Domain-specific working memory, but not dopamine-related genetic variability, shapes reward-based motor learning

Running title: Mechanisms of reward-based motor learning

Peter Holland*[1], Olivier Codol*[1], Elizabeth Oxley[1], Madison Taylor[1], Elizabeth Hamshere[1], Shadiq Joseph[1], Laura Huffer[1], Joseph M. Galea[1]

[1]*School of Psychology and Centre for Human Brain Health, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*

* = These authors provided equal contribution to this work

Number of Figures: 11

Number of Tables: 2

Number of Pages: 47

Words: Abstract = 249 Introduction = 650 Discussion = 1494

Keywords: Explicit control, Motor adaptation, Reinforcement, Retention

Conflict of Interest: The authors declare no competing financial interests

Acknowledgements: This work was supported by the European Research Council starting grant: MotMotLearn (637488)

Correspondence should be addressed to:

Peter Holland

Email: P.J.Holland@bham.ac.uk

Abstract

The addition of rewarding feedback to motor learning tasks has been shown to increase the retention of learning, spurring interest in its possible utility for rehabilitation. However, motor tasks employing rewarding feedback have repeatedly been shown to lead to great inter-individual variability in performance. Understanding the causes of such variability is vital for maximising the potential benefits of reward-based motor learning. Thus, using a large human cohort of both sexes (n=241), we examined whether spatial (SWM), verbal (VWM) and mental rotation (RWM) working memory capacity and dopamine-related genetic profiles were associated with performance in two reward-based motor tasks. The first task assessed participant's ability to follow a slowly shifting reward region based on hit/miss (binary) feedback. The second task investigated participant's capacity to preserve performance with binary feedback after adapting to the rotation with full visual feedback. Our results demonstrate that higher SWM is associated with greater success and an enhanced capacity to reproduce a successful motor action, measured as change in reach angle following reward. In contrast, higher RWM was predictive of an increased propensity to express an explicit strategy when required to make large reach angle adjustments. Therefore, SWM and RWM were reliable but dissociable predictors of success during reward-based motor learning. Change in reach direction following failure was also a strong predictor of success rate, although we observed no consistent relationship with working memory. Surprisingly, no dopamine-related genotypes predicted performance. Therefore, working memory capacity plays a pivotal role in determining individual ability in reward-based motor learning.

Significance statement

Reward-based motor learning tasks have repeatedly been shown to lead to idiosyncratic behaviours that cause varying degrees of task success. Yet, the factors determining an

individual's capacity to use reward-based feedback are unclear. Here, we assessed a wide range of possible candidate predictors, and demonstrate that domain-specific working memory plays an essential role in determining individual capacity to use reward-based feedback. Surprisingly, genetic variations in dopamine availability were not found to play a role. This is in stark contrast with seminal work in the reinforcement and decision-making literature, which show strong and replicated effects of the same dopaminergic genes in decision-making. Therefore, our results provide novel insights into reward-based motor learning, highlighting a key role for domain-specific working memory capacity.

Introduction

When performing motor tasks under altered environmental conditions, adaptation to the new constraints occurs through the recruitment of a variety of systems (Taylor and Ivry, 2014). Arguably the most studied of those systems is cerebellum-dependent adaptation, which consists of the implicit and automatic recalibration of mappings between actual and expected outcomes through sensory prediction errors (Morehead et al., 2017; Tseng et al., 2007). Besides cerebellar adaptation, other work has demonstrated the involvement of a cognitive, deliberative process whereby motor plans are adjusted based on structural understanding of the task (Bond and Taylor, 2015; Taylor and Ivry, 2011). We label this process 'explicit control' (Codol et al., 2018; Holland et al., 2018), although it has also been referred to as strategy (Taylor and Ivry, 2011) or explicit re-aiming (Morehead et al., 2015). Recently it has been proposed that reinforcement learning, whereby the memory of successful or unsuccessful actions are strengthened or weakened, respectively, may also play a role (Huang et al., 2011; Izawa and Shadmehr, 2011; Shmuelof et al., 2012). Such reward-based reinforcement has been assumed to be an implicit and automatic process (Haith and Krakauer, 2013). However, recent evidence suggests that phenomena attributed to reinforcement-based learning during visuomotor rotation

tasks can largely be explained through explicit processes (Codol et al., 2018; Holland et al., 2018).

One outstanding feature of reinforcement-based motor learning is the great variability expressed across individuals (Codol et al., 2018; Holland et al., 2018; Therrien et al., 2016, 2018). What factors underlie such variability is unclear. If reinforcement is explicitly grounded, it could be argued that individual working memory capacity (WMC), which is reliably related to the propensity to employ explicit control in classical motor adaptation tasks (Anguera et al., 2010, 2012; Christou et al., 2016; Holland et al., 2018; Sidarta et al., 2018), would also predict performance in reinforcement-based motor learning. Anguera et al. (2010) demonstrated that mental rotation WMC (RWM), unlike other forms of working memory such as verbal WMC (VWM), correlates with explicit control. Recently, Christou et al. (2016) reported similar results with spatial WMC (SWM). If this extends to reward based motor learning, this would strengthen the proposal that it bears a strong explicit component.

Another potential contributor to this variability is genetic profile. In previous work (Codol et al., 2018; Holland et al., 2018), we argue that reinforcement-based motor learning performance relies on a balance between exploration and exploitation of the task space, a feature reminiscent of structural learning and reinforcement-based decision-making (Daw et al., 2005; Frank et al., 2009; Sutton and Barto, 1998). A series of studies from Frank and colleagues suggests that individual tendencies to express explorative/exploitative behaviour can be predicted based on dopamine-related genetic profile (Doll et al., 2016; Frank et al., 2007, 2009). Reinforcement has consistently been linked to dopaminergic function in a variety of paradigms, and thus, such a relationship could also be expected in reward-based motor learning (Pekny et al., 2015). Specifically, Frank and colleagues focused on Catecholamine-O-Methyl-Transferase (COMT), Dopamine- and cAMP-Regulated neuronal Phosphoprotein (DARPP32) and Dopamine

Receptor D2 (DRD2), and suggest a distinction between COMT-modulated exploration and DARPP32- and DRD2-modulated exploitation (Frank et al., 2009).

Consequently, we investigated the influence of WMC (RWM, SWM, and VWM) and genetic variations in dopamine metabolism (DRD2, DARP32, and COMT) on individuals' ability to perform reward-based motor learning. We examined this using two established reward-based motor learning tasks. First, a task analogous to a gradually introduced rotation (Holland et al., 2018) required participants to learn to adjust the angle at which they reached to a slowly and secretly shifting reward region (Acquire); second, a task with an abruptly introduced rotation (Codol et al., 2018; Shmuelof et al., 2012) required participants to preserve performance with reward-based feedback after adapting to a visuomotor rotation (Preserve). The use of these two tasks enabled us to examine whether similar predictors of performance explained participant's capacity to acquire and preserve behaviour with reward-based feedback.

Methods

Prior to the start of data collection, the sample size, variables of interest and analysis method were pre-registered. The pre-registered information, data and analysis code can be found online at https://osf.io/j5v2s/ and https://osf.io/rmwc2/ for the Preserve and Acquire tasks, respectively.

Participants

121 (30 male, mean age: 21.06, range: 18-32) and 120 (16 male, mean age: 19.24, range: 18-32) participants were recruited for the Acquire and Preserve tasks, respectively. All participants provided informed consent and were remunerated with either course credit or money (£7.50/hour). All participants were free of psychological, cognitive, motor or uncorrected

visual impairment. The study was approved by and performed in accordance with the local research ethics committee of the University of Birmingham, UK.

Experimental design

Participants were seated before a horizontally fixed mirror reflecting a screen placed above, on which visual stimuli were presented. This arrangement resulted in the stimuli appearing at the level on which participants performed their reaching movements. The Acquire (gradual) and Preserve (abrupt) tasks were performed on two different stations, with a KINARM (BKIN Technology, London, Ontario; sampling rate: 1000Hz) and a Polhemus 3SPACE Fastrak tracking device (Colchester, Vermont; sampling rate: 120Hz), employed respectively. The Acquire task was run using Simulink (The Mathworks, Natwick, MA) and Dexterit-E (BKIN Technology), while the Preserve task was run using Matlab (The Mathworks, Natwick, MA) and Psychophysics toolbox (Brainard, 1997). The Acquire task employed the same paradigm and equipment as in Holland et al. (2018), with the exception of the maximum reaction time (RT), which was increased from 0.6s to 1s, and the maximum movement time, which was reduced from 1s to 0.6s. The Preserve task used the same setup and display as in Codol et al. (2018); however, the number of 'refresher' trials during the binary feedback (BF) blocks was increased from one to two in every 10 trials. The designs were kept as close as possible to their respective original publications to promote replication and comparability across studies. In both tasks reaching movements were made with the dominant arm. Both the Acquire and Preserve tasks have previously been examined in isolation from each other (Acquire Task: Cashaback et al., 2017, 2019; Holland et al., 2018; Therrien et al., 2016, 2018; Preserve: Codol et al., 2018; Shmuelof et al., 2012) and we maintain this distinction here. However, it should be noted that the two tasks are essentially visuomotor rotation tasks. One of the aims of this study was to determine if similar mechanisms underly the use of binary feedback in both the

learning of a gradual rotation and maintenance of a previously learnt abrupt rotation. Therefore, despite the similarities we analyse the results of each task in isolation in addition to comparing the results across tasks.

Reaching tasks

Acquire task. Participants performed 670 trials, each of which followed a stereotyped timeline. The starting position for each trial was in a consistent position roughly 30cm in front of the midline and was indicated by a red circle (1cm radius). After holding the position of the handle within the starting position, a target (red circle, 1cm radius) appeared directly in front of the starting position at a distance of 10cm. Participants were instructed to make a rapid 'shooting' movement that passed through the target. If the cursor position at a radial distance of 10cm was within a reward region (±5.67°, initially centred on the visible target; grey region in Figure 1a) the target changed colour from red to green and a green tick was displayed just above the target position, informing participants of the success of their movement. However, if the cursor did not pass through the reward region, the target disappeared from view and no tick was displayed, signalling failure (binary feedback). After each movement, the robot returned to the starting position and participants were instructed to passively allow this.

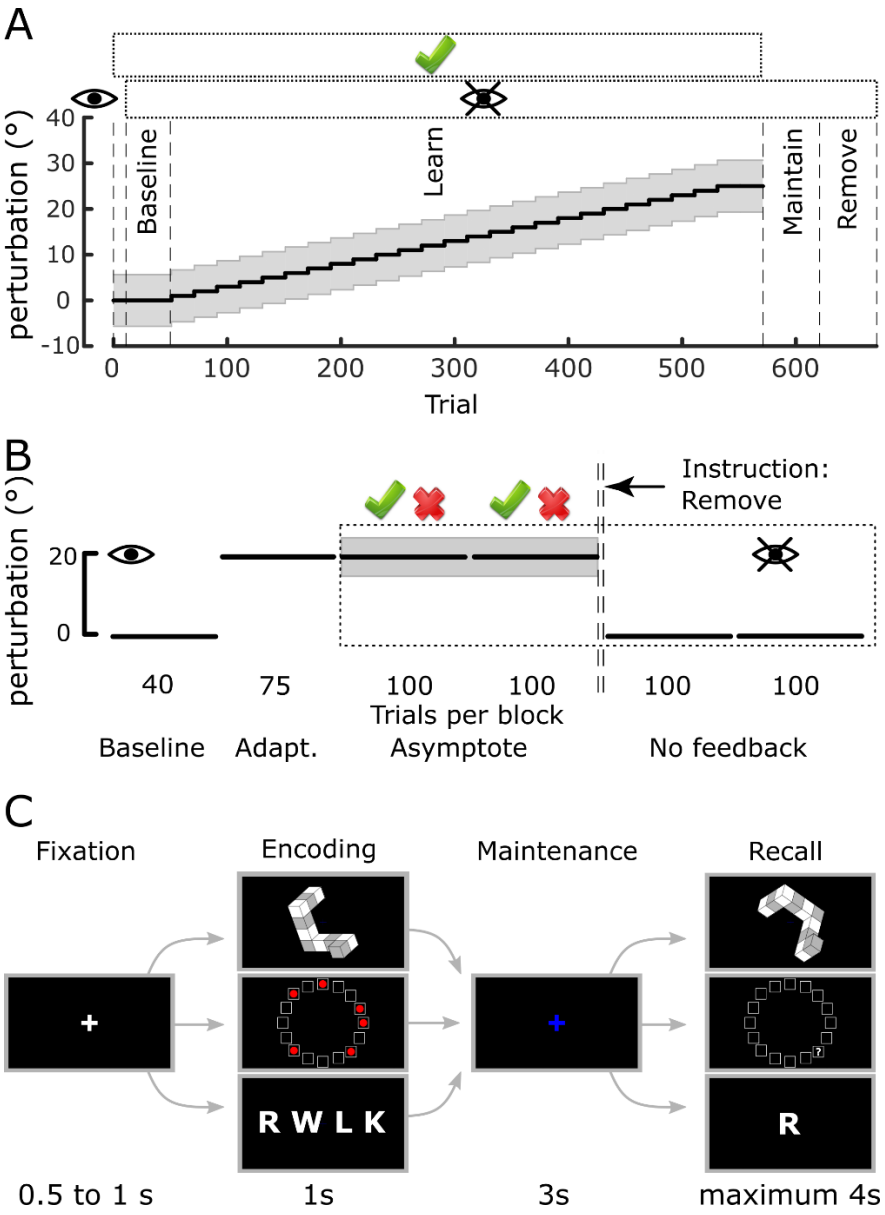
For the first 10 trials, the position of the robotic handle was displayed as a white cursor (0.5 cm radius) on screen. Following this practice block, the cursor was extinguished for the remainder of the experiment and participants only received binary feedback. The baseline block consisted of the first 40 trials under binary feedback. During this period the reward region remained centred on the visible target. Subsequently, unbeknownst to the participant, the reward region rotated in steps of 1° every 20 trials; the direction of rotation was counterbalanced across participants. After reaching a rotation of 25°, the reward region was held constant for an additional 20 trials. Performance during these last 20 trials was used to

determine overall task success. Subsequently, binary feedback was removed, and participants were instructed to continue reaching as they were (maintain block) for the following 50 trials. Following this, participants were then informed that the reward region shifted during the experiment but not of the magnitude or the direction of the shift. They were then instructed to return to reaching in the same manner as they were at the start of the experiment (remove block, 50 trials). During the learning phase of the task participants were given a 1-minute rest after trials 190 and 340.

Preserve task. Participants performed 515 trials in total. On each trial participants were instructed to make a rapid 'shooting' movement that passed through a target (white circle, radius: 0.125cm) visible on the screen. The starting position for each trial was indicated by a white square (width: 1cm) roughly 30cm in front of the midline and the target was located at angle of 45° from the perpendicular in a counter clockwise direction at a distance of 8cm. The position of the tracking device attached to the fingertip was displayed as a cursor (green circle, radius: 0.125cm). When the radial distance of the cursor from the starting position exceeded 8cm, the cursor feedback disappeared, and the end position was displayed instead.

First, participants performed a baseline period of 40 trials, during which the position of the cursor was visible, and the cursor accurately reflected the position of the fingertip. In the adaptation block (75 trials), participants were exposed to an abruptly introduced 20° clockwise visuomotor rotation of the cursor feedback (Figure 1b). Subsequently, all visual feedback of the cursor was removed, and participants received only binary feedback. If the end position of the movement fell within a reward region, the trial was considered successful and a tick was displayed; otherwise a cross was displayed. The reward region was centred at a clockwise rotation of 20° with respect to the visual target with a width of 10°, that is, it was centred on the direction that successfully accounted for the previously experienced visuomotor rotation. Binary feedback was provided for 200 trials divided into 2 blocks of 100 trials (asymptote

198 blocks). Furthermore, participants experienced 2 ‘refresher’ trials for every 10 trials, where
199 rotated visual feedback of the cursor position was again accessible (Codol et al., 2018;
200 Shmuelof et al., 2012). This represents an increase compared to Codol et al. (2018) because
201 participants in this study tended to have poorer performance under binary feedback, possibly
202 due to a fatigue effect following the WM tasks (Anguera et al., 2012; see discussion). Finally,
203 two blocks (100 trials each) with no performance feedback were employed in order to assess
204 retention of the perturbation (no-feedback blocks). Before the first of those two blocks,
205 participants were informed of the visuomotor rotation, asked to stop accounting for it through
206 aiming off target and to aim straight at the target.
207



208

209

210 **Figure 1: Experimental design. A**: Time course of the Acquire task with the different
211 experimental periods labelled. The grey region represents the reward region, which gradually
212 rotated away from the visual target after the initial baseline period. The rectangle enclosing the
213 green tick above the axes represents trials in which reward was available, and the rectangle
214 with the 'eye' symbol indicates when vision was not available. **B**: Time course of the Preserve
215 task. After adapting to an initial rotation with vision available, vison was removed (eye symbol)
216 and reward-based feedback was introduced (tick and cross above the axes). Prior to the no-
217 feedback blocks participants were instructed to remove any strategy they had been using. **C**:
218 WMC tasks, the three tasks followed a stereotyped timeline with only the items to be
219 remembered differing. Each trial consisted of 4 phases (Fixation, Encoding, Maintenance, and
220 Recall) with the time allocated to each displayed below.

221

222

Working memory tasks

Participants performed three WM tasks, all of which followed the same design with the exception of the nature of the items to be remembered (Figure 1c). All WM tasks were run using Matlab (The Mathworks, Natwick, MA) and Psychophysics toolbox (Brainard, 1997). At the start of each trial, a white fixation cross was displayed in the centre of the screen for a period of 0.5 to 1s randomly generated from a uniform distribution (fixation period in Figure 1c). In the encoding period, the stimuli to be remembered was displayed for 1s and then subsequently replaced with a blue fixation cross for the maintenance period which persisted for 3s. Finally, during the recall period, participants were given a maximum of 4s to respond by pressing one of three keys on a keyboard with their dominant hand. The '1' key indicated that the stimuli presented in the recall period was a 'match' to that presented in the encoding period, the '2' key indicated a 'non-match', and '3' indicated that the participant was unsure as to the correct answer. Each WM task contained 5 levels of difficulty with the 12 trials presented for each; 6 of which were trials in which 'match' was the correct answer and 6 in which 'non-match' was the correct answer. Consequently, each WM task consisted of 60 trials and the order in which the tasks were performed was pseudorandomised across participants. Prior to the start of each task participants performed 10 practice trials to familiarise themselves with the task and instructions. For both the Acquire and Preserve tasks, the WM tasks were performed in the same experimental session as the reaching. However, in the case of the Acquire task the WM tasks were performed after the reaching task whereas for the Preserve task the WM tasks were performed first.

In the RWM task (Figure 1c, top row), the stimuli consisted of six 2D representations of 3D shapes drawn from an electronic library of the Shepard and Metzler type stimuli (Peters and Battista, 2008). The shape presented in the recall period was always the same 3D shape presented in the encoding period after undergoing a screen-plane rotation of 60°, 120°, 180°,

240° or 300°. In 'match' trials, the only transform applied was the rotation; however, in 'non-match' trials an additional vertical-axis mirroring was also applied. The difficulty of mental rotation has been demonstrated to increase with larger angles of rotation (Shepard and Metzler, 1971) and therefore the different degrees of rotation corresponded to the 5 levels of difficulty. However, given the symmetry of two pairs of rotations (60 and 300, 120 and 240), these 5 levels were collapsed to 3 for analysis.

In the SWM task (Figure 1c, middle row), stimuli in the encoding period consisted of a variable number of red circles placed within 16 squares arranged in a circular array (McNab and Klingberg, 2008). In the recall period, the array of squares was presented without the red circles and instead a question mark appeared in one of the squares. Participants then answered to the question '*Was there a red dot in the square marked by a question mark?*' by pressing a corresponding key. In 'match' trials the question mark appeared in one of the squares previously containing a red circle and in 'non-match' trials it appeared in a square that was previously empty. Difficulty was scaled by varying the number of red circles (i.e. the number of locations to remember) from 3 to 7.

In the VWM task (Figure 1c, bottom row), participants were presented with a list of a variable number of consonants during the encoding period. In the recall period a single consonant was presented, and participants answered to the question '*Was this letter included in the previous array?*'. Thus, the letter could either be drawn from the previous list ('match' trials) or have been absent from the previous list ('non-match' trials). Difficulty in this task was determined by the length of the list to be remembered, ranging from 5 to 9.

Both the SWM and RWM tasks have been suggested to fall under the general umbrella term of spatial ability (Buszard and Masters, 2018). However, Miyake et al. (2001) suggest that although both RWM and short term storage of spatial information (i.e. SWM) are within the spatial domain, RWM appears to rely more heavily on executive function and SWM on basic

short term storage of spatial information. Furthermore, previous studies have found relationships between motor learning and this SWM task (Christou et al., 2016; Vandevoorde and Orban de Xivry, 2019) and tasks similar to our RWM task (Anguera et al., 2010). Therefore, we included both tasks to investigate if there was any severability in their relationships with reaching performance and leveraged our use of two separate reaching tasks and large cohorts to probe if this was due to specific task parameters.

Genetic sample collection and profiling

COMT is thought to affect DA function mainly in the prefrontal cortex (Egan et al., 2001; Goldberg et al., 2003), a region known for its involvement in WM and strategic planning (Anguera et al., 2010; Doll et al., 2015), whereas DARPP32 and DRD2 act mainly in the basal ganglia to promote exploitative behaviour, possibly by promoting selection of the action to be performed (Frank et al., 2009). Consequently, we focused here on single-nucleotide polymorphisms (SNP) related to those genes: RS4680 (COMT) and RS907094 (DARPP32). Regarding DRD2, there are two potential SNPs available, RS6277 and RS1800497. Although previous studies focusing on exploration and exploitation have assessed RS6277 expression (Doll et al., 2016; Frank et al., 2007, 2009), it should be noted that this SNP varies greatly across ethnic groups, with some allelic variations being nearly completely absent in non-Caucasian-European groups (e.g. see RS6277 in 1000 Genomes Project (The 1000 Genomes Project Consortium et al., [2015]). This has likely been inconsequential in previous work, as Caucasian-European individual represented the majority of sampled groups; here however, this represents a critical shortcoming, as we aim at investigating a larger and more representative population including other ethnic groups. Consequently, we based our analysis on the RS1800497 allele of the DRD2 gene (Pearson-Fuhrhop et al., 2013).

At the end of the task, participants were asked to produce a saliva sample which was collected, stabilized and transported using Oragene.DNA saliva collection kits (OG-500, DNAgenotek, Ontario, Canada). Participants were requested not to eat or drink anything except water for at least two hours before sample collection. Once data collection was completed across all participants, the saliva samples were sent to LGC (Hoddeson, Hertfordshire; https://www.lgcgroup.com/) for DNA extraction (per Oragene protocols: https://www.dnagenotek.com/) and genotyping. SNP genotyping was performed using the KASP SNP genotyping system. KASP is a competitive allele-specific PCR incorporating a FRET quencher cassette. Specifically, the SNP-specific KASP assay mix (containing two different, allele specific, competing forward primers) and the universal KASP master mix (containing FRET cassette plus Taq polymerase in an optimised buffer solution) were added to DNA samples and a thermal cycling reaction performed, followed by an end-point fluorescent read according to the manufacturer's protocol. All assays were tested on in-house validation DNA prior to being run on project samples. No-template controls were used, and 5% of the samples had duplicates included on each plate to enable the detection of contamination or non-specific amplification. All assays had over 90% call rates. Following completion of the PCR, all genotyping reaction plates were read on a BMG PHERAStar plate reader. The plates were recycled until a laboratory operator was satisfied that the PCR reaction had reached its endpoint. In-house Kraken software then automatically called the genotypes for each sample, with these results being confirmed independently by two laboratory operators. Furthermore, the duplicate saliva samples collected from 5% of participants were checked for consistency with the primary sample. No discrepancies between primary samples and duplicates were discovered.

Data analysis

Acquire task: Reach trials containing movement times over 0.6s or less than 0.2s were removed from analysis (6.9% of trials). The end point angle of each movement was defined at the time when the radial distance of the cursor exceeded 10cm. This angle was defined in relation to the visible target with positive angles indicating clockwise rotations. End point angles and target angles for participants who experienced the counter clockwise rotations were sign-transformed. The explicit component of retention was defined as the difference between the mean reach angle of the maintain block and the remove block, while the implicit component was the difference between the mean reach angle of the remove block and baseline (Werner et al., 2015). Participants that achieved a mean reach angle within the reward region during the final 20 trials before the maintain block were considered '*successful*' in learning the rotation; otherwise they were considered '*unsuccessful*'. As in Holland et al. (2018), for unsuccessful participants, the largest angle of rotation at which the mean reach angle fell within the reward region was taken as the end of successful performance, and only trials prior to this point were included for further analysis. Success rate was defined as the percentage of trials during the learning blocks in which the end point angle was within the reward region. In order to examine the effect of reward on the change in end point angle on the subsequent trial, we examined the magnitude and variability of changes in end point angle between consecutive trials (Holland et al., 2018; Sidarta et al., 2018; Therrien et al., 2016, 2018). To calculate the median absolute change following rewarded (ΔR) and unrewarded (ΔP) trials we extracted the changes in reach angle following each trial type and calculated the median of the absolute values of these changes for each participant. These measures therefore represent the median of the magnitude of changes in reach angle, regardless of direction. Furthermore, in order to examine the variability of trial-by-trial adjustments (MAD[ΔR] and MAD[ΔP] for rewarded and unrewarded trials, respectively) we calculated the median absolute deviation of the changes in reach angle. It is

important to note that ΔR and ΔP are calculated from the absolute magnitude of the changes in reach angle, whereas, MAD[ΔR] and MAD[ΔP] are calculated from the non-absolute values (including the direction of change).

Preserve task: Reach trials containing movement times over 1s were removed from analysis (2.38% of trials). The end point angle for each movement was defined at the time that the radial distance of the cursor from the start position exceeded 8cm. Trials in which the error was greater than 80° were excluded from further analysis (0.94% of trials). As in Codol et al. (2018), learning rate was calculated by fitting an exponential function to the angular error between cursor and target for trials in the adaptation block, with the β value taken as the learning rate (mean R^2=0.34±0.15). The β estimates attained from all fits were first sign transformed and then log-transformed to counteract skewness prior to entering the regression analysis. Using this method, a value close to 0 indicated faster learning, whereas more negative values indicated slower learning. Similar to Codol et al (2018), success rate, corresponding to percentage of rewarded trials, was measured separately in the first 30 and last 170 trials of the asymptote blocks and labelled early and late success rate, respectively. This reflects a dichotomy between a dominantly exploration-driven early phase and a later exploitation-driven phase. The analysis of changes in reach angle (ΔR and ΔP) was confined to the last 170 trials of the asymptote blocks. Implicit retention was defined as the difference between the average baseline reach direction and the mean reach direction of the last 20 trials of the last no-feedback block (Codol et al., 2018). Analysis of changes in reach angle following rewarded trials were not pre-registered but were included *post-hoc*.

Exploratory analysis of reaching data: In order to understand which outcome variables in the reaching tasks were predictive of overall task success, we split the learning period into two sections for every participant. We assessed trial-by-trial changes in end point angle in the first section and compared them to success rate in the second section. For the Acquire task, we

assessed trial-by-trial adjustments during the learning block, excluding the final 20 trials, and compared them to success rate in the last 20 trials of the learning block. In the Preserve task, we measured adjustments in the first 100 trials of the asymptote blocks and compared them to success rate in the last 100 trials of the asymptote blocks.

WM tasks: WM performance was defined as the average percentage of correct responses across the 3 highest levels of difficulty for each task. In the case of the RWM task, the symmetrical arrangement of the angles of rotation in effect produced three levels of difficulty and therefore all trials were analysed.

Genetics: Genes were linearly encoded, with heterozygote alleles being 0, homozygote alleles bearing the highest dopaminergic state being 1, and homozygote alleles bearing the lowest dopaminergic state being -1 (Table 1). All groups were assessed for violations of the Hardy-Weinberg equilibrium. The participant pool in the Preserve task was in Hardy-Weinberg equilibrium for all three genes considered. In the Acquire task population, COMT and DRD2 were in Hardy-Weinberg equilibrium, but DARPP32 was not ($p=0.002$), with too few heterozygotes. Therefore, the DARPP32 alleles were recoded, with the heterozygotes (0) and the smallest homozygote group (C:C, -1) combined and recoded as 0.

SNP	location	Allele code -1	Allele code 0	Allele code 1
rs4680	COMT	G:G (val:val)	A:G (met:val)	A:A (met:met)
		31, 33	68, 61	17, 21
rs1800497	DRD2	T:T (lys:lys)	T:C (lys:glu)	C:C (glu:glu)
		8, 7	48, 51	64, 62
rs907094	DARPP32	C:C	C:T	T:T
		10, 21	54, 38	56, 62

Table 1: Coding for SNPs. The name of the SNP is provided along with the code assigned to each allele. The numbers represent the counts for the specific allele in the two tasks (Preserve, Acquire).

Statistical analysis

Regressions were performed using the linear Lasso method (Tibshirani, 1996; *lasso* function in MatLab's *Statistics and Machine Learning* Toolbox). Lasso regression employs a shrinkage method that allows for some predictors to be shrunk to a value of 0, effectively removing them from the regression model. Therefore, the method acts as a selection method for predictors in an analogous way to stepwise regression. We used a 10-fold cross validation approach to calculate the Mean Squared Error (MSE) over a range of values of a penalty term λ. Specifically, as λ increases, the shrinkage of predictor values increases. For λ=0, the model reduced to a standard linear regression, as all predictors were included without any shrinkage. Cross validation protects against the problem of over-fitting by calculating the MSE on data 'unseen' by the model during fitting. For any given outcome variable, if its MSE(λ) function exhibited a minimum value within its defined boundaries, the model associated with that minimum value was considered selected. If no minimum was observed, this signified that an empty model was a better fit than any other possible model. If such minimum was detected in the MSE(λ) function, the β estimates from that model (i.e. at that value of λ) were taken. We repeated this

procedure 1000 times to obtain the distribution of the true β from the estimates (Hastie et al., 2015). In order for a potential variable to be considered a selected predictor, that predictor should be selected (i.e. $\beta \neq 0$) in at least 80% of the repetitions. The threshold of 80% was chosen as to maintain sufficient sensitivity whilst still returning relatively sparse models. We report the median β estimate in the text for all selected predictors.

In order to understand what genetic and WM factors are predictive of performance in the Acquire task, we performed a lasso regression of the seven predictors (three allelic variations, three WM and ethnicity) onto each of several outcome measures representative of performance: success rate, implicit and explicit retention, ΔR, MAD[ΔR], ΔP, MAD[ΔP].. For the Preserve task, we performed separate lasso regressions using the same seven predicators for the following outcome variables: baseline reach direction as a control variable, learning rate in the adaptation block, early and late success rate in the asymptote blocks (first 30 and last 170 trials; Codol et al., 2018), retention in the no-feedback blocks, and ΔR and ΔP during the asymptote blocks. We adopted a parsimonious approach when interpreting the results of the regression analysis and gave particular credence to results reproduced by the analysis across both tasks.

Prior to the regression analysis, all predictors and predicted variables were standardised (z-scored). For all non-ordinal variables, individual data were considered outliers if further than 3 standard deviations from the mean and were removed prior to standardisation. Multicollinearity of predictors was also assessed before regression with Belsley Collinearity Diagnositcs (*collintest* function in MatLab's *Econometrics Toolbox*) and no predictors were found to exhibit condition indexes over 30, indicating acceptable levels of collinearity. When considering retention for both tasks, unsuccessful participants were removed from the regression analysis. We further characterised the relationships between predictor variables by combining the data for the two tasks for the working memory (WM) tasks and the genetic codes

(N=241). We analysed relationships between the WM tasks with correlations and between genetics and WM tasks with one-way ANOVAs.

Exploratory mediation analysis: We performed a mediation analysis to test if the relationship between SWM and SR was mediated by ΔR. Our hypothesis was that higher SWM enables smaller changes after correct trials (ΔR) and this then explains the relationship between SWM and SR. To ensure that separate trials were used in the calculation of ΔR and SR, we split the trials into two equally sized folds. The SR was then calculated for one-fold as a percentage of correct trials, and ΔR was calculated as the median absolute change of reach angle after correct trials in the other fold. For the Acquire task only successful subjects were included in the mediation analysis. We employed Baron & Kenny's three step mediation analysis (Baron and Kenny, 1986): first regress SR on SWM, then regress ΔR on SWM, and finally regress SR on both SWM and ΔR. Subsequently, we performed a Sobel test to determine if there was a significant reduction in the relationship between SWM and SR when including ΔR. The Sobel test examines if the amount of variance in SR explained by SWM is significantly reduced by including the mediator (Sobel, 1986). For a significant effect to be found, SWM must be a significant predictor of ΔR and ΔR must also be a significant predictor of SR after controlling for the effect of SWM on SR. We repeated this procedure 1000 times with the allocation of trials to each fold randomised on each repetition. We present results in terms of the 95% confidence intervals for the R^2 values for each of the regressions and the median p-value of the Sobel test, along with the associated 95% confidence intervals. An alternative possibility to the hypothesized model is that the relationship between SWM and ΔR is mediated by SR. In order to compare the size of the mediation effect for these alternate models, we follow the Mackinnon and Dwyer (1993) procedure and normalize the size of the indirect effect by dividing it by the sum of the direct and indirect effects. This analysis allows to express the mediation effect in

terms of percentage of the total effect. We present the median of the normalized value for the 1000 repetitions on both the hypothesized and alternate models.

Results

Acquire task

In the Acquire task, participants had to learn to compensate for a secretly shifting reward region in order to obtain successful feedback (Figure 2, 3). As in Holland et al. (2018), about a quarter (28.1%) of participants failed to learn to compensate for the full extent of the rotation (Figure 3a). The inability of a significant proportion of participants to learn the full extent of the rotation is also consistent with previous reports in reward-based motor learning paradigms (Cashaback et al., 2019; Codol et al., 2018; Saijo and Gomi, 2010; Therrien et al., 2016, 2018). Successful participants retained most of the learnt rotation (mean 80.7% ± 28.0% SD) in the maintain block. This level of retention is in accordance with that reported previously in similar paradigms (Holland et al., 2018; Therrien et al., 2016). However, upon being asked to remove any strategy they had been employing, their performance returned to near-baseline levels. Consequently, a large explicit component to retention was found for successful participants (Figure 3b), whereas both successful and unsuccessful participants manifest a small but non-zero implicit component ($t(86)=9.90$, $p=7.43\times10^{-16}$, $d=1.061$ and $t(33)=4.53$, $p=7.39\times10^{-5}$, $d=0.776$, respectively; Figure 3c). The persistent implicit retention is a common finding of retention periods in which no visual feedback is provided and may reflect a combination of implicit reinforcement (Shmuelof et al., 2012), use-dependent plasticity (Diedrichsen et al., 2010), perceptual bias (Vindras et al., 1998), or perceptual recalibration (Modchalingam et al., 2019). Furthermore, in accordance with Holland et al (2018), we found that participants made larger ($t(120)=15.80$, $p=4.32\times10^{-31}$, $d=1.900$) and more variable changes in reach angle following unrewarded trials ($t(120)=14.54$, $p=3.144\times10^{-28}$, $d=1.667$; Figure 3d-g). However, in

participants who would go on to fail, the post-error adjustments were smaller than in successful participants (t(119)=3.33, p=0.001, d=0.672; Figure 3d). Changes following rewarded trials were similar between the groups (t(119)=0.71, p=0.48, d=0.143; Figure 3f,g). The results obtained in this sample (N=121) therefore replicate results from a previous study (N=30) and provides further confirmation that performance in this task is fundamentally explicitly driven (Holland et al., 2018).

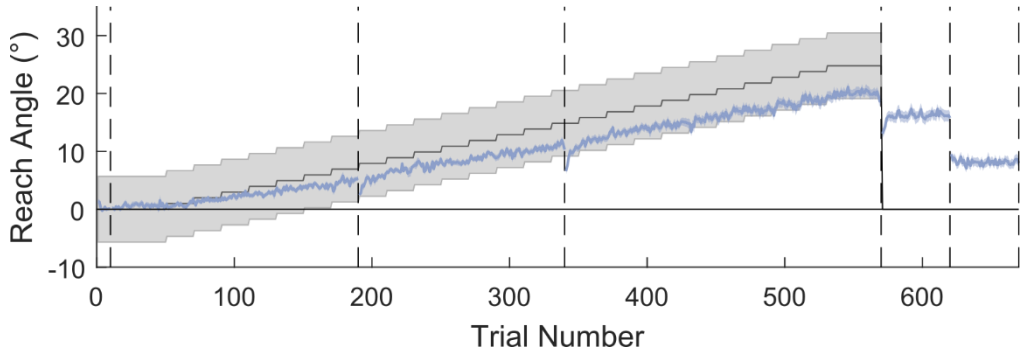


Figure 2: Reaching performance in the Acquire task. The grey region represents the gradually rotating rewarded region, the blue line represents mean reach angle for each trial, and the shaded blue region represent SEM. Vertical dashed lines represent different experiment blocks or breaks. Average performance for the full cohort falls within the reward region and demonstrates a clear reduction in reach angle when asked to remove strategy. N=121.

In order to understand what genetic and WM factors are predictive of performance in the reaching task, we performed a lasso regression of the seven predictors (three allelic variations, three WM and ethnicity) onto each of several outcome measures representative of performance: success rate, implicit and explicit retention, ΔR, MAD[ΔR], ΔP, MAD[ΔP].

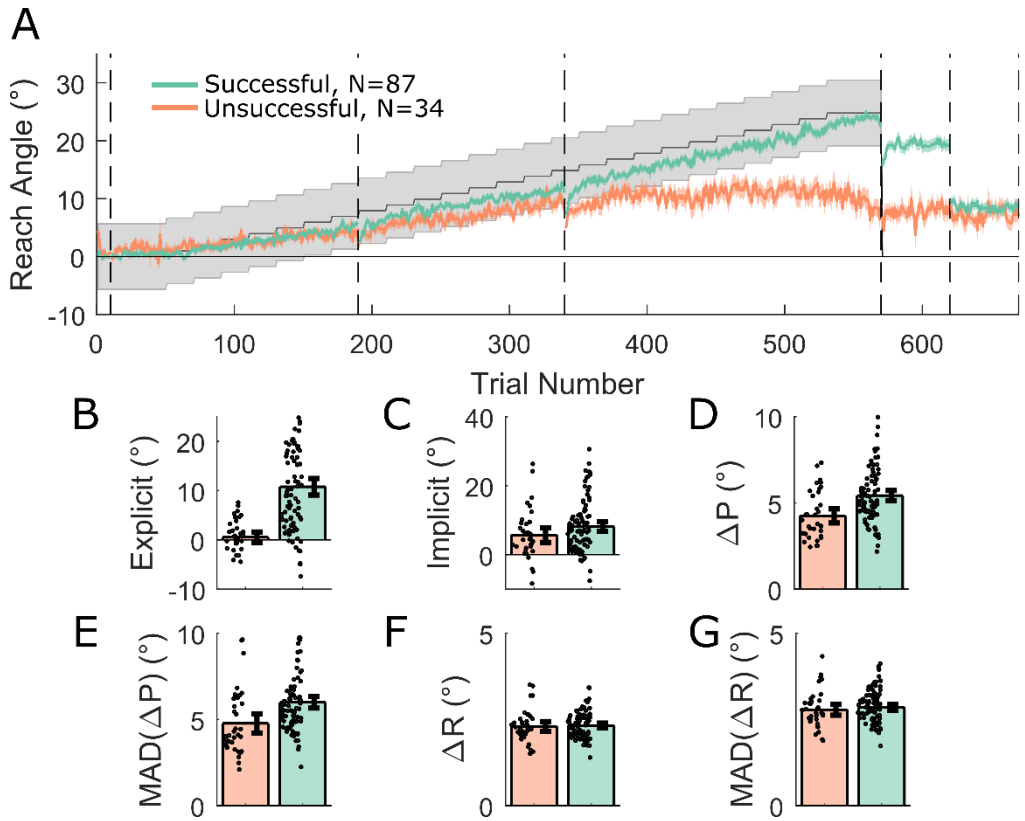


Figure 3: Acquire task split by success at final angle. A: Average reach angle for the successful (green) and unsuccessful (orange) groups, shaded regions represent SEM and grey shaded region represents the rewarded region. Despite similar initial performance, a clear divergence can be seen between the two groups and an explicit component to retention is only visible in the successful group, whereas implicit retention is similar between groups. **B-G**: subplots displaying derived measures, which acted as outcome variables for the regression analysis, separated into successful and unsuccessful participants overlaid with individual data points. Error bars represent 95% bootstrapped confidence intervals. ΔR and ΔP refer to changes made in reach angle after rewarded and unrewarded trials respectively. The bar plots in panels D and F display the median absolute change and panels E and G display the median absolute deviation of the changes in angle after each trial type.

For success rate, SWM, RWM and DRD2 were selected as predictors (median β=0.31, 0.06, and 0.03, respectively; Figure 4a), with the strongest predictor being SWM. Figure 5 displays the effect of the strongest predictor selected for each outcome variable and shows that there was a positive relationship between SWM and success rate (Figure 5a). To ensure that the relationship between SWM and success rate was not due to failure at a later point in the task, success rate was measured during the initial period in which subjects who could not fully

account for the displacement are still successful; for those who could, the full learning block was included.

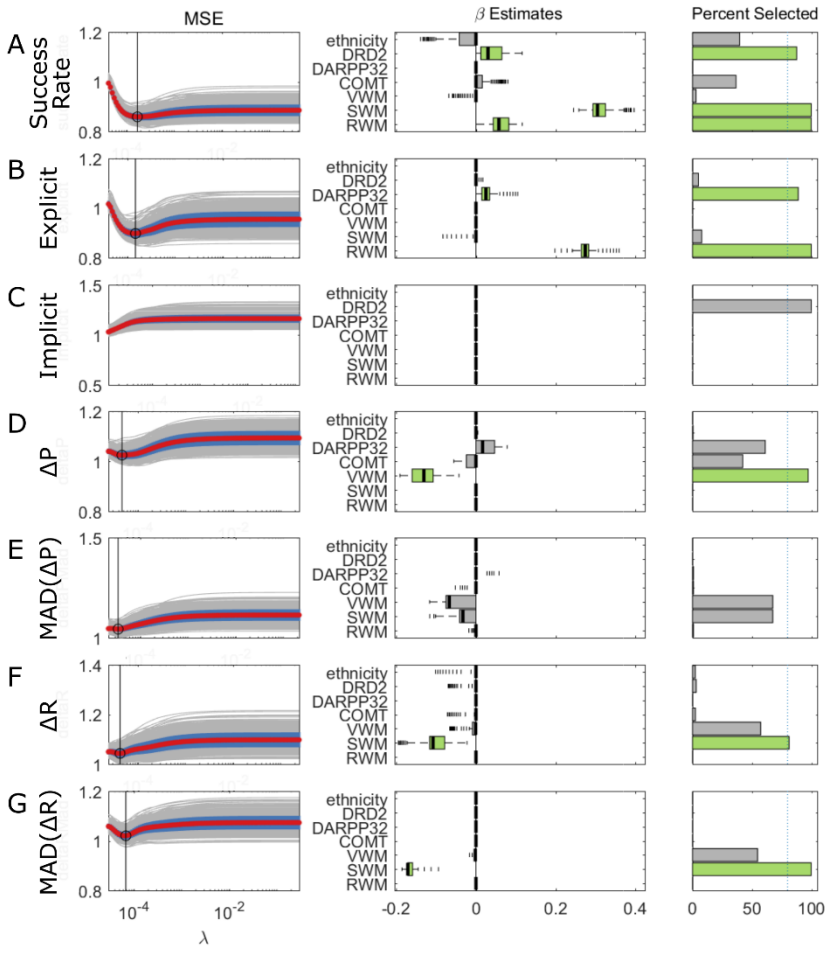


Figure 4: Lasso regression results for the Acquire task. Each row (A-G) represents the results from one outcome variable. The left column indicates the MSE as a function of changing the shrinkage parameter λ, with larger values of λ representing greater penalization and sparser models. A minimum in the MSE within its defined boundaries indicates the suitability of that choice of λ and is indicated with a vertical line. Given the presence of a minimum, the values of the β for each predictor are taken. We performed 1000 repetitions of the lasso regression for each outcome variable and box plots indicating the distribution of the coefficient estimates are displayed in the middle panel. The rightmost column indicates the percentage of times that the individual predictors were assigned non-zero coefficients. We employed a threshold of 80% (indicated with a dashed vertical line) to signify that a particular predictor was robustly selected, and these variables are highlighted in green. Median absolute change in reach angle after rewarded (ΔR) and unrewarded (ΔP) trials. Median absolute deviation of change in reach angle after rewarded (MAD[ΔR]) and unrewarded (MAD[ΔP]) trials.

Next, retention was assessed by splitting up the explicit and implicit components such as in Holland et al. (2018). No predictor was related to the implicit component, but the explicit component was strongly and positively associated with RWM (β=0.27; Figure 4b, 5b) with a weaker association between DARPP32 and explicit retention (β=0.03). These results suggest positive relationships for both RWM and SWM with task performance: greater RWM predicts a greater contribution of explicit processes to learning, whereas greater SWM predicts a greater percentage of correct trials.

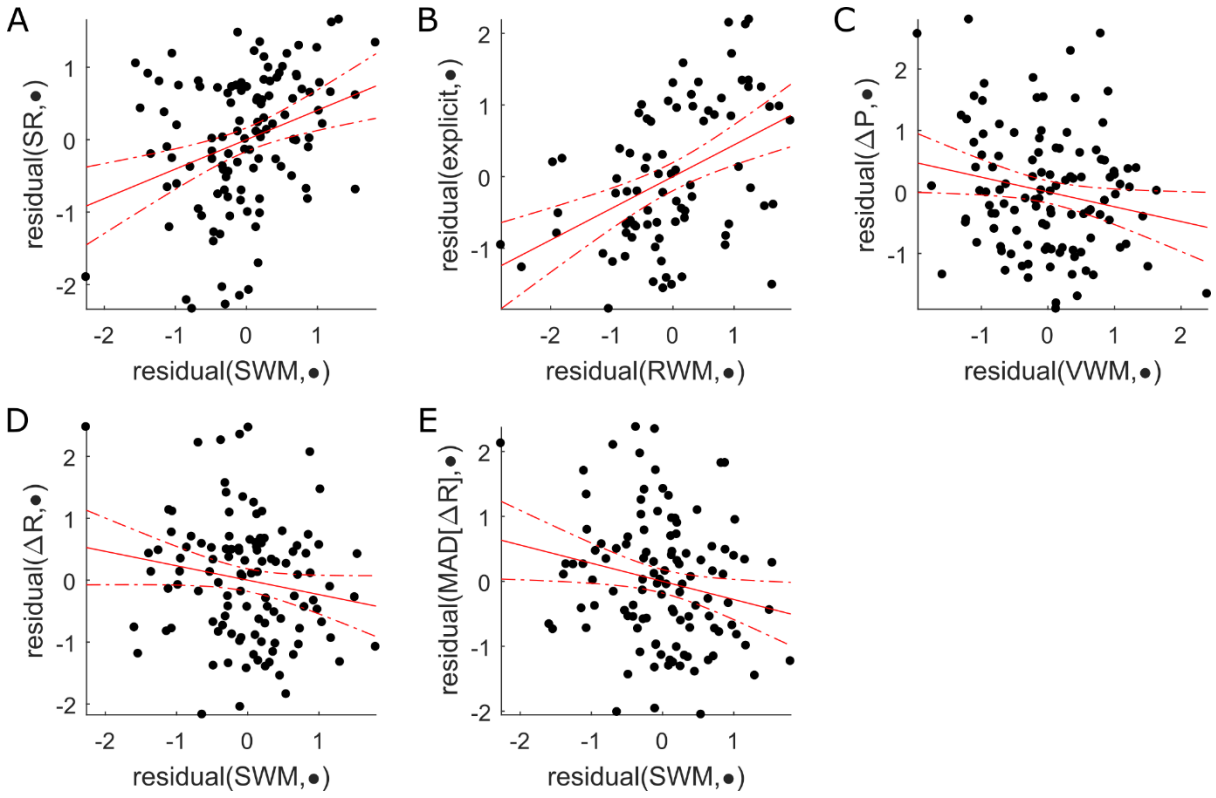


Figure 5: Added variable plots for selected predictors in the Acquire task. Each plot displays the relationship between the strongest predictor selected by the lasso regression (x-axis), and the corresponding outcome variable (y-axis). Added variable plots display the residuals of regressing the response variable with all remaining independent variables, and the residuals of the regression of the selected predictor to the remaining predictors. The resulting relationship corresponds to the effect of the selected predictor on the outcome measure after controlling for the remaining predictors. SR: Success Rate. Median absolute change in reach angle after rewarded (ΔR) and unrewarded (ΔP) trials. MAD(ΔR): Median absolute deviation of change in reach angle after rewarded trials.

In Holland et al (2018), the amplitude of the changes in reach angle participants made following unrewarded trials was found to be predictive of task success, that is, greater ΔP was predictive of an increased chance of overall task success. Thus, it could be that RWM and SWM, that are shown to associate with performance in this study, are themselves predictors of changes in reach angle. Conformingly, the regression results demonstrated that a large ΔR was inversely related to SWM (β=-0.11; Figure 4f, 5d), as was MAD[ΔR] (β=-0.17; Figure 4g, 5e). The results indicate that greater SWM was predictive of smaller and less variable changes in reach angle after successful trials, suggesting high SWM enables the maintenance of rewarding reach angles. It was also found that changes in reach angle following unrewarded trials (ΔP) were negatively associated with VWM (β=-0.13, Figure 4d, 5c). This result was unexpected as it suggests that greater WMC predicts smaller changes following unrewarded trials, whereas previous results suggest a positive relationship between the amplitude of these changes and overall task success. Although the difference may be due to the domain of WM under consideration, it is unclear as to the reason for this relationship. Another important aspect of the analysis of trial-to-trial changes to control for is that the numbers of trials analysed and their phase in the experiment differs between successful and unsuccessful subjects. Therefore, we repeated the Lasso regression while only including successful subjects. The predictors that were selected were identical to those selected when using the full participant pool.

Overall, WM (in particular RWM and SWM) successfully predicted various aspects of performance in the Acquire task, while genetic predictors generally failed to do so. Specifically, greater SWM predicted smaller and less variable changes after correct trials. This suggests that SWM underlies one's capacity to preserve and consistently express an acquired reach direction to obtain reward. Furthermore, SWM also directly predicted success rate. In addition, greater RWM was a strong predictor of explicit control. The inverse relationship between VWM and the magnitude of changes after unrewarded trials was unexpected. However, one possible

explanation is that participants with poorer WMC make larger errors which require larger corrections.

Preserve task

In this task, we addressed how well participants can maintain a previously learnt adaptation after transitioning to binary feedback. As participants are unable to compensate for a large abrupt displacement of a hidden reward region (van der Kooij and Overvliet, 2016; Manley et al., 2014), participants first adapted to an abruptly introduced 20° clockwise rotation with full vision of the cursor available. Subsequently, visual feedback of the cursor position was replaced with binary feedback; participants were rewarded if they continued reaching towards the same angle that resulted in the cursor hitting the target during the adaptation phase. Overall, participants adapted to the visuomotor rotation successfully (Figure 6, 7a-c) before transitioning to the binary feedback-based asymptote blocks. However, from the start of the asymptote blocks onward, participants exhibited very poor performance, expressing an average 45.0 ± 24.2 SD% success rate when considering all 200 asymptote trials (Figure 6, 7a, d,e). We have previously shown in (Codol et al., 2018) that this drop in performance (Shmuelof et al., 2012) represents exploratory behaviour that arises due to a lack of transfer of the cerebellar memory between the two contexts. Separating successful and unsuccessful participants (40% success rate cut-off; Figure 7a) revealed that successful participants expressed behaviour greatly similar to that observed in Codol et al. (2018), in which unsuccessful participants were excluded, using the same cut-off (40% success rate). The 'spiking' behaviour observed in reach angles during the asymptote blocks (Figure 7a) is due to the presence of the 'refresher' trials, with the large positive changes in reach angle corresponding to trials immediately following the refresher trials. This pattern of behaviour is particularly pronounced in the unsuccessful participants. Finally, participants demonstrated at least a residual level of retention even after

being instructed to remove any strategy they had employed ($t(69)=7.268$, $p=3.345x10^{-10}$, $d=0.869$; Figure 7a,f). Therefore, the results obtained in this sample (N=120) replicate results from a previous study (Codol et al., 2018; N=20, BF-Remove group) and provides further confirmation that performance in this task is fundamentally explicitly driven. It should also be noted that the successful group displayed higher implicit retention than the unsuccessful participants. As with the Acquire task successful participants displayed larger changes in angle after unrewarded trials than their unsuccessful counterparts ($t(117)=3.847$, $p=1.952x10^{-4}$, $d=0.717$; Figure 7h). However, in contrast to the Acquire task, successful participants also displayed smaller changes in angle after rewarded trials ($t(115)=-7.534$, $p=1.218x10^{-11}$, $d=1.421$; Figure 7g).

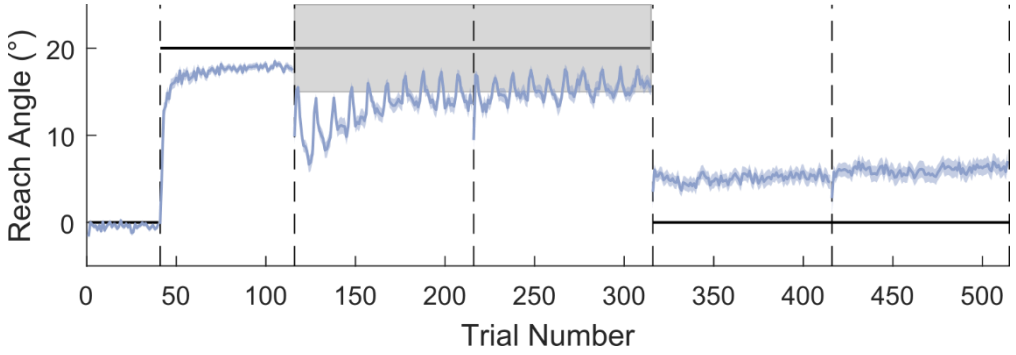


Figure 6: Reaching performance in the Preserve task. The grey shaded area represents the rewarded region, and the thick black line represents the perturbation. The vertical dashed lines represent block limits. The blue line indicates mean reach angle for every trial and blue shaded areas represent SEM. After successfully adapting to the visuomotor rotation performance deteriorates at the onset of binary feedback, subsequently success rate increases towards the end of the asymptote blocks. Following the removal of all feedback, and the instruction to remove any strategy, a small amount of implicit retention remains. N=120.

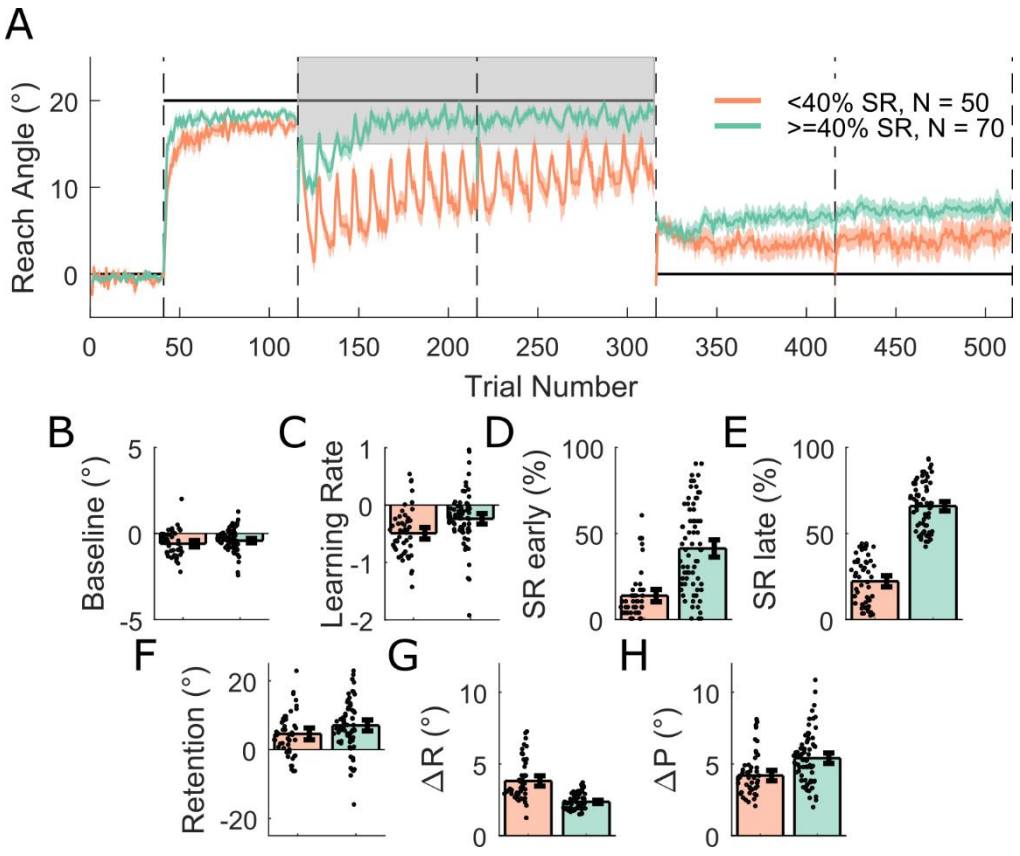


Figure 7: Preserve task split into two groups on the basis of success rate. A: Shaded regions represent SEM. **B-H**: Derived variables, which acted as outcome variables for the regression analysis, for the two groups, error bars on the bars represent 95% bootstrapped confidence intervals and individual data points are displayed. SR: Success Rate. Median absolute change in reach angle after rewarded (ΔR) and unrewarded (ΔP) trials.

As in the Acquire task, we examined if performance in any of the WM tasks or genetic profile could predict participants' behaviour in the reaching task. We performed separate lasso regressions for the following outcome variables: baseline reach direction as a control variable, learning rate in the adaptation block, early and late success rate in the asymptote blocks (first 30 and last 170 trials; Codol et al., 2018), retention in the no-feedback blocks, and ΔR and ΔP during the asymptote blocks. The most striking result was that both early and late success rate could be reliably predicted by RWM (early: β=0.17, late: β=0.12; Figure8c,d, and 9a,b), with greater RWM associated with increased success rates. An additional positive relationship was found between SWM and success rate but only during the later period (β=0.02; Figure8c).

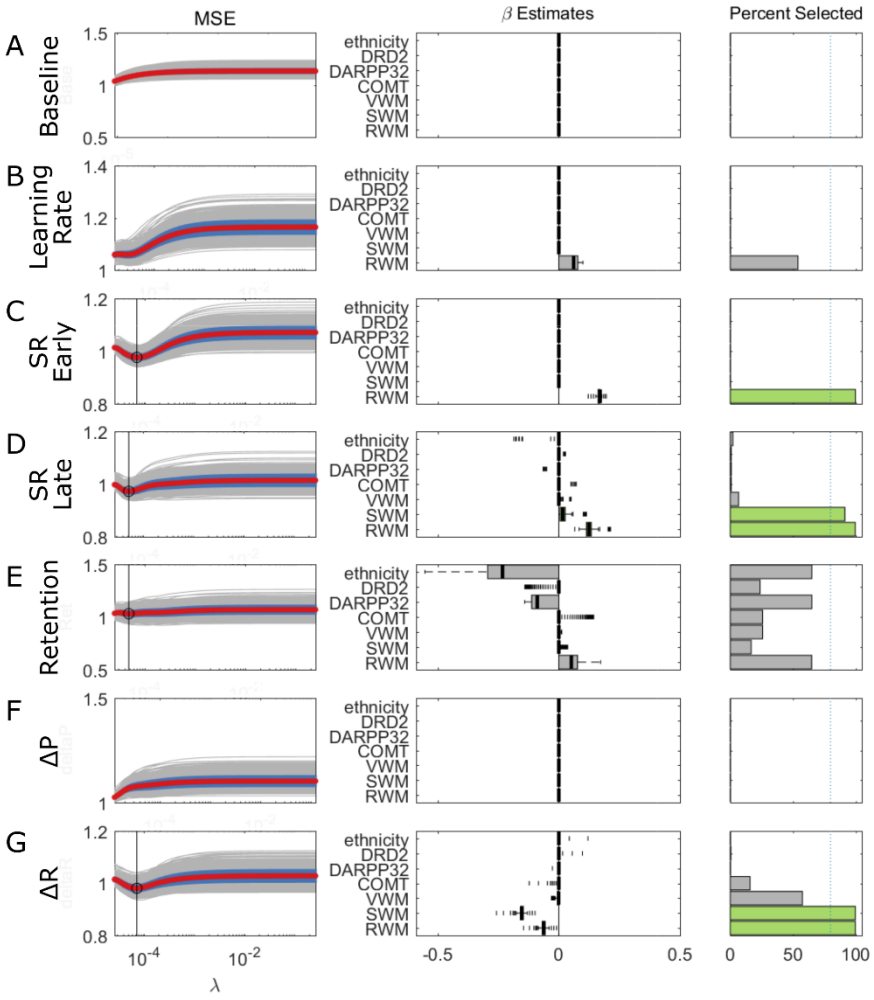


Figure 8: Results of the Lasso regressions for the Preserve task. The format is identical to Figure 4 with each row (A-G) representing the predictors of a single outcome measure. Selected predictors are highlighted in green, with the middle panels displaying the β estimates and the right panels displaying the probability of each predictor being selected. SR: Success Rate. Median absolute change in reach angle after rewarded (ΔR) and unrewarded (ΔP) trials.

Genetic profile did not predict any aspect of performance. In contrast, greater SWM successfully predicted reduced ΔR (β=-0.15; Figure 8g, 9c) similarly to the Acquire task. Additionally, there was a weaker relationship between RWM and ΔR (β=-0.06; Figure8g) which was absent in the Acquire task. Despite the presence of a local minimum in the MSE for the regression involving retention, no individual predictor was consistently selected in more than 80% of repetitions (Figure 8e).

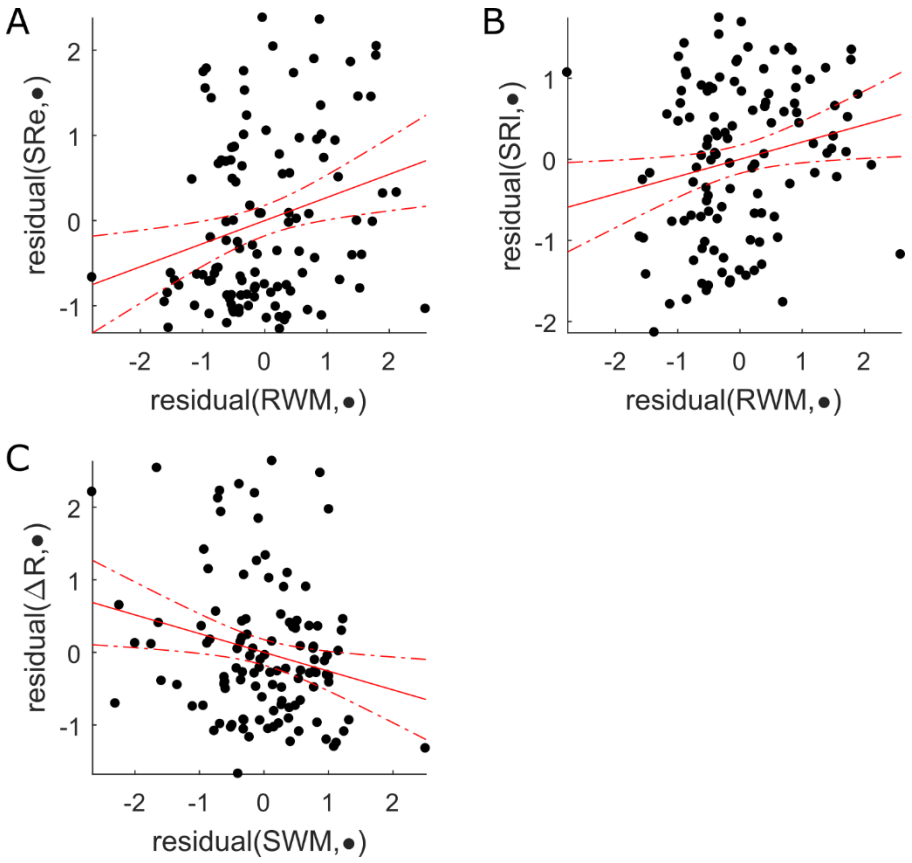


Figure 9: Added variable plots for selected predictors in the Preserve task. Each panel (A-C) displays the effect of the considered predictor when accounting for the effect of all other predictors. Results are displayed for the strongest selected predictor for each outcome measure. SRe: Early Success Rate. SRl: Late Success Rate. ΔR: Median absolute change in reach angle after rewarded trials.

Overall, the regression results across both tasks exhibited a pattern similar, with greater RWM predicting improved performance on the reaching task and greater SWM predicting smaller changes in reach angle after rewarded trials. The weak relationships found between genetic variables and performance measures in the Acquire task (DRD2-Success rate and DARPP32-Explicit retention) were not replicated in the Preserve task, questioning the reliability of these relationships.

Furthermore, we analysed the data using group lasso (Boyd, 2010; Yuan and Lin, 2006) regression in order to check for the possibility that our analysis was insensitive to categorical predictors (the genetic variables). The group lasso is an extension to lasso regression in which predictor variables can be assigned to groups. Although each member of a group can be

assigned a different β, the group lasso applies the regularisation penalty to all members of the group, leading to the removal of all members of the group from the model at the same value of λ. We employed reference dummy variable coding for each genetic variable and treated the dummy variables representing each SNP as a group for the purposes of the group lasso; this ensures that the dummy variables representing each genetic factor are removed from the regression at the same time. The results of the group lasso analysis replicate those of the standard lasso and furthermore no genetic predictors were found for any outcome variable in either task. The results obtained for both tasks via the lasso regression methods are similar to those obtained using a stepwise regression procedure. All data and code are available online, including the procedures, results, and significance tests of the lasso and stepwise regression analysis.

Relationships between predictors

In the full sample (n=241), we assessed the relationship between the predictor variables. Despite the collinearity of the variables being within recommended values for use in regression (See methods section), we did find significant relationships between all three WM tasks. VWM and SWM were the most closely correlated ($r=0.393$, $p=3.153\times10^{-10}$), followed by SWM and RWM ($r=0.384$, $p=7.491\times10^{-10}$), and finally RWM and VWM ($r=0.189$, $p=0.003$). When examining the relationships between genetics and WM tasks, only one relationship was significant (DRD2 and SWM, $F(236,2)=3.927$, $p=0.021$). However, this relationship did not survive correction for multiple comparisons.

Partial Correlation Analysis

In order to understand if the RWM and SWM measures have separable effects on the outcome measures considered here, we performed a partial correlation analysis examining the

relationships between RWM, SWM, and success rate in both tasks. After controlling for the effect of RWM, SWM remained significantly correlated with success in both tasks ($r=0.343$, $p=0.005$ Preserve, $r=0.488$, $p=6.823\times10^{-6}$ Acquire). However, the partial correlation between RWM and success rate was not significant for either task, indicating that even in the Preserve task SWM plays a dominant role in determining success rate.

Exploratory analysis

As a relationship exists between SWM and ΔR in both the Acquire and Preserve paradigms, we ran exploratory regressions to assess the relationship between ΔR and success rate across both tasks. Since ΔR and success rate are conceptually strongly related variables, and measuring on the same data set would render them non-independent, we split each individual's reaching data into two sections and assessed whether ΔR or ΔP in the first section could reliably predict success rate in the second (see methods for details). Although we found no predictors of ΔP in our primary analysis, results here in combination with previous work (Holland et al., 2018) has demonstrated a link between ΔP and task success, with a greater ΔP indicative of greater success. Therefore, we also performed the same analysis for ΔP.

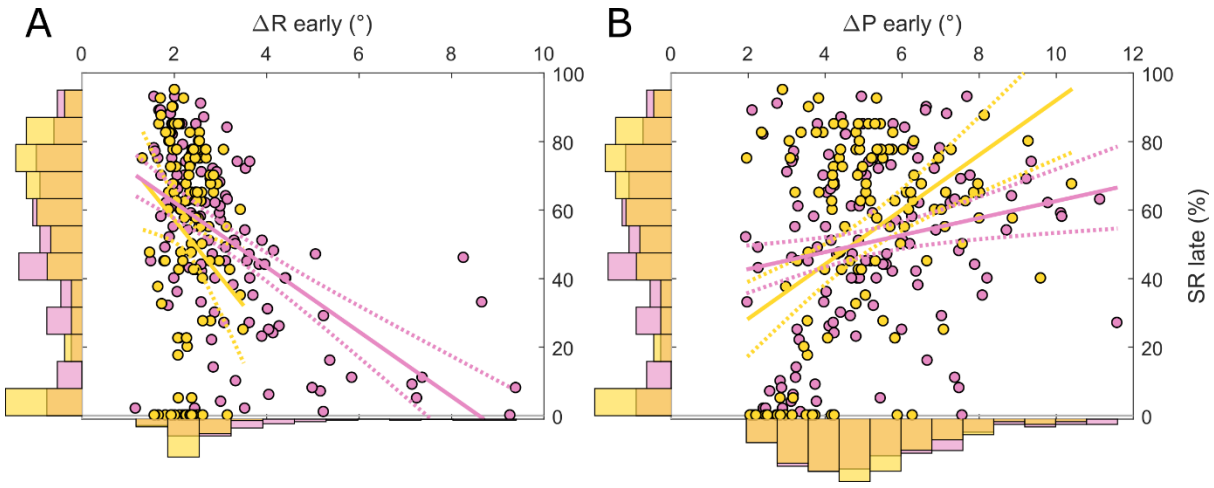


Figure 10: Slice plots showing regression results for prediction of late success rate (SR) by changes in reach angle following rewarded (A) and unrewarded (B) trials during the early learning period. The central axis of each panel displays the individual data from the Acquire (yellow) and Preserve (pink) task, a histogram displaying the distribution of the data in each dimension is presented on the corresponding axis. Solid lines represent the prediction of the regression model when the other predictor is held at its mean value. SR: Success Rate. Median absolute change in reach angle after rewarded (ΔR) and unrewarded (ΔP) trials.

In the Acquire task, ΔR and ΔP in the first section of learning trials predicted success rate in the final twenty trials, though ΔP appeared as the strongest predictor (ΔR: β=-0.274, p=0.015; ΔP: β=0.581, p=3.89x10^{-6}; Figure 10a,b, yellow; Table 2). Similarly, for the Preserve task, ΔR and ΔP in the first half of asymptote trials predicted success rate in the second half (ΔR: β=-0.750, p=1.07x10^{-12}; ΔP: β=0.229, p=0.007; Figure 10a,b, pink; Table 2). In both tasks, the directions of these relationships were opposite; greater success rate was predicted by smaller ΔR and greater ΔP. In summary, we found that for both tasks the magnitude of changes in behaviour in response to rewarded and unrewarded trials early in learning were strongly predictive of future task success across both the Acquire and Preserve tasks.

		ΔR	ΔP	Model
Acquire	β	-0.274	0.581	
				F(115,2)=11.9
	SE	0.111	0.120	
				$p=2.09\times10^{-5}$
	p	0.015	3.89×10^{-6}	
Preserve	β	-0.750	0.229	
				F(112,2)=35.3
	SE	0.093	0.084	
				$p=1.28\times10^{-12}$
	p	1.07×10^{-12}	0.007	

Table 2: Regression results for split data for both the Acquire and Preserve tasks. Ordinary least squares linear regressions were performed with both ΔR and ΔP included as predictors. The regression coefficient, standard error and p value for each predictor are reported along with the significance of the comparison between the model and an intercept only model. In both tasks there is an opposing relationship between ΔR and ΔP and success rate, with smaller changes after rewarded trials and larger changes after unrewarded trials predictive of success. SR: Success Rate. Median absolute change in reach angle after rewarded (ΔR) and unrewarded (ΔP) trials.

Mediation analysis

Finally, to test whether the effect observed between SWM and SR was explained by an indirect effect through ΔR, we performed an exploratory mediation analysis on both tasks. For both the Acquire and Preserve tasks, the results indicate a significant proportion (median $p=7.10\text{x}10^{-4}$ and p=0.04 respectively) of the relationship between SWM and SR can be explained by a mediation from SWM via ΔR to SR (Figure 11). However, in the case of the Acquire task (Figure11a), a significant relationship between SWM and SR also remained, indicating that not all of the effect of SWM on SR could be explained by the indirect pathway. Of note, in the Preserve task (Figure 11b) the SWM-ΔR relationship was weaker and was not significant on every repetition, occasionally leading to an insignificant mediation effect, despite the median p-value indicating an effect when considering all repetitions. We also examined an alternative possibility to the hypothesized model in which relationship between SWM and ΔR is mediated by SR. We found that 31.20% of the total effect is mediated in the Acquire task using the

hypothesized model, in contrast to only 0.17% in the alternative model. Similarly, in the Preserve task the hypothesized model displayed a substantially larger mediation effect (44.77%) than the alternative model (5.02%). These results support the application of the hypothesized model.

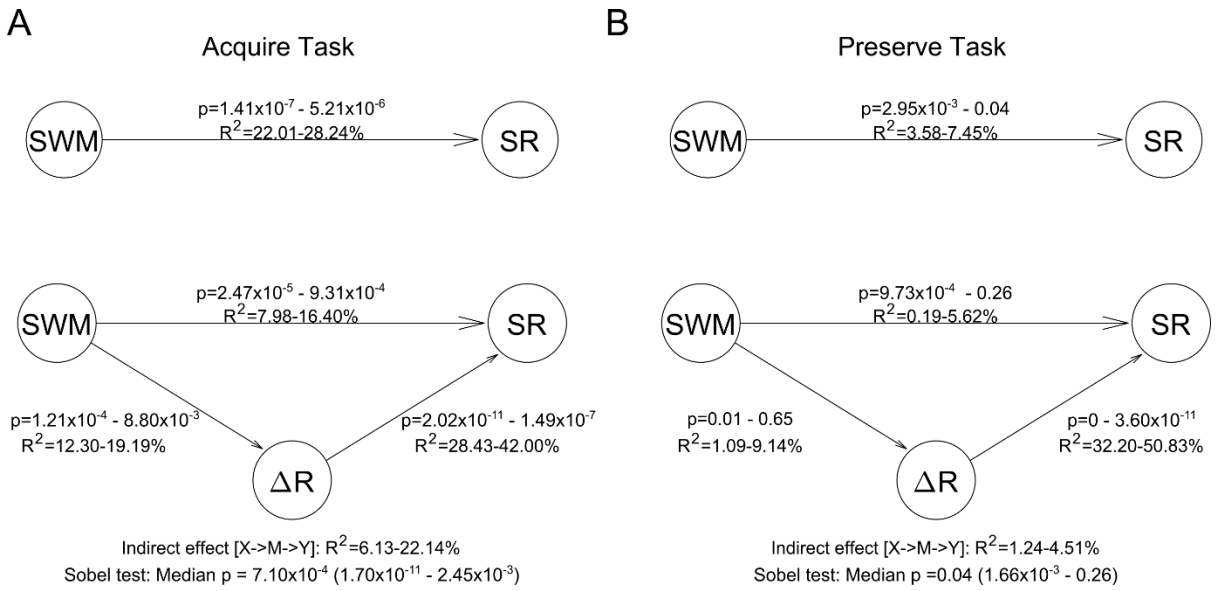


Figure 11: Mediation Analysis for both the Acquire (A) and Preserve (B) tasks. The numbers associated with each arrow display the 95% confidence intervals for each of the relationships (R^2 and p-values) across the 1000 repetitions. Below the figure, the results of the Sobel test are displayed indicating the amount of variance explained by the indirect pathway and the 95% confidence intervals and median p-value. SR: Success Rate. ΔR: Median absolute change in reach angle after rewarded trials.

Discussion

In this study, we sought to identify if genetic background or specific domains of WMC could explain the variability observed in performance levels during reward-based motor learning tasks. We found that RWM and SWM predicted different aspects of the Acquire and Preserve tasks, whereas VWM only related to one performance measure (ΔP), but not consistently across tasks. Specifically, RWM predicted the explicit component of retention in the Acquire task and

success rate in the Preserve task, whereas SWM predicted success rate in the Acquire task and the late period of the Preserve task. Furthermore SWM negatively predicted ΔR in both tasks. Conversely, allelic variations of the three dopamine-related genes (DRD2, COMT and DARPP32) did not consistently predict any behavioural variables across both tasks. This suggests that SWM predicts a participant's capacity to reproduce a rewarded motor action, while RWM predicts a participant's ability to express an explicit strategy when making large behavioural adjustments. Therefore, we conclude that WMC plays a pivotal role in determining individual ability in reward-based motor learning.

Recently, Wong et al. (2019) described a positive relationship between SWM and the development of explicit strategies in visuomotor adaptation, complementing previous reports (Anguera et al., 2012; Christou et al., 2016; Vandevoorde and Orban de Xivry, 2019). However, in contrast to the current findings the previous experiments employed relatively small sample sizes, which may render correlations unreliable. The large group sizes employed here, and the confirmation of relationships across two tasks, provides strong evidence that these relationships are robust, replicable, and extend from visuomotor adaptation to reward-based motor learning.

An interesting dichotomy was the reliance on SWM and RWM for the Acquire and Preserve task, respectively. While the Preserve task required the maintenance of a large, abrupt behavioural change, the Acquire task required the gradual adjustment of behaviour considering the outcomes of recent trials. Therefore, RWM may underscore one's capacity to express a large correction consistently over trials with binary feedback, whereas SWM reflects one's capacity to maintain a memory of previously rewarded actions and adjust behaviour accordingly. Accordingly, McDougle and Taylor (2019) demonstrated a mental rotation process is employed in countering a visuomotor rotation, and Sidarta et al. (2018) reported that higher SWM is associated with reduced movement variability in a reward-based motor learning task. Here, the magnitude of ΔR was negatively related to SWM but not RWM in both tasks,

suggesting high SWM enables the maintenance of rewarding actions. Additionally, explicit retention, an element of the Acquire task requiring a large, sudden change in reach direction, was predicted by RWM rather than SWM. Notably, RWM and SWM were often selected as predictors simultaneously. The overlapping but distinct pattern of relationships between RWM, SWM, and outcome measures considered here supports the view that they share substrates but have different patterns of dependency on executive functions (Miyake et al., 2001).

A notable feature of the Preserve task is the 'spiking' behaviour observed immediately following 'refresher' trials, suggesting a central role of refresher trials in binary feedback-based performance when included (Codol et al., 2018; Shmuelof et al., 2012). The transient nature of this decrease in error demonstrates this is insufficient to promote generalisation to binary feedback trials, at least in unsuccessful participants. It remains an open question whether superior performance of successful participants was partly due to a capacity to generalise information from 'refresher' trials. McDougle and Taylor (2019) suggest that two separate strategies are employed in visuomotor adaptation: response-caching and mental rotation. The balance between the two strategies is a function of task demands. The relationships between RWM and SWM to success rate in the Preserve and Acquire tasks respectively may reflect a different balance of the use of these strategies. Visual feedback in 'refresher' trials in the Preserve task may engage mental rotation processes, whereas the slow updating of behaviour in the Acquire task engages the response-caching memory system. This would imply that response-caching is associated with SWM.

Surprisingly, although ΔP was a strong predictor of success in both tasks, it was not consistently predicted by any variable across both tasks. The lack of a consistent predictor of ΔP was unexpected given the importance of errors for the induction of structural learning in reinforcement learning (Daw et al., 2011; Manley et al., 2014; Sutton and Barto, 1998) and reward-based motor learning (Maxwell et al., 2001; Sidarta et al., 2018).

If RWM is important for explicit control and the main element predicting success in the Preserve task, it is worth considering whether gradual designs (as in the Acquire task) are more suitable to engage implicit reinforcement learning, at least initially. However, the Acquire task still bears a strong explicit component (Holland et al., 2018). How can these two views be reconciled? In reward-based motor learning tasks, it is observed that participants begin to reflect upon task structure and develop strategies upon encountering negative outcomes (Leow et al., 2016; Loonis et al., 2017; Maxwell et al., 2001), which occurs nearly immediately in the Preserve task after the introduction of binary feedback, due to a lack of generalisation of cerebellar memory (Codol et al., 2018). In contrast, in the Acquire task, participants experience an early learning phase with mainly rewarding outcomes, possibly suppressing development of explicit control and allowing for this early window of implicit reward-based learning. Other studies have demonstrated that minor adjustments in reach direction under reward-based feedback can occur, though none has assessed their explicitness directly in the very early stages (Izawa and Shadmehr, 2011; Pekny et al., 2015; Therrien et al., 2016). Notably, Izawa and Shadmehr, (2011) observed that after 8° shifts of a similarly-sized reward region, participants indeed noticed the perturbation, but awareness was not assessed for smaller shifts.

In Holland et al., (2018), the addition of a RWM-like dual-task was very effective in preventing explicit control, leading to participants invariably failing at the reaching task. Therefore, it may seem surprising that RWM does not predict success rate in the Acquire task. A possible explanation is that RWM and SWM share the same memory buffer (Anguera et al., 2010; Beschin et al., 2005; Cohen et al., 1996; Jordan et al., 2001; Suchan et al., 2006). Similarly, in force-field adaptation the early component of adaptation – considered as bearing a strong explicit element – is selectively disrupted with a VWM dual-task (Keisler and Shadmehr, 2010). However, we found no consistent relationship with VWM across our reward-based motor tasks.

It may be possible that reward-based motor performance relies more on spatial instances of WM as opposed to tasks such as force-field adaptation.

The absence of DA-related genetic relationships with behaviour is a surprising result as a substantial body of literature points to a relationship between dopamine and performance in reward-based tasks, including those with motor components (Deserno et al., 2015; Doll et al., 2016; Frank et al., 2007, 2009; Gershman and Schoenbaum, 2017; Izawa and Shadmehr, 2011; Nakahara and Hikosaka, 2012; Pekny et al., 2015; Therrien et al., 2016). There is a growing appreciation of the links between decision-making and motor learning (Chen et al., 2017, 2018; Haith and Krakauer, 2013). However, the results presented here suggest that genetic predictors of exploration and exploitation in decision-making tasks are not also predictive of similar behaviours in reward-based motor learning.

Our sample sizes were defined *a priori* for 90% power based on previous work (Doll et al., 2016; Frank et al., 2009; see pre-registrations), and are unlikely to be underpowered. Another possibility is that we employed the wrong variables to assess behaviour. However, given the informative and coherent relationships between WM and motor learning, it could be that the SNPs we selected do not meaningfully relate to performance in reward-based motor tasks compared to WM. A similar claim was made in the decision-making literature (Collins and Frank, 2012). In line with this, a recent study showed that DA pharmacological manipulation did not alter reward effects in a visuomotor adaptation task (Quattrocchi et al., 2018). However, previous work has shown that Parkinson's disease patients show impaired reward-based motor performance (Pekny et al., 2015). It is possible that genetic variations may simply not impact reward-based motor learning significantly, especially compared to the wide depletion of dopaminergic neurons in Parkinson's disease. It is also important to note that while we refer to both of our tasks as reward-based motor learning, they are both in essence visuomotor rotation

paradigms. In future it is important to investigate if these findings extend to more complex reward-based motor learning paradigms.

In summary, despite employing two distinct tasks and an independent participant pool on different devices, we find strikingly similar results in reward-based motor learning. While SWM strongly predicted a participant's capacity to reproduce successful motor actions, RWM predicted a participant's ability to express an explicit strategy when required to make large behavioural adjustments. Surprisingly, no dopamine-related genotypes predicted performance. Therefore, WMC plays a pivotal role in determining individual ability in reward-based motor learning. This could have important implications when using reward-based feedback in applied settings as only a subset of the population may benefit.

References

Anguera, J.A., Reuter-Lorenz, P.A., Willingham, D.T., and Seidler, R.D. (2010). Contributions of spatial working memory to visuomotor learning. Journal of Cognitive Neuroscience *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. Behavioural Brain Research *228*, 107–115.

Baron, R.M., and Kenny, D.A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. Journal of Personality and Social Psychology *51*, 1173–1182.

916 Beschin, N., Denis, M., Logie, R.H., and Della Sala, S. (2005). Dissociating mental
917 transformations and visuo-spatial storage in working memory: Evidence from
918 representational neglect. Memory *13*, 430–434.

919 Bond, K.M., and Taylor, J.A. (2015). Flexible explicit but rigid implicit learning in a
920 visuomotor adaptation task. Journal of Neurophysiology *113*, 3836–3849.

921 Boyd, S. (2010). Distributed Optimization and Statistical Learning via the Alternating
922 Direction Method of Multipliers. FNT in Machine Learning *3*, 1–122.

923 Brainard, D.H. (1997). The Psychophysics Toolbox. Spat Vis *10*, 433–436.

924 Buszard, T., and Masters, R.S.W. (2018). Adapting, correcting and sequencing movements:
925 does working-memory capacity play a role? International Review of Sport and Exercise
926 Psychology *11*, 258–278.

927 Cashaback, J.G.A., McGregor, H.R., Mohatarem, A., and Gribble, P.L. (2017). Dissociating
928 error-based and reinforcement-based loss functions during sensorimotor learning. PLoS
929 Comput Biol *13*, e1005623.

930 Cashaback, J.G.A., Lao, C.K., Palidis, D.J., Coltman, S.K., McGregor, H.R., and Gribble,
931 P.L. (2019). The gradient of the reinforcement landscape influences sensorimotor learning.
932 PLoS Comput Biol *15*, e1006839.

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

935 Chen, X., Holland, P., and Galea, J.M. (2018). The effects of reward and punishment on
936 motor skill learning. Current Opinion in Behavioral Sciences *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. Scientific Reports *6*.

Codol, O., Holland, P.J., and Galea, J.M. (2018). The relationship between reinforcement and explicit control during visuomotor adaptation. Scientific Reports *8*.

Cohen, M.S., Kosslyn, S.M., Breiter, H.C., DiGirolamo, G.J., Thompson, W.L., Anderson, A.K., Bookheimer, S.Y., Rosen, B.R., and Belliveau, J.W. (1996). Changes in cortical activity during mental rotation A mapping study using functional MRI. Brain *119*, 89–100.

Collins, A.G.E., and Frank, M.J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis: Working memory in reinforcement learning. European Journal of Neuroscience *35*, 1024–1035.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. Neuron *69*, 1204–1215.

Deserno, L., Huys, Q.J.M., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A.A., Dolan, R.J., Heinz, A., and Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proceedings of the National Academy of Sciences *112*, 1595–1600.

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

Doll, B.B., Duncan, K.D., Simon, D.A., Shohamy, D., and Daw, N.D. (2015). Model-based choices involve prospective neural activity. Nature Neuroscience *18*, 767–772.

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. Journal of Neuroscience *36*, 1211–1222.

Egan, M.F., Goldberg, T.E., Kolachana, B.S., Callicott, J.H., Mazzanti, C.M., Straub, R.E., Goldman, D., and Weinberger, D.R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proceedings of the National Academy of Sciences *98*, 6917–6922.

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. Proceedings of the National Academy of Sciences *104*, 16311–16316.

Frank, M.J., Doll, B.B., Oas-Terpstra, J., and Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nature Neuroscience *12*, 1062–1068.

Gershman, S.J., and Schoenbaum, G. (2017). Rethinking dopamine prediction errors. BioRxiv 239731.

Goldberg, T.E., Egan, M.F., Gscheidle, T., Coppola, R., Weickert, T., Kolachana, B.S., Goldman, D., and Weinberger, D.R. (2003). Executive Subprocesses in Working Memory: Relationship to Catechol-O-methyltransferase Val158Met Genotype and Schizophrenia. Arch Gen Psychiatry *60*, 889–896.

Haith, A.M., and Krakauer, J.W. (2013). Model-Based and Model-Free Mechanisms of Human Motor Learning. In Progress in Motor Control, M.J. Richardson, M.A. Riley, and K. Shockley, eds. (New York, NY: Springer New York), pp. 1–21.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). Chapter 6. Statistical Inference: The Bootstrap. In Statistical Learning with Sparsity: The Lasso and Generalizations, (Boca Raton: Chapman and Hall/CRC), p.

Holland, P.J., Codol, O., and Galea, J.M. (2018). The contribution of explicit processes to reinforcement-based motor learning. Journal of Neurophysiology.

Huang, V.S., Haith, A., Mazzoni, P., and Krakauer, J.W. (2011). Rethinking Motor Learning and Savings in Adaptation Paradigms: Model-Free Memory for Successful Actions Combines with Internal Models. Neuron *70*, 787–801.

Izawa, J., and Shadmehr, R. (2011). Learning from Sensory and Reward Prediction Errors during Motor Adaptation. PLoS Computational Biology *7*, e1002012.

Jordan, K., Heinze, H.-J., Lutz, K., Kanowski, M., and Jäncke, L. (2001). Cortical Activations during the Mental Rotation of Different Visual Objects. NeuroImage *13*, 143–152.

Keisler, A., and Shadmehr, R. (2010). A Shared Resource between Declarative Memory and Motor Memory. Journal of Neuroscience *30*, 14817–14823.

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

Leow, L.-A., de Rugy, A., Marinovic, W., Riek, S., and Carroll, T.J. (2016). Savings for visuomotor adaptation require prior history of error, not prior repetition of successful actions. Journal of Neurophysiology *116*, 1603–1614.

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.

Mackinnon, D.P., and Dwyer, J.H. (1993). Estimating Mediated Effects in Prevention Studies. Eval Rev *17*, 144–158.

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. The Quarterly Journal of Experimental Psychology A *54*, 1049–1068.

McDougle, S.D., and Taylor, J.A. (2019). Dissociable cognitive strategies for sensorimotor learning. Nature Communications *10*.

McNab, F., and Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. Nature Neuroscience *11*, 103–107.

Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., and Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. J Exp Psychol Gen *130*, 621–640.

Modchalingam, S., Vachon, C.M., 't Hart, B.M., and Henriques, D.Y.P. (2019). The effects of awareness of the perturbation during motor adaptation on hand localization. PLOS ONE *14*, 20.

1021 Morehead, J.R., Qasim, S.E., Crossley, M.J., and Ivry, R. (2015). Savings upon Re-Aiming in
1022 Visuomotor Adaptation. Journal of Neuroscience *35*, 14386–14396.

1023 Morehead, J.R., Taylor, J.A., Parvin, D., and Ivry, R.B. (2017). Characteristics of Implicit
1024 Sensorimotor Adaptation Revealed by Task-irrelevant Clamped Feedback. Journal of
1025 Cognitive Neuroscience 1–14.

1026 Nakahara, H., and Hikosaka, O. (2012). Learning to represent reward structure: A key to
1027 adapting to complex environments. Neuroscience Research *74*, 177–183.

1028 Pearson-Fuhrhop, K.M., Minton, B., Acevedo, D., Shahbaba, B., and Cramer, S.C. (2013).
1029 Genetic Variation in the Human Brain Dopamine System Influences Motor Learning and Its
1030 Modulation by L-Dopa. PLoS ONE *8*, e61197.

1031 Pekny, S.E., Izawa, J., and Shadmehr, R. (2015). Reward-Dependent Modulation of
1032 Movement Variability. Journal of Neuroscience *35*, 4015–4024.

1033 Peters, M., and Battista, C. (2008). Applications of mental rotation figures of the Shepard and
1034 Metzler type and description of a mental rotation stimulus library. Brain and Cognition *66*,
1035 260–264.

1036 Quattrocchi, G., Monaco, J., Ho, A., Irmen, F., Strube, W., Ruge, D., Bestmann, S., and
1037 Galea, J.M. (2018). Pharmacological Dopamine Manipulation Does Not Alter Reward-Based
1038 Improvements in Memory Retention during a Visuomotor Adaptation Task. Eneuro *5*,
1039 ENEURO.0453-17.2018.

1040 Saijo, N., and Gomi, H. (2010). Multiple Motor Learning Strategies in Visuomotor Rotation.
1041 PLoS ONE *5*, e9399.

1042 Shepard, R., N., and Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects.
1043 Science *171*, 701–703.

1044 Shmuelof, L., Huang, V.S., Haith, A.M., Delnicki, R.J., Mazzoni, P., and Krakauer, J.W.
1045 (2012). Overcoming Motor "Forgetting" Through Reinforcement Of Learned Actions.
1046 Journal of Neuroscience *32*, 14617–14621a.

1047 Sidarta, A., van Vugt, F., and Ostry, D.J. (2018). Somatosensory Working Memory in Human
1048 Reinforcement-Based Motor Learning. Journal of Neurophysiology 41.

1049 Sobel, M.E. (1986). Some New Results on Indirect Effects and Their Standard Errors in
1050 Covariance Structure Models. Sociological Methodology *16*, 159.

1051 Suchan, B., Botko, R., Gizewski, E., Forsting, M., and Daum, I. (2006). Neural substrates of
1052 manipulation in visuospatial working memory. Neuroscience *139*, 351–357.

1053 Sutton, R.S., and Barto, A. (1998). Reinforcement Learning: An Introduction (Cambridge,
1054 Mass: A Bradford Book).

1055 Taylor, J.A., and Ivry, R.B. (2011). Flexible Cognitive Strategies during Motor Learning.
1056 PLoS Computational Biology *7*, e1001096.

1057 Taylor, J.A., and Ivry, R.B. (2014). Cerebellar and Prefrontal Cortex Contributions to
1058 Adaptation, Strategies, and Reinforcement Learning. In Progress in Brain Research,
1059 (Elsevier), pp. 217–253.

1060 The 1000 Genomes Project Consortium, Gibbs, R.A., Boerwinkle, E., Doddapaneni, H., Han,
1061 Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., et al. (2015). A global reference
1062 for human genetic variation. Nature *526*, 68–74.

1063 Therrien, A.S., Wolpert, D.M., and Bastian, A.J. (2016). Effective reinforcement learning
1064 following cerebellar damage requires a balance between exploration and motor noise. Brain
1065 *139*, 101–114.

1066 Therrien, A.S., Wolpert, D.M., and Bastian, A.J. (2018). Increasing Motor Noise Impairs
1067 Reinforcement Learning in Healthy Individuals. Eneuro *5*, ENEURO.0050-18.2018.

1068 Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal
1069 Statistical Society. Series B (Methodological) *58*, 267–288.

1070 Tseng, Y. -w., Diedrichsen, J., Krakauer, J.W., Shadmehr, R., and Bastian, A.J. (2007).
1071 Sensory Prediction Errors Drive Cerebellum-Dependent Adaptation of Reaching. Journal of
1072 Neurophysiology *98*, 54–62.

1073 Vandevoorde, K., and Orban de Xivry, J.-J. (2019). Internal model recalibration does not
1074 deteriorate with age while motor adaptation does. Neurobiology of Aging *80*, 138–153.

1075 Vindras, P., Desmurget, M., Prablanc, C., and Viviani, P. (1998). Pointing Errors Reflect
1076 Biases in the Perception of the Initial Hand Position. Journal of Neurophysiology *79*, 3290–
1077 3294.

1078 Werner, S., van Aken, B.C., Hulst, T., Frens, M.A., van der Geest, J.N., Strüder, H.K., and
1079 Donchin, O. (2015). Awareness of Sensorimotor Adaptation to Visual Rotations of Different
1080 Size. PLOS ONE *10*, e0123321.

1081 Wong, A.L., Marvel, C.L., Taylor, J.A., and Krakauer, J.W. (2019). Can patients with
1082 cerebellar disease switch learning mechanisms to reduce their adaptation deficits? Brain *142*,
1083 662–673.

Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. J Royal Statistical Soc B *68*, 49–67.