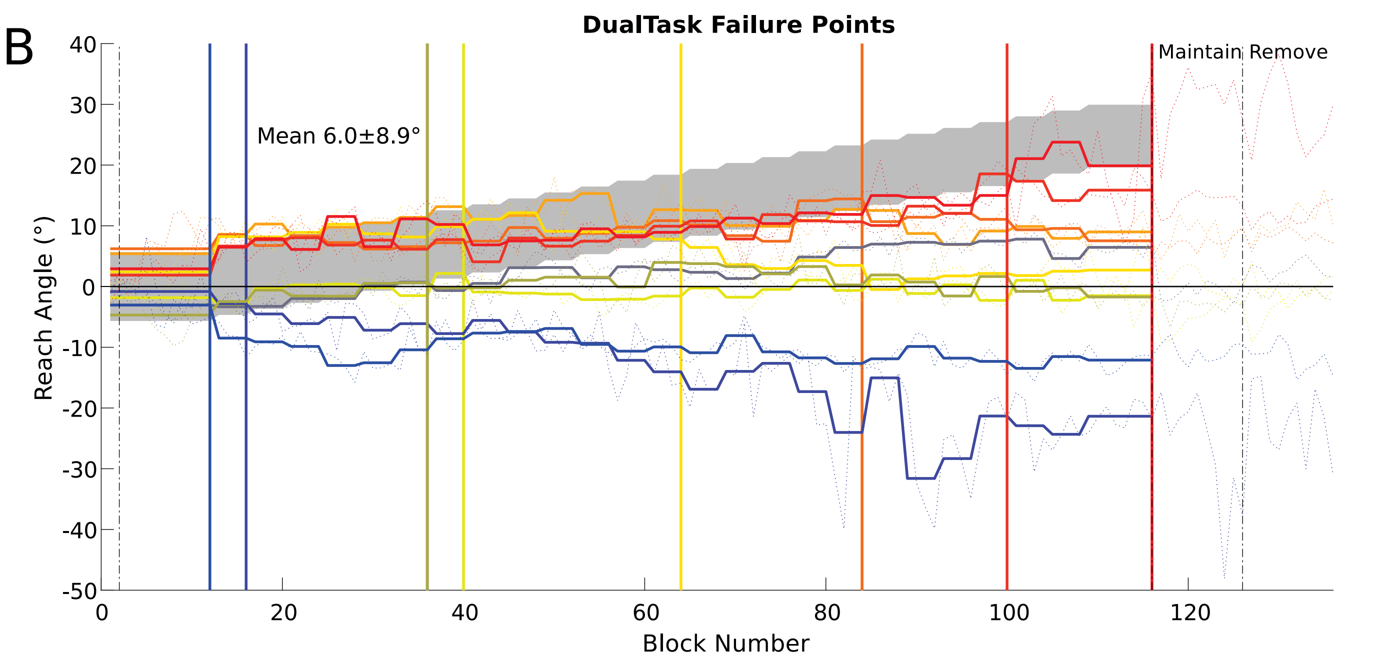
A
15° Paradigm
Maintain
Remove
15Rot (N = 10)
Reach Angle (°)
B
25° Paradigm
25Rot (N = 30)
25RotSuccess (N = 20)
25RotFail (N = 10)
Block Number
C
Learning
Base
At 15°
At 25°
Maintain
Remove
D
Retention
Δ Reach Angle (°)
Maintain-Remove
(Explicit)
Remove-Base
(Implicit)
E
25° Paradigm (Control)
25RotControl (N = 10)
Maintain 1
Maintain 2

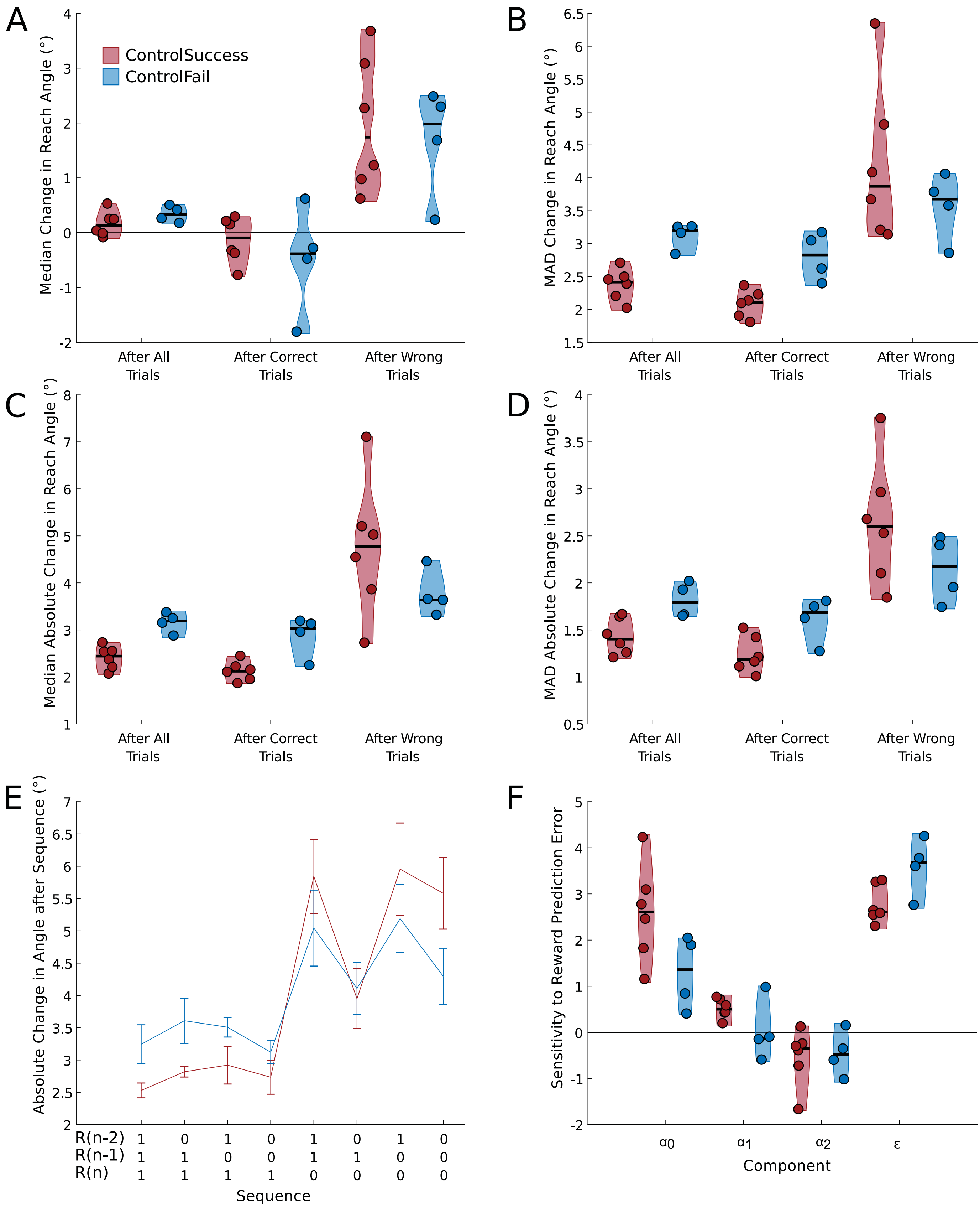

A
B
C
Maintain
Remove
Reach Angle (°)
Trial Number
25RotFail Failure Points
Mean 13.0±5.1°
Block Number
30
25
20
15
10
5
0
-5
-10
-20
-30
100
200
300
400
500
600
700
40
60
80
120


A
B
C
D
E
F
15Rot
25RotSuccess
25RotFail
Median Change in Reach Angle (°)
MAD Change in Reach Angle (°)
Median Absolute Change in Reach Angle (°)
MAD Absolute Change in Reach Angle (°)
Absolute Change in Angle after Sequence (°)
Sensitivity to Reward Prediction Error
After All Trials
After Correct Trials
After Wrong Trials
R(n-2) 1 0 1 0 1 0 1 0
R(n-1) 1 1 0 0 1 1 0 0
R(n) 1 1 1 1 0 0 0 0
Sequence
α0
α1
α2
ε
Component


A
Dual Task Paradigm
Control (N = 10)
DualTask (N = 10)
Maintain
Remove
Reach Angle (°)
Block Number
B
Dual Task Paradigm (Control)
Control (N = 10)
ControlSuccess (N = 6)
ControlFail (N = 4)
Maintain
Remove
Reach Angle (°)
Block Number
C
Learning
Reach Angle (°)
Baseline
At 15°
At 25°
Maintain
Remove
D
Retention
Δ Reach Angle (°)
Maintain-Remove
(Explicit)
Remove-Base
(Implicit)


A
Maintain
Remove
Reach Angle (°)
Trial Number
30
25
20
15
10
5
0
-5
-10
-15
0
100
200
300
400
500
600
700


B
DualTask Failure Points
Maintain
Remove
Mean 6.0±8.9°
Reach Angle (°)
Block Number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120


A
ControlSuccess
ControlFail
Median Change in Reach Angle (°)
After All Trials
After Correct Trials
After Wrong Trials
B
MAD Change in Reach Angle (°)
C
Median Absolute Change in Reach Angle (°)
D
MAD Absolute Change in Reach Angle (°)
E
Absolute Change in Angle after Sequence (°)
R(n-2) 1 0 1 0 1 0 1 0
R(n-1) 1 1 0 0 1 1 0 0
R(n) 1 1 1 1 0 0 0 0
Sequence
F
Sensitivity to Reward Prediction Error
α0
α1
α2
ε
Component