J Neurophysiol 107: 1111–1122, 2012.
First published November 30, 2011; doi:10.1152/jn.00635.2011.

Facilitation of learning induced by both random and gradual visuomotor task variation

Edward J. A. Turnham, Daniel A. Braun, and Daniel M. Wolpert
Department of Engineering, University of Cambridge, Cambridge, United Kingdom

Submitted 11 July 2011; accepted in final form 22 November 2011

Turnham EJ, Braun DA, Wolpert DM. Facilitation of learning induced by both random and gradual visuomotor task variation. *J Neurophysiol* 107: 1111–1122, 2012. First published November 30, 2011; doi:10.1152/jn.00635.2011.—Motor task variation has been shown to be a key ingredient in skill transfer, retention, and structural learning. However, many studies only compare training of randomly varying tasks to either blocked or null training, and it is not clear how experiencing different nonrandom temporal orderings of tasks might affect the learning process. Here we study learning in human subjects who experience the same set of visuomotor rotations, evenly spaced between $-60°$ and $+60°$, either in a random order or in an order in which the rotation angle changed gradually. We compared subsequent learning of three test blocks of $+30° \rightarrow -30° \rightarrow +30°$ rotations. The groups that underwent either random or gradual training showed significant ($P < 0.01$) facilitation of learning in the test blocks compared with a control group who had not experienced any visuomotor rotations before. We also found that movement initiation times in the random group during the test blocks were significantly ($P < 0.05$) lower than for the gradual or the control group. When we fit a state-space model with fast and slow learning processes to our data, we found that the differences in performance in the test block were consistent with the gradual or random task variation changing the learning and retention rates of only the fast learning process. Such adaptation of learning rates may be a key feature of ongoing meta-learning processes. Our results therefore suggest that both gradual and random task variation can induce meta-learning and that random learning has an advantage in terms of shorter initiation times, suggesting less reliance on cognitive processes.

sensorimotor learning; visuomotor rotation; contextual interference; learning to learn; structural learning

WHEN HUMANS OR OTHER ANIMALS practice one or more tasks they are often subsequently able to learn related tasks more rapidly. This phenomenon of "transfer learning" or "learning to learn" has been demonstrated for cognitive tasks in animals (Harlow 1949; Warren 1965; Mackintosh and Little 1969; Schrier 1984; Langbein et al. 2007) and humans (Duncan 1960; Halford et al. 1998; Tenenbaum and Griffiths 2001; Kemp and Tenenbaum 2009) and more recently for sensorimotor control tasks in humans (Welch et al. 1993; Roller et al. 2001; Seidler 2004, 2007; Cohen et al. 2005; Seidler 2007; Braun et al. 2009a,b; Mulavara et al. 2009).

Many of these studies used a random training sequence of tasks. Such frequent changes in training task can contribute to greater retention and better transfer, even though performance during practice might be adversely affected-termed "contextual interference" (Battig 1966, 1972, 1979; Shea and Morgan 1979; Magill and Hall 1990; Brady 2008; Schweighofer et al. 2011). For example, Shea and Morgan (1979) trained subjects on different motor sequences, either with each sequence experienced repeatedly within a block of trials or with the sequence randomly changing from trial to trial. The random group experienced larger errors during training but performed better in a retention test and generalized better to more complex motor sequences. Here we ask whether random ordering of the training tasks is crucial for the induction of meta-learning (i.e., adaptation of the learning process itself, such that learning of a novel task proceeds differently before and after meta-learning) that manifests as faster learning of novel tasks in the above studies or whether the same tasks ordered in nonrandom ways could induce similar meta-learning. The paradigm of visuomotor rotation allows us to present tasks in a randomly or gradually varying order, since the rotation angle is a continuous variable, whereas movement sequences such as those employed by Shea and Morgan (1979) can only be varied in a discrete parameter space.

While in some of the above studies a generic speed up in motor learning was observed after practice (e.g. Seidler 2004), other studies reported facilitation effects that were more specific for tasks that shared similar structural features (See Braun et al. 2010 for review). Braun et al. (2009b) proposed that the latter category of facilitation effects was due to subjects learning the structure that was common to the tasks ("structural learning," a form of meta-learning). Once this structure is learned, adaptation to a task conforming to that structure requires the learning of fewer parameters than if the structure were not known and so is faster. Transitioning between two tasks conforming to the structure leads to less interference, because fewer parameters have to be changed. Braun et al. showed that randomly ordered exposure to 100 rotation angles between $-90°$ and $+90°$ increased learning rates on subsequent rotations. Further, they showed that random training with examples from a structure led to preferential exploration along that structure when confronted with a new task. This demonstrated that much of the increase in learning rates from random rotation training is due to structural learning.

Smith et al. (2006) developed a state-space model to describe the time course of visuomotor and force field learning. In this model, there are one fast and one slow learning processes that contribute to the total motor output. The fast process both learns and forgets quickly, while the slow process learns and forgets more slowly. This two rate model can in some paradigms explain such diverse phenomena as savings, anterograde interference, spontaneous recovery, and rapid unlearning. However, this model cannot capture more complex processes of meta-learning that are likely to require changes to the learning and forgetting rates (Zarahn et al. 2008), and

Address for reprint requests and other correspondence: E. Turnham, Dept. of Engineering, Trumpington St., Cambridge CB2 1PZ, United Kingdom (e-mail: ejat3@cam.ac.uk).

forgetting rates have been shown to be modulated by prior experience (Huang and Shadmehr 2009). It is an open question whether meta-learning processes, such as learning-to-learn or structural learning, can be understood within such a two rate learning model in which the parameters of the model might adapt over time. The current study, therefore, both compares random vs. nonrandom task variation in facilitating new learning and asks whether such facilitation can be understood as the learning of parameters within the two rate state-space model framework.

In a planar reaching task, we exposed three groups of subjects to a training phase in which they experienced visuomotor rotations than changed in angle every 16 trials. The 100 rotation angles experienced in the training phase were evenly spaced between −60° and +60°. In the "random" group, the order of the rotations was random whereas in the two "gradual" groups the order was selected so that the rotation angle, starting and ending at 0°, changed gradually in either a positive or negative saw-tooth pattern. A final "control" group made reaches with veridical feedback, that is without a rotation. All subjects then received a washout session with veridical feedback and were then exposed to three blocks of trials: a +30° rotation block, followed by a −30° block and a +30° block. These three test blocks were used to measure adaptation rate and the extent of interference between the learning of opposite rotations as well as the time taken to initiate movements. This allows a comparison of performance between the groups who had experienced different temporal orders of learning during the training phase. Adaptation rate is defined as the rate at which reaching directions, relative to target directions, change during learning; it can be expressed in absolute terms (e.g., degrees per trial) or in relative terms (e.g., as a proportion of the initial reach direction error, per trial). Interference occurs when performance on learning or relearning of a rotation is degraded by the preceding presentation of another rotation; it can manifest as an aftereffect of adaptation to the interfering rotation (meaning that compensatory behavior suitable for the interfering rotation continues during exposure to the new rotation for which it is inappropriate) or as a reduction in adaptation rate from what the rate would have been if the interfering rotation had not been presented. A recent study (Fernandez-Ruiz et al. 2011) suggests that slower initiation of movements is a characteristic of cognitive processes that quickly reduce directional error in rotation learning. We examine initiation times to look for a correlation with learning rate suggestive of cognitive strategies. The "contextual interference" hypothesis predicts that the increased difficulty of random training will lead to better test performance than for the gradual group. However, it would be expected that each training task would be more completely learned if they were presented in a gradually varying order. By comparing the performance of the two groups, our study also investigates whether structural learning effects are driven mainly by errors (high in random) or driven by the sequence of visited adaptation states (more broadly distributed in gradual).

METHODS

Experimental procedure. Twenty-seven subjects, naïve to the purpose of the experiment, participated after giving written informed consent in accordance with the requirements of the local ethics committee that approved the methods.

Subjects used their right hand to grasp the handle of a vBOT manipulandum, a custom-built planar robotic interface that allows movement in the horizontal plane and can measure the position of its handle with a 1-kHz update rate. Using a mirror-monitor system, subjects were prevented from seeing their own arm or the robot handle, and we overlaid virtual visual feedback of the hand position (a 0.5-cm radius cursor, which was visible throughout all trials) and targets into the plane of the movement. For full details of the robotic and virtual reality setup, see Howard et al. (2009).

Each trial consisted of an out-and-back reaching movement that started and ended at a starting circle (1 cm radius), ~30 cm in front of the chest in the midline. To start a trial, subjects had to place the hand cursor inside the starting circle for 0.5 s. A circular target (0.5-cm radius) then appeared 8 cm from the starting circle in one of eight directions spaced 45° apart. A cycle of movements consisted of eight trials, in which each target was presented once in a pseudorandom order. Subjects were instructed to make a fast movement to the target and back in a straight line and not to try to correct for errors within the movement. Subjects were required to start the movement such that the cursor left the starting circle within 0.4 s of target appearance; otherwise the trial was terminated and an error signaled. If the target was hit (defined as any part of the cursor overlapping any part of the target) within 0.2 s of the cursor leaving the start circle, the subject was rewarded with the explosion of the target. Otherwise the target changed color to indicate failure and remained visible for up to a further 1.3 s, until the cursor had reentered the starting circle. The next trial could then begin, subject to a minimum 1.5-s gap between the appearance of consecutive targets. Depending on the condition and trial, the cursor either veridically represented the hand position or represented the hand position rotated around the starting circle by a rotation angle. The rotation angle was always fixed within a cycle and changes in the rotation angle only occurred while subjects were in the starting circle, at the time of target appearance.

Subjects were assigned to four possible groups (Fig. 1 shows the paradigm and rotation orders for the groups). All groups initially performed five cycles with veridical feedback. They then performed a training phase of 200 cycles in which the nature of any visuomotor rotations depended on the group. All subjects then performed a washout session of five cycles with veridical feedback to reestablish a baseline. They were then exposed to a test phase consisting of seven cycles of +30°, followed by seven cycles of −30°, and finally another seven cycles of +30° rotations. That is they experienced three test

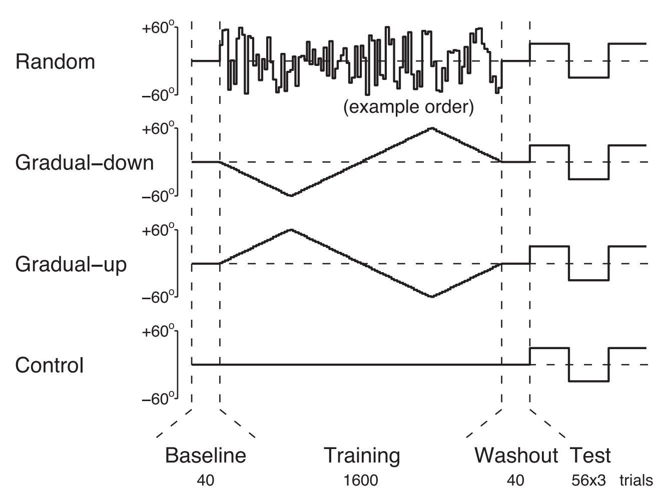

Fig. 1. Sequences of rotation angles for the four experimental groups. All groups experienced 5 cycles (with 8 trials per cycle) of veridical reaching, followed by 200 cycles of training with rotations (random and gradual groups) or more veridical reaching (null group), then a washout of 5 veridical cycles, and then 7 cycles each of +30°, −30°, and +30° test rotations.

blocks of a clockwise, anticlockwise, and then clockwise visuomotor rotation.

The control group (7 subjects) experienced veridical feedback during the entire training phase, meaning that the cursor tracked the actual hand position. The other three groups experienced visuomotor rotations during the training phase that changed every two cycles. Across the training phase, these subjects experienced 100 rotation angles evenly spaced between −60° and +60° (1.2° spacing). The difference between the groups was the order in which these visuomotor rotations were experienced. For each of the 10 subjects in the random group, the 100 rotation angles were experienced in a different pseudorandom order. For the gradual-down group (5 subjects), the rotation angles were ordered so that the rotation gradually decreased from 0° to −60° and then increased to +60° before decreasing to 0° (that is changing in steps of 2.4° on both the up and down portions but with the two parts 1.2° out of phase). The gradual-up group (5 subjects) experienced the same order but with the rotation direction reversed.

Thus all experiments were 231 cycles, that is 1,848 trials long, and took on average 56 min (range 50–77 min). A rest of ≥1 min was enforced after the 105th cycle for all subjects, halfway through the training phase.

Analysis. The position and velocity of the hand were recorded at 1,000 Hz. Trials in which the hand speed did not exceed 10 cm/s were excluded from further analysis (0.48% of all trials; the maximum number excluded for 1 subject was 25). For all remaining trials, we calculated the adaptive state (a) that represents the visuomotor rotation for which the reach would have been accurate. This is based on the position of the hand when it had reached its maximum velocity in the first 8 cm of the reach or 120 ms after movement onset (hand speed first exceeding 10 cm/s), whichever occurred first. The 120-ms limit ensured our measure of adaptive state did not include any visual feedback corrections (Shabbott and Sainburg 2009; Saunders and Knill 2005). The adaptive state was calculated as the angle between the line joining the hand to the starting hand position and the line joining the target to the starting cursor position. The reach error is calculated by subtracting the adaptive state from the rotation angle.

We analyzed how learning at one target generalized to other targets by calculating a spatial generalization measure, which was the absolute reach error on the previous trial minus the absolute reach error on the current trial. For this purpose, reach error on the current trial was calculated using the rotation angle of the previous trial, since a change of rotation angle cannot affect the first trial because our measure of adaptive state only uses kinematic data from the part of the trial before visually mediated corrections are possible. Individual measurements were thresholded at ±20° before averaging to reduce the effect of outliers.

The movement initiation time on each trial was the time between the target being displayed on the screen and the hand speed first exceeding 3 cm/s.

Modeling. To perform statistical analysis of the noisy adaptive state data from the test phase (that is the final 3 rotation blocks of +30°, −30°, and +30°), we fit a nonmechanistic exponential-based model to these time series and extracted measurements of learning from these functions rather than from the raw data. The model was a piecewise function with four pieces: a flat line segment for the last 15 trials of the washout phase, and one exponential function for each test block (see Fig. 6, *top*, for examples). Each exponential was of the form: $\hat{a}_i = \hat{a}_\infty + (\hat{a}_O - \hat{a}_\infty)e^{-qi}$, where i is the trial number (starting at 0 on the first trial of a block), $\hat{a}_i$ is the predicted adaptive state on trial i in degrees, $\hat{a}_O$ is the adaptive state on the first trial, $\hat{a}_\infty$ is the asymptotic adaptive state, and q is the rate constant. As explained above, $\hat{a}_O$ cannot be affected by the current rotation angle, and therefore the function was constrained to be continuous, such that adjacent pieces make the same prediction for the first trial of a test block. This approach is superior to fitting an exponential to each block independently of the surrounding trials, since the adaptive states on the few trials immediately preceding a block provide considerable information about the adaptive state at the start of the block. The constraints leave seven free parameters, $\hat{a}_0^1$, $\hat{a}_0^2$, $\hat{a}_0^3$, $\hat{a}_\infty^3$, q_1, q_2, and q_3, where superscripts denote test block number.

We used a robust fitting procedure so that our results would be less sensitive to outliers, which can arise from subjects occasionally losing concentration or adopting cognitive strategies. Thus we used an error model in which each measured a_i was assumed to be drawn with probability α from a Gaussian distribution with mean $\hat{a}_i$ and SD σ (the standard squared error model), or with probability $1 - \alpha$ from a uniform distribution around the circle (to represent more random outlier responses).

Thus:

$$p(a_i|\hat{a}_i) = \alpha\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(a_i - \hat{a}_i)^2}{2\sigma^2}\right] + (1 - \alpha)\frac{1}{360}.$$

The total likelihood, the product of $p(a_i|a_i^*)$ over all blocks and trials, was maximized by simultaneously optimizing σ, α, and the seven parameters of the exponential-based function, by conjugate gradient descent, separately for each subject. Although adaptive state is a circular variable, the values it takes in this experiment are highly concentrated in a narrow range around zero (compared to the extreme of 180°) and therefore we can approximate the error measure as Gaussian. Consistent with this view, in the fits σ varied between 4.2° and 9.6°, with median 6.6°, and α varied between 0.92 and 1, with median 0.99.

From the fitted parameters, we calculated two measures of the learning rate by assessing the drop in error over the first 10 trials of each block. The first measures the absolute value of the drop, $\text{sign}[r](\hat{a}_{10} - \hat{a}_0)$, where r is the rotation angle of the block. The second is the drop in error as proportion of the error at the start of the block, $(\hat{a}_{10} - \hat{a}_0)/(r - \hat{a}_0)$. The purpose of dividing the 10-trial drop by the initial error is to more fairly compare between cases where the initial error is very different. A faster drop in absolute error is expected for larger initial errors under popular linear state-space models of adaptation (Thoroughman and Shadmehr 2000; Donchin et al. 2003; Smith et al. 2006; Huang et al. 2011). The choice of 10 trials was arbitrary; we also calculated absolute and proportional drops over the first 5, 15, or 20 trials and found that none of the conclusions of the study were affected by changing the number of trials used.

We also implemented the dual-process state-space model of adaptation proposed by Smith et al. (2006). We use similar nomenclature to Smith et al. (2006) to define the model:

$$x_f(i + 1) = A_f \cdot x_f(i) + B_f \cdot e(i)$$
$$x_s(i + 1) = A_s \cdot x_s(i) + B_s \cdot e(i)$$
$$B_f > B_s, A_s > A_f, x(i) = x_f(i) + x_s(i)$$

where i is the trial number, x represents the adaptive state, x_f and x_s represent the contributions of the fast and slow processes respectively, A is the "retention factor," B is the "learning rate," and e is the reach error. Note that a large retention factor means slow forgetting.

As the parameters of the model could change during the training phase we fit the model by running it on the full length of the experiment but only optimizing to maximize the likelihood for the test phase, by which time the parameters are likely to have stabilized. The effects on x_f and x_s of any errors in the parameters for the initial part of training are likely to have been washed out by the end of the training. Consistent with this is that only starting the simulation from the middle of training gave very similar results and did not affect the significant differences in parameters across groups reported in RESULTS.

We fit each subject's data separately and used the same robust error model described above, with parameters fixed at the median values found during fitting of the exponential-based functions ($\alpha = 0.99$; $\sigma = 6.6°$). The fit was implemented in MATLAB by the fminsearch function with the four parameters (learning and retention factors for the slow and fast processes) each curtailed such that they lay between zero and one.

RESULTS

Twenty-seven naïve subjects participated in a center-out discrete reaching task with eight targets evenly spread around the circle. All subjects started with reaching under veridical feedback conditions. In a subsequent "training" phase, three groups were given experience of a fixed set of 100 visuomotor rotations (each for a 16-trial block) spaced evenly from −60° (anticlockwise) to +60°, while a further control group was given experience of veridical reaching (Fig. 1 shows the rotation orders for the groups). One of the rotation-trained groups (the random group) experienced the rotations in random order, while for the other two groups (gradual) the angle gradually increased from 0° to one extreme (±60°), then gradually to the other extreme (∓60°), and then back to 0°. These two groups experienced the rotations in opposite order. All four groups were finally tested, after a further veridical-feedback washout, in a standard A→B→A paradigm, i.e., +30° rotations followed by −30° rotations, and then +30° rotations again.

Figure 2 shows outward cursor trajectories from one typical subject of each group. Figure 2, *left*, shows the accurate reaches in the final cycle of the veridical-feedback washout. Figure 2, *middle*, shows trajectories from the first cycle of the first test block; the reaches are sent off-target by the clockwise +30° rotation. Figure 2, *right*, shows the more accurate reaches in the last cycle of this block, after subjects have learned to compensate for the rotation (particularly subjects from the random and gradual groups).

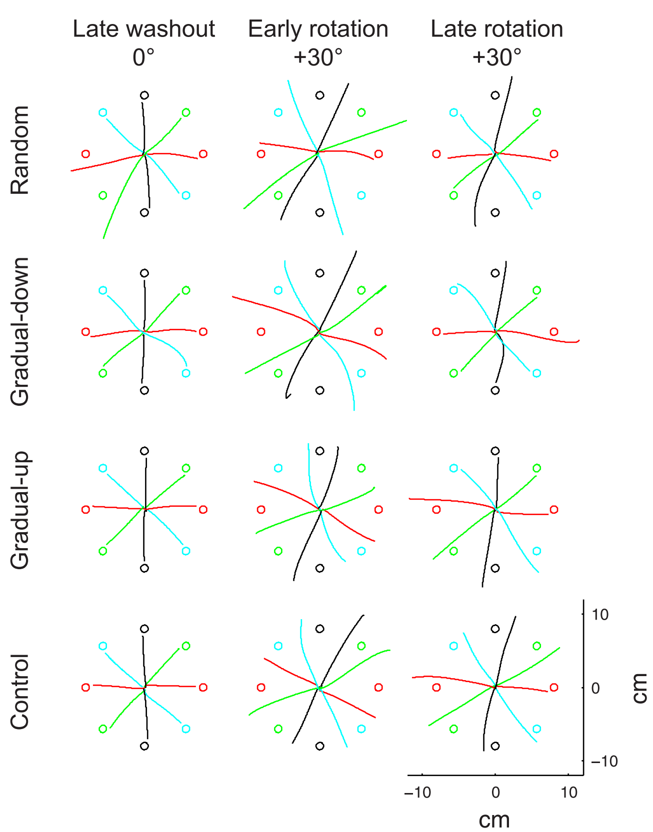

Fig. 2. Cursor trajectories under veridical and rotated reaching. For the median subject of each group (selected on the basis of the reduction in reach error from the first to the last cycle of the first test block), outward reach trajectories from 3 cycles of 8 trials are shown. *1*) Late washout cycle is the last cycle before the test rotations. *2*) Early rotation cycle is the first cycle of the first test (+30°) block. *3*) Late rotation cycle is the last cycle of the first test block.

A subject's hand trajectory on each trial was characterized with a scalar "adaptive state" representing the visuomotor rotation for which the reach would have been accurate. The difference between the adaptive state and the actual rotation angle gives the reach error. Figure 3, *left*, shows the adaptive state for example subjects during random and gradual training. This shows that subjects do not fully adapt to the perturbations when they are ordered randomly but adapt more fully when they are ordered gradually, although gradual-trained subjects also tend to under-adapt to the largest rotations. Figure 3, *middle*, shows the distribution across trials of adaptive states for all subjects of a group. To calculate the range of adaptive states experienced before the test phase, we dealt with outliers by taking the circular mean within each cycle. The range of states was on average 13.3° for the control group, 102.8° for the gradual group (in which the 2 gradual groups are combined) and 51.7° for the random group. All three values are significantly different from each other at the $P < 0.001$ level, and the 95% confidence limits on the difference between the gradual and random groups are 40.0 to 62.2°. This shows that the gradual group tracked the perturbations, adapting more fully than the random group. Figure 3, *right*, shows the proportion of targets hit, as a function of rotation angle. Across all rotations, the hit rate was 0.22 for the random group and 0.61 for the (combined) gradual group (difference significant at $P = 0.0002$, Wilcoxon rank sum test). To examine experience at hitting targets under rotations similar to the test rotations, we took the mean hit rate for the eight rotation angles closest to +30° (from 26.1° to 34.5°) and the eight closest to −30°. The means were 0.10 for the random group and 0.66 for the (combined) gradual group ($P = 0.0002$, Wilcoxon rank sum test). A clear tendency exists for hit rates to be greater for rotations in the direction experienced first during gradual training; this may be understood simply as being due to anterograde interference exerted by rotation in one direction on rotation in the opposite direction (Krakauer et al. 1999; Bock et al. 2001; Wigmore et al. 2002; Miall et al. 2004; Imamizu et al. 2007). To measure the smoothness with which the adaptive states change across blocks in the training phase, we calculated the mean of the absolute difference between adaptive states in the last cycle of a block and the last cycle of the subsequent block, in the part of the experiment before the test phase. The result was an average change of 1.79° for the control group, 3.08° for the (combined) gradual group, and 11.99° for the random group. All three values are significantly different from each other at the $P < 0.001$ level. Thus during training the gradual-trained subjects experience on average twice the range of adaptive states as the random-trained subjects but experienced an adaptive state that changed much more slowly.

Figure 4, *A*–*D*, shows that the random group experience much larger errors (median magnitude: 20.9°) than the combined gradual group (median: 6.4°) and the control group (median: 3.0°). Figure 4*E* shows, for each group, the change in error magnitude across the 16 trials of a block, averaged across the 100 blocks. The fall in error from the first to the last trial of a block is significant for the random group (*t*-test: $P = 6 \times$

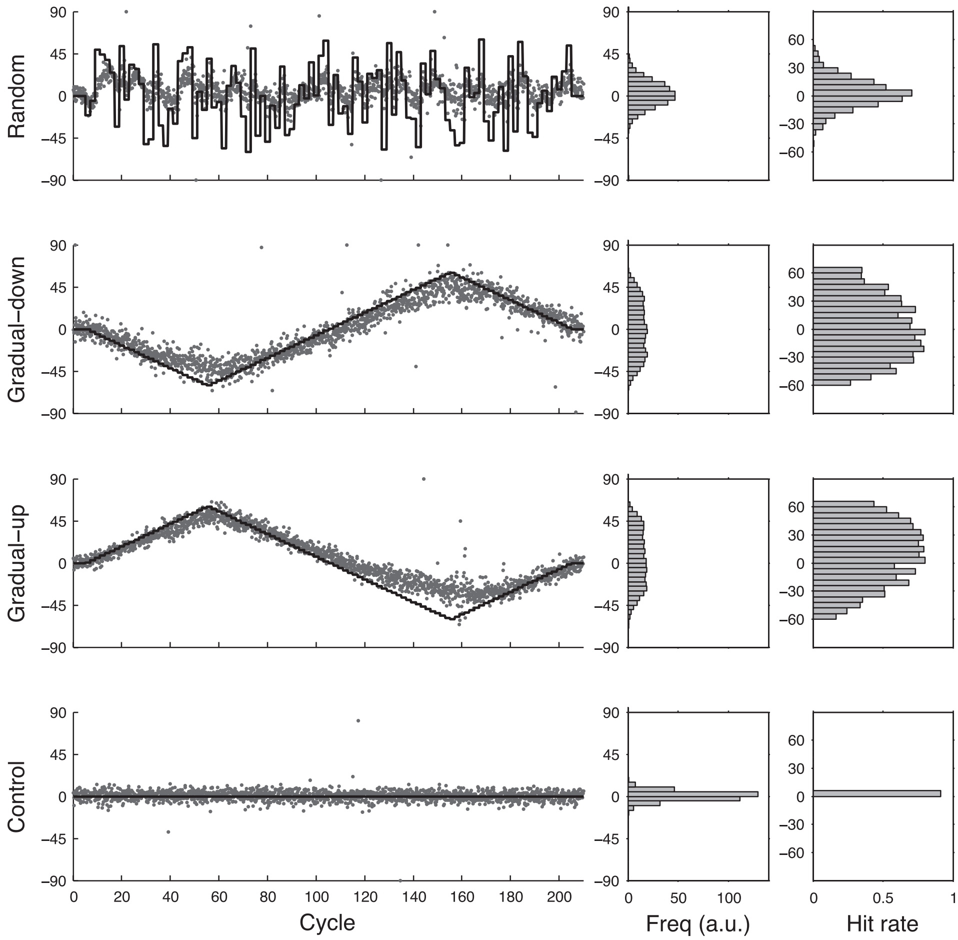

Fig. 3. Adaptation during the training phase and veridical reaching blocks. *Left*: for each trained group the rotation angle (lines) and adaptive state on each trial (dots) in the baseline, training and washout phases are shown for an example subject (the same subject as in Fig. 2). Adaptive state represents the rotation angle for which the reach would have been accurate; adaptive states are shown here constrained within a ±90° limit for visibility. *Middle*: distribution of adaptive states for individual trials during the training phase and veridical reaching blocks, across all subjects in a group. *Right*: the proportion of targets hit during this period, as a function of the rotation angle. a.u., Arbitrary units.

10^{-5}, 95% confidence limits: 8.1° to 15.8°) but not for any other group. Therefore, only the random group, who generally have a large error at the start of a block, shows notable improvement over the course of the 16 trials with a new rotation. Figure 4*F* shows the change in error magnitude across the training phase, averaged across all 16 trials of each block and within groups of 10 blocks. The error on the last 20 blocks is significantly lower than for the first 20 blocks in the random group ($P = 0.0002$, 95% confidence limits on the difference: 5.0° to 10.7°). The gradual group shows peaks in error when the rotation angle is large (marked with asterisks).

Adaptation to a visuomotor rotation experienced when reaching to one target in a center-out continuous feedback task only leads to limited adaptation for reaches to other targets, obeying an approximately Gaussian generalization function with $\sigma \approx 30°$ (Krakauer et al. 2000). An increase in generalization across targets might be one mechanism by which training on rotations leads to faster adaptation to later rotations in a multitarget paradigm. Figure 5 shows generalization functions that are the reduction in absolute reach error from one trial to the next as a function of the difference in target angles between the two trials. Figure 5, *A* and *B*, shows generalization across the whole training phase and the whole test phase. All functions show narrow generalization, with the height of the function varying across groups depending on the size of the errors experienced or the general speed of adaptation. Figure 5, *C* and *D*, shows that the width of generalization does not increase from the first half of the training phase (solid lines) to the second half (dashed lines) in either the random group or the (combined) gradual group (no data points significantly different at the $P < 0.01$ level between early and late training).

Characterizing adaptation with exponential-based functions. To measure the rate of adaptation during the first few trials of exposure to a test rotation in a way robust to the noise on the adaptive state measures, the time courses of adaptation were fit with exponential-based functions. Figure 6, *top*, shows this fit for a typical subject from each group. The exponential-based fits are constrained so that the whole function is continuous across the test blocks (see METHODS). Figure 6, *middle*, shows the exponential-based functions fit to all subjects of the four groups. The random and gradual groups show considerable within-group variation, with some subjects adapting rapidly and some very slowly, whereas all subjects of the control group adapt relatively slowly and show considerable interference between the learning of the opposite rotations, with the adaptive state during the middle −30° block never getting far below 0°. Means across subjects, for each of the four groups, are shown in the bottom graph of Fig. 6.

Variations in adaptation across groups. Our primary measure of the rate of adaptation is the drop in error over the first ten trials, as a proportion of the error on the first trial. This measure avoids confounding the rate of adaptation to the current rotation with the aftereffect of the previously learned rotation and was calculated from the exponential-based functions shown in the middle row of Fig. 6. Figure 7, *left*, shows, for each subject group, the proportional drop for each of the

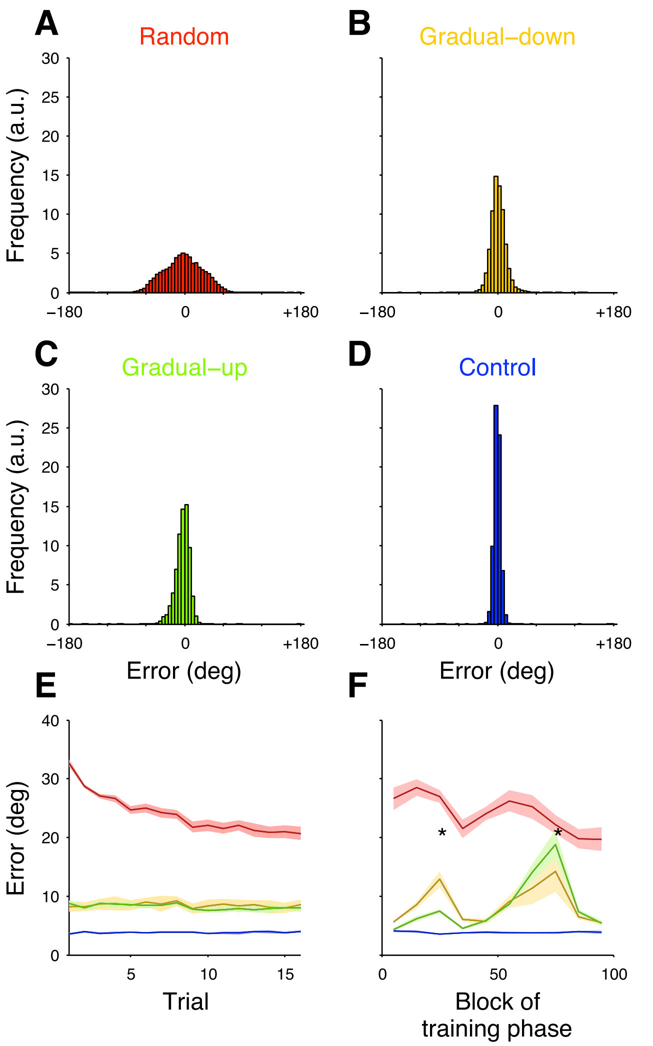

Fig. 4. Reach errors during the training phase. *A–D*: distribution of errors (adaptive state minus rotation angle) across the training phase, across all subjects in an experimental group. *E*: time course of error (means ± 1 SE across subjects) over the 16 trials of a training block, averaged across all 100 blocks, for each of the 4 groups (groups colored as in *A–D*). *F*: time course of error over the whole training phase, averaged over all 16 trials within a block and over groups of 10 blocks. *Blocks of maximal (±60°) rotation for the gradual groups.

three test blocks (means ± SE across subjects), and Fig. 7, *right*, shows the mean across blocks for each subject group. The mean drop for the random group is significantly greater than for the control group ($P = 0.004$ on *t*-test, 95% confidence limits on the difference: 0.10 to 0.43), and the difference is significant at the $P < 0.05$ level for each block. The mean drop for the (combined) gradual group is significantly greater than for the control group ($P = 0.003$, 95% confidence limits: 0.12 to 0.49), and the difference is significant at the $P < 0.05$ level for each block. Comparing the drop for the gradual and random groups shows there is no significant difference on any block or on the across-block average.

Examining the absolute 10-trial drop gives the same qualitative differences between groups. The average drop for the random group (19.4°) is significantly greater ($P = 0.004$) than for the control group (7.6°) and the difference is significant at the $P < 0.05$ level for each block. The same is true when comparing the gradual group (mean drop 22.9°) with the control group. Again, the drop is never significantly different between the random and gradual groups. To examine the completeness of adaptation, we measured how far a subject's average adaptive state (using the circular mean) fell short of the rotation angle in the last two cycles of a test block. We found this shortfall, averaged across all three blocks, to be less for the random ($P = 0.012$) and combined gradual ($P = 0.0005$) groups than for the control group. It tends to be less for the gradual group than for the random group, albeit not significantly (mean difference: 3.6°, $P = 0{:}15$). The same results are found if the final adaptive state is instead taken as that predicted for the last trial of the block by the exponential-based function fits.

Variations in adaptation across test blocks. We examined the rate of adaptation to the three test blocks to assess interference between the learning of the +30° and −30° rotations. Interference might be expected to reduce the rate on the −30° block compared with the first +30° block, but we find that the 10-trial drop in error is greater on the −30° block for both the control and random and (combined) gradual groups ($P < 0.05$ for each). When the 10-trial drop is normalized by the initial error the difference is not significant for any group, although the trend in all groups is for faster adaptation in the −30° block. This suggests that any interference of the first +30° block on adaptation to the −30° rotation consists only of an aftereffect (higher initial error).

Interference of the −30° block on learning of the +30° rotation could affect the second +30° block both through an aftereffect and through the prevention of savings, meaning a faster learning rate on the second exposure to the +30° rotation. Our paradigm does not reveal how much savings would have been seen if the −30° block had been instead a washout or break, but it can be shown that the rate of adaptation on the second +30° block is at least not less than that on the first. The absolute 10-trial drop is greater on the last block for all three groups ($P < 0.05$), and as a proportion of the initial error it is greater for the random group ($P < 0.05$) and also tends to be greater for the gradual and control groups (albeit not significantly).

Fitting test adaptation using state-space models. Smith et al. (2006) proposed a model of visuomotor adaptation composed of two adaptive processes; one that learned fast but with poor retention, and one that learned slowly but with good retention. This model cannot capture meta-learning effects and is unable to account for the different rates seen across our groups of subjects under the assumption that the groups have similar and fixed parameter settings for the learning rates and retention factors. Here we ask what parameters of the two-process model could be altered by meta-learning that may occur during random and gradual training.

We fit the model to the test phase for each subject separately (although the model is run on the whole experiment, so that the processes are appropriately adapted before the test phase begins). The model output, averaged across subjects, is shown in the top row of Fig. 8, and the means and ranges of parameters are shown in Table 1. The model is able to closely approximate the adaptive state for all subjects, and Fig. 8, *bottom*, shows (in the style of Fig. 7) that the drop in error over the first 10 trials of each test block shows a similar trend in the model fits to that measured using the exponential-based fits.

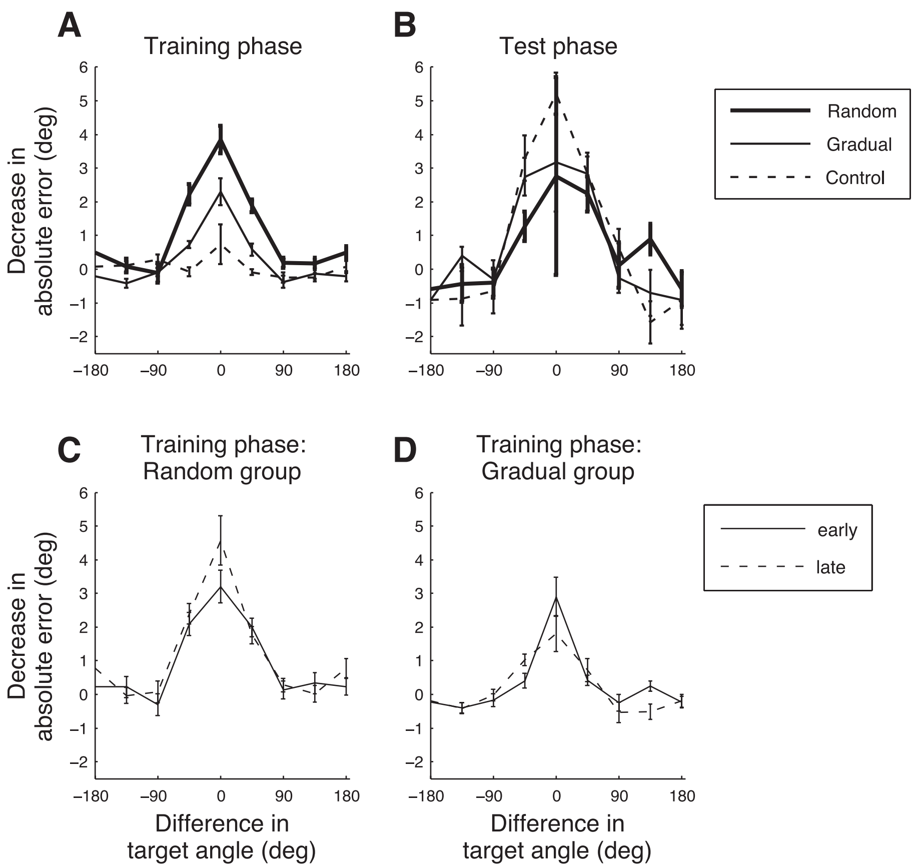

Fig. 5. Generalization of learning across target directions. Reduction in absolute reach error from one trial to the next ($|e_{t+1}| - |e_t|$) is plotted as a function of the difference in target angle between the 2 trials ($\theta_{t+1} - \theta_t$). Error bars are ± 1 SE across subjects. *A*: generalization functions calculated from all pairs of trials within the training phase. *B*: generalization functions from the test phase. *C-D*: generalization functions from the first and second halves of the training phase, for the random and (combined) gradual groups.

Examining the slow process parameters shows that neither the retention factor nor the learning rate is significantly different among the random, (combined) gradual, and control groups (Kruskal-Wallis test, $P = 0.28$ and $P = 0.53$, respectively). Examining the fast process rates shows that the retention factor varies significantly between groups ($P = 0.03$), and post hoc tests reveal the gradual group's retention factor is significantly higher than for the control group. In addition, the fast process learning rate also varies significantly between the groups ($P = 0.01$), and post hoc tests reveal the random and gradual group rates are both significantly higher than for the control group. When the two control groups are considered separately, there is

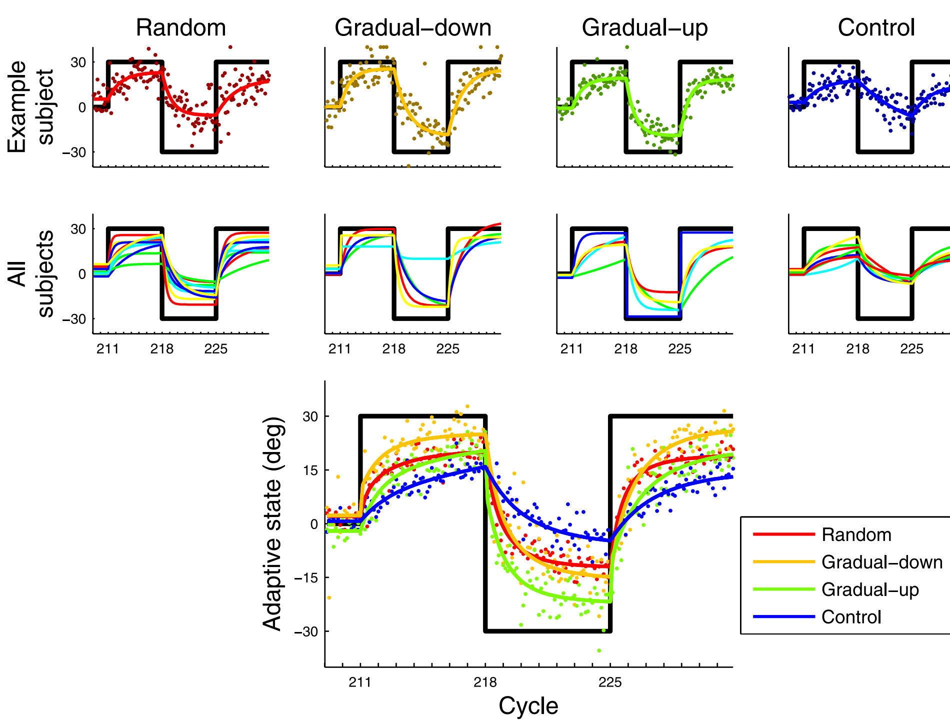

Fig. 6. Adaptive state measures and exponential adaptation functions. *Top*: adaptive state measures in the last 2 cycles of the washout and the 3 test blocks (shown here constrained within a ±40° limit for visibility), and the exponential-based functions fit to this data, for an example subject from each group (the same subjects shown in Fig. 2). Straight black lines show the rotation angle. *Middle*: exponential-based functions for all subjects of each experimental group. *Bottom*: mean adaptive state measures across subjects (dots) and mean adaptive state predicted by the exponential-based functions fit to individual subjects (curves).

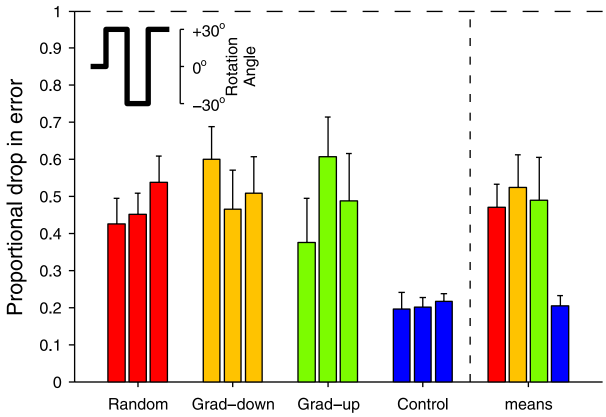

Fig. 7. Drop in reach error over the first 10 trials of test blocks, as a proportion of the initial error. This measure is computed from the exponential-based functions fit to individual subjects' adaptive state measures. Error bars show 1 SE of the mean across subjects. The 4 *leftmost* groups of bars show the drop in each of the 3 test blocks. *Rightmost* group shows the mean drop across test blocks for each of the 4 subject groups.

very little difference between them for any parameter (A_f means: 0.86 for gradual-down group and 0.93 for gradual-up group; A_s: 1.00 and 1.00; B_f : 0.098 and 0.094; and B_s: 0.016 and 0.020).

Variations in initiation time. Figure 9*A* shows, for each group, the change in movement initiation time across the 16 trials of a training block, averaged across the 100 blocks. Although the average error for the random group falls considerably from the first to the last trial of a block ($P = 6 \times 10^{-5}$; Fig. 4*E*), there is no such change in initiation time ($P = 0.38$). Figure 9*B* shows the change in initiation time across the training phase, averaged across all 16 trials of each block and within groups of 10 blocks. No group shows a significant change in initiation time from the first to the last 10 blocks, although the gradual groups tend to show increases in initiation time that parallel the increases in error that occur as the ±60° extremes of rotation angle are reached. Figure 9*C* shows initiation time during the washout and test blocks (separated by vertical black lines). Mean initiation time during the washout shows no significant differences across groups. In the first three cycles of each test block, initiation times for the control group are on average raised from that in the washout (mean increase of 21 ms, $P = 0.03$), and the same is true for the gradual group (mean increase of 31 ms, $P = 0.01$), while the random group shows no increase ($P > 0.15$ for each block). This initiation time increment is also significantly greater across blocks for both the gradual and control group compared with the random group ($P < 0.05$ in either case), but there is no significant difference between the gradual and control groups ($P = 0.35$). We found no significant correlation across subjects, within any group, between this measure and the proportional 10-trial drop in error.

DISCUSSION

In our study, we found that subjects trained on visuomotor rotations spaced between −60° and +60° adapt faster to subsequent +30° and −30° rotations than control subjects who are only trained with veridical reaching. This facilitation was seen regardless of whether the training rotations are experienced in random order or in gradual order where the rotation angle changes only slowly. The random and gradual groups learn the test rotations at similar rates, suggesting that both types of experience are effective at enhancing the mechanisms involved in adaptation to visuomotor rotations. However, the random group initiate reaches more rapidly than the gradual group when exposed to the test rotations. Unlike the gradual and control group, they show no significant increase in initiation time during the test phase. This shows that random and gradual exposures do not affect the visuomotor system identically. Furthermore, by fitting a two rate state-space model we

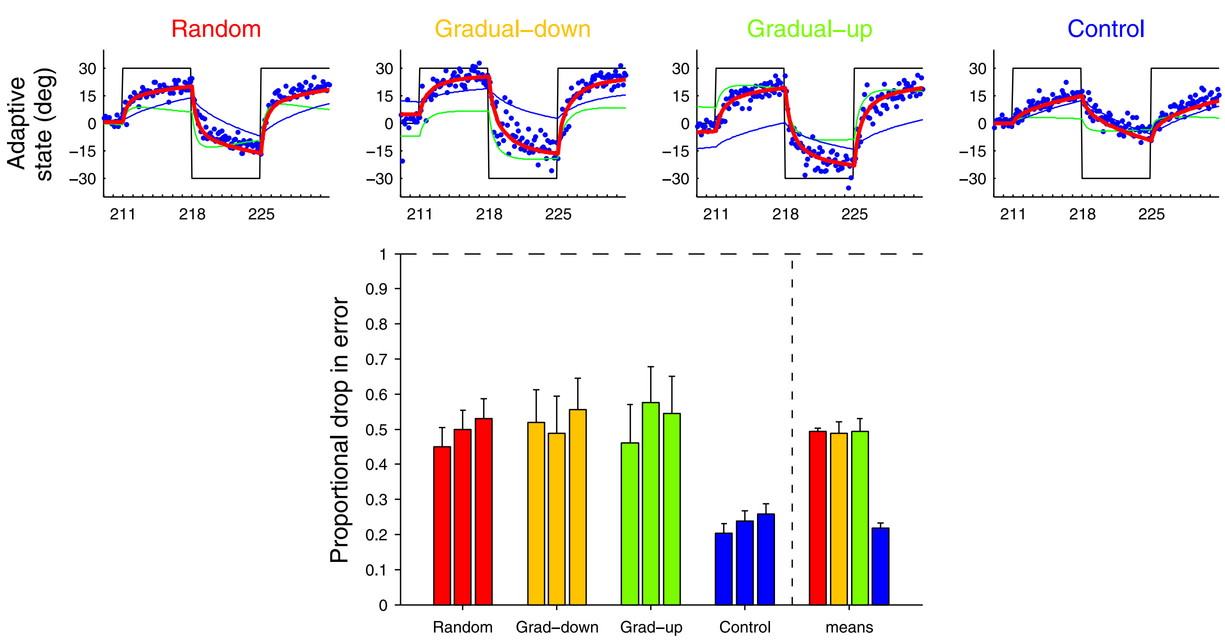

Fig. 8. Fits of the dual-process model. *Top*: adaptive states and model fits to the test phase, for each of the four subject groups. Model was fit to the test phase for all subjects individually and predictions averaged across subjects. Blue dots are adaptive state measures; blue curves represent the slow process state; green curves represent the fast process state; and red curves represent the predicted motor output, which is the combination of the 2 states. *Bottom*: drop in error over the first 10 trials of each block (as a proportion of the initial error) predicted by the model fits, plotted as for Fig. 7.

Table 1. *Parameters of dual-process model fits*

	A_f	A_s	B_f	B_s
Random group	0.77 ± 0.09	1.00 ± 0.00	0.098 ± 0.019	0.021 ± 0.005
Gradual group	0.89 ± 0.05	1.00 ± 0.00	0.096 ± 0.024	0.018 ± 0.005
Control group	0.74 ± 0.06	1.00 ± 0.00	0.034 ± 0.007	0.012 ± 0.002

Values are means ± SE across subjects with parameters averaged across fits to each individual subject of a group. Parameters for the dual-process model of Smith et al. (2006) fit to the test phase adaptive state data.

found that there were significant differences between the learning rates and retention factors of the fast process that could account for the differences in the performance of the groups in the test phase. That is the learning rate was greater in the gradual and random groups compared with the control and the retention factor was significantly larger in the gradual group. This suggests that facilitation of learning can be seen after both gradual and random task variation and that these may change parameters of the fast process of the two rate state-space model.

Contextual interference. In our study, we compared groups that underwent randomly and gradually changing tasks, whereas previous studies compared groups that underwent blocked and random tasks during training and investigated the subsequent effect on retention and transfer (e.g., Shea and Morgan 1979). The superior retention and transfer of the random group in such studies have often been attributed to "contextual interference" (Battig 1966, 1972, 1979). The hypothesis of contextual interference suggests that increased difficulty experienced during random practice leads to the activation of multiple processing mechanisms that allow for improved performance on later testing. However, if we take the reach error as the measure of difficulty, our findings suggest that task difficulty measured by large errors is not necessary for enhancing adaptivity. That is although the control, gradual, and random groups had average errors of 3.0°, 6.3°, and 20.9°, respectively, during the training phase, the gradual and random group showed similar facilitation on testing despite the large difference in errors previously experienced. Indeed, a recent study (Abe et al. 2011) showed that learning under conditions of reward leads to better long-term retention than neutral conditions or punishment; since the gradual group experiences more implicit reward than the random group during training, in the form of target hits, one might expect them to perform better on the test rotations (Pekny et al. 2011; Izawa and Shadmehr 2011).

Similarly, our results suggest that rapid changes in task parameters, or in the adaptive state of the subject, are not necessary conditions for facilitation of learning, since the rotation angle and adaptive state change much slower during gradual training than during random training, and yet both groups perform similarly in the test phase. Indeed, rapid changes in task parameters might reduce later adaptation rates; several studies have shown that adaptation to a given error is reduced in a randomly changing environment vs. a slowly changing environment (Cheng and Sabes 2007; Donchin et al. 2003; Smeets et al. 2006), perhaps due to reduced retention of learning in a rapidly changing environment (Huang and Shadmehr 2009). This may explain the nonsignificant trend in our results for gradual-trained subjects to learn the test rotations more completely than random-trained subjects.

Another notable difference between gradual and random training is the experienced range of adaptive states, which during gradual training is roughly twice that during random training, since gradual subjects adapt more fully to the training rotations. This may lead to more thorough learning of the structure or to training of a wider variety of internal models (Wolpert and Kawato 1998; Haruno et al. 2001) leading to faster switching between these models during the test phase and therefore to faster adaptation. It may be that this advantage of gradual training over random training compensates for the absence of other stimuli to the adaptive mechanisms, such as large errors, rapid changes in task parameters, and rapid changes in adaptive state, such that the net difference in

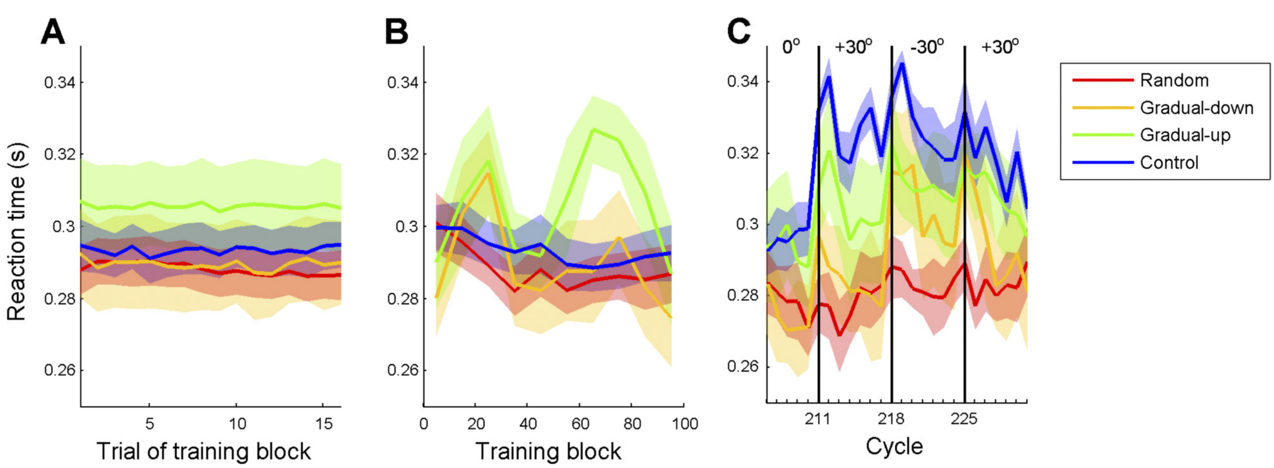

Fig. 9. Initiation time for reaches. *A*: time course of initiation time (means ± 1 SE across subjects) over the 16 trials of a training block, averaged across all 100 blocks, for each of the four groups. Initiation times are measured from target appearance until the time the hand speed first exceeds 3 cm/s. *B*: time course of initiation time over the whole training phase, averaged over all 16 trials within a block and over groups of 10 blocks. *C*: time course of initiation time over the washout and test phases. Black lines represent the onset of test rotations.

adaptation rate during the test phase between the two subject groups is small. Thus it is possible that several of the features of training discussed here lead to considerable enhancement of adaptation, despite the lack of difference between the learning rates of the two groups.

Structural learning. Braun et al. (2009b) showed that subjects trained on a random sequence of visuomotor rotations adapted faster to test rotations and showed less interference between opposite rotations than controls trained with veridical reaching. Their experiment also demonstrated that some of the adaptation benefit of random rotation training was specific for rotations. This demonstrates that the increase in adaptation rate and reduction in interference does not arise purely from a generic increase in adaptability (e.g., Seidler 2004). This leaves open two possibilities: *1*) that random rotation training caused subjects to learn the structure of rotations so that their sensorimotor system could rapidly move to any point along that structure, or *2*) that rotation training induced the learning of two or more internal models (Wolpert and Kawato 1998; Osu et al. 2004; Lee and Schweighofer 2009), between which subjects were able to switch on the basis of a few trials of contextual evidence. To distinguish between these possibilities, Braun et al. (2009b) examined subjects who were trained on both the test rotations intermingled with random linear transformations that were not restricted to rotations. These subjects did not demonstrate improved performance on the test rotations even though they had experienced a large number of the test rotations during training. These results argue in favor of *hypothesis 1* of structural learning, since under *hypothesis 2* subjects would be expected to learn internal models suitable for these test rotations even though the random linear transformations did not correspond to a structure. However, despite this result it is still possible that *hypothesis 2* accounts in part for the superior performance of the trained subjects compared with the control subjects; that is, the trained subjects adapt faster to the −30° and +30° rotations because they have experienced them before. However, the main focus of the current study is to investigate the role of temporal ordering during learning rather than to test between these two hypotheses.

Although the gradual and random groups show similar facilitation of learning as assessed by reach direction, we found important differences between the initiation times of these groups. The initiation time of the random group remained low during the test phase unlike the control or gradual groups whose initiation times increased. A recent study (Fernandez-Ruiz et al. 2011) showed that when subjects' initiation times are unconstrained by the task there is a strong correlation between initiation time and reduction in directional movement error during visuomotor learning. However, when subjects were required to initiate movement within 350 ms of target appearance, they showed only very gradual improvements in error and very little variation in reaction times. In our study, we also found no significant correlation between reach error and reaction time, as our subjects were required to initiate movement within 400 ms. Fernandez-Ruiz et al. (2011) have related larger initiation times to more cognitive strategies that possibly involve spatial working memory (mental rotation). Therefore, despite the similarity in spatial performance, there is an advantage to random training in terms of initiation time and this would suggest that the random group might be employing a less cognitive strategy during the test phase. Random-trained subjects might be expected to make greater use of cognitive strategies, since these subjects were aware of perturbations throughout training whereas gradual-trained subjects often reported being unaware until the latter half of the training phase. However, the faster initiation times in the random group compared with the gradual groups during testing suggest that implicit adaptation mechanisms eventually dominate such cognitive strategies. Whether the reduction in initiation time might also be a signature of structural learning remains to be explored.

To investigate whether training or its temporal order had any effect on spatial generalization, we examined the generalization function across different target directions for the different groups and for early and late training (Fig. 5). However, we found no changes in the width of the generalization for the random and gradual groups over the course of training. This suggests that the possible underlying processes of structural learning do not affect spatial generalization.

Task setup and error measures. There are notable methodological differences between our current study and Braun et al. (2009b). First, the present study used discrete movements with short initiation time (0.4 s), similarly to a number of other studies (Miall et al. 2004; Sing et al. 2009), whereas Braun et al. (2009b) imposed only a 2-s limit for reaching the target (although subjects were asked to move swiftly). The short initiation time in the present study, as well as the use of eight randomized targets common to both studies, should reduce the possibility of subjects making use of cognitive strategies (Weiner et al. 1983; Bedford 1993; Redding and Wallace 2002, 2006a,b; Ghilardi et al. 2009; Fernandez-Ruiz et al. 2011).

Second, we used 30° test rotations rather than the 60° used by Braun et al. (2009b), and a −60° to +60° rather than −90° to +90° range of exposure rotations. Since large rotations (especially 90° and larger) may induce the use of more cognitive strategies (Imamizu et al. 1995; Abeele and Bock 2001; Miall et al. 2004), it is useful to demonstrate here that the apparent structural learning effect is replicated under conditions where cognitive strategies are less likely to be used.

Third, our method of characterizing the time course of adaptation with exponential functions allows estimates of the magnitude of the improvement in learning rate brought about by prior exposure to rotations. Whereas the mean fall in error over the first 10 trials of the 3 test blocks is just 7.6° for the control group, this measure for the random exposure group is 19.5° (95% confidence limits on the difference: 4.4° to 19.3°), showing an effect of considerable magnitude. Direct comparison of trial 11 with trial 1 would provide much more noisy estimates of this learning measure.

Zarahn et al. (2008) also used exponential functions to empirically compare adaptation across conditions. This study compared rate constants (q in our parameterization, see METHODS) rather than trial-by-trial changes in performance. The 10-trial drop in error is closer to such a measure and is more robust to measurement noise than q, which varied in our fits over a factor of 9,000. Since most of the adaptation in a block takes place over just 10–20 trials, the noisy data are often consistent with a wide variety of functions that yield similar values of the 10-trial drop but have very different values of q. Further, Zarahn et al. (2008) only used data from the current test block when fitting exponential functions, although these blocks followed straight on from previous blocks without breaks, making it valid to fix the starting adaptation level of the exponential at the average of the last few trials of the previous block, as in the present study. The starting level of the exponential is particularly poorly constrained by the data in the current block, so our approach reduces the sampling error on the estimate of this parameter and therefore on estimates of the other parameters.

Our approach of analyzing the speed of adaptation (Sing and Smith 2010; Krakauer 2009; Zarahn et al. 2008; Smith et al. 2006; Kojima et al. 2004; Miall et al. 2004; Bock et al. 2001; Abeele and Bock 2001), rather than averaging errors over the first few or all the trials of exposure to a task (Cunningham 1989; Brashers-Krug et al. 1996; Shadmehr and Brashers-Krug 1997; Imamizu et al. 2007; Braun et al. 2009b), allows us to separate the aftereffect of previous learning from the rate of adaptation. We find that in all subject groups the rate of adaptation to the −30° test rotation is at least as great as that to the preceding +30° rotation, and the rate of adaptation on the second +30° block is at least as great as that on the first +30° block. These results show that anterograde interference between the opposing rotations consists only of an aftereffect rather than a reduction in learning rate. It is possible that retrograde interference of the −30° rotation on the memory of the +30° rotation prevented savings that would have otherwise increased the rate of learning on the second +30° block, but our experiment did not test this.

State-space models of facilitation. Zarahn et al. (2008) showed that linear state-space models of adaptation (e.g., Smith et al. 2006) cannot explain processes of meta-learning and argued that variable-rate adaptive processes are necessary to account for such phenomena. More specifically, Huang and Shadmehr (2009) showed that gradual introduction of a dynamic perturbation increased the retention factor of adaptive processes later in the experiment and that rapid introduction of a perturbation reduced the retention factor. Similarly, we show that parameters of the Smith et al. (2006) model during the test phase are altered by previous experience of visuomotor rotations. With fits to individual subjects, we find that the learning rate of the fast process is increased considerably by both random and gradual learning. Such a rate change could be driven by a number of meta-learning processes, for example, by a generic increase in adaptation rate in response to error magnitude or variability or mapping uncertainty (e.g., Burge et al. 2008), or structural learning (e.g., Braun et al. 2009b). We also found the retention parameter of the fast process to be increased by gradual learning but not random learning. This is in accordance with the aforementioned finding of Huang and Shadmehr (2009); both studies suggest that properties of adaptive processes are adapted within minutes to the properties of the environment, with learning in rapidly changing environments being rapidly forgotten and learning in slowly changing environments being retained over longer timescales. However, it should be remembered that the dual-rate mechanism of the Smith et al. (2006) model does not necessarily reflect the actual mechanism of meta-learning in our study; as discussed above, it may be that faster adaptation to test rotations is actually due to switching between multiple internal models that were learned during the training phase. In the future, it will be interesting to develop variable rate models that can capture meta-learning processes as reported in our study and to fit them to trial-by-trial movement data.

In conclusion, we found facilitation of visuomotor rotation learning after subjects underwent training with either randomly or gradually varying rotations. Additionally, subjects that experienced random training showed a significantly lower movement initiation time compared with subjects who experienced gradual or no training. When fitting a state-space model with a fast and a slow learning process to our data, we found that random and gradual training had adapted parameters of the fast learning module. These differences between subject groups suggest that learning rates are adapted through experience as part of meta-learning processes that can depend on previous errors, task variability, mapping uncertainty, and structural similarity between tasks. While the way in which errors lead to changes in the internal state are well understood, future studies will need to understand how different features of experience change the learning rates and retention parameters of the processes underlying learning.

GRANTS

This work was funded by the Wellcome Trust and the European SENSOPAC project (IST-2005 028056). D. A. Braun was supported by the German Academic Exchange Service (DAAD) and an Emmy Noether grant (DFG).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

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

Abeele S, Bock O. Sensorimotor adaptation to rotated visual input: different mechanisms for small versus large rotations. *Exp Brain Res* 140: 407–410, 2001.

Battig WF. Facilitation and interference. In: *Acquisition of Skill*, edited by Bilodeau E. New York: Academic, 1966.

Battig WF. Intratask interference as a source of facilitation in transfer and retention. In: *Topics in Learning and Performance*, edited by Thompson R, Voss J. New York: Academic, 1972.

Battig WF. The flexibility of the human memory. *Levels of Processing in Human Memory*, edited by Cermak L, Craik F. Hillsdale, NJ: Erlbaum, 1979.

Bedford FL. Perceptual learning. In: *The Psychology of Learning and Motivation*, edited by Medin D. New York: Academic, 1993, vol. 30, p. 1–60.

Bock O, Schneider S, Bloomberg J. Conditions for interference versus facilitation during sequential sensorimotor adaptation. *Exp Brain Res* 138: 359–365, 2001.

Brady F. The contextual interference effect and sport skills. *Percept Mot Skills* 106: 461–472, 2008.

Brashers-Krug T, Shadmehr R, Bizzi E. Consolidation in human motor memory. *Nature* 382: 252–255, 1996.

Braun DA, Aertsen A, Wolpert DM, Mehring C. Learning optimal adaptation strategies in unpredictable motor tasks. *J Neurosci* 29: 6472–6478, 2009a.

Braun DA, Aertsen A, Wolpert DM, Mehring C. Motor task variation induces structural learning. *Curr Biol* 19: 352–357, 2009b.

Braun DA, Mehring C, Wolpert DM. Structure learning in action. *Behav Brain Res* 206: 157–165, 2010.

Burge J, Ernst MO, Banks MS. The statistical determinants of adaptation rate in human reaching. *J Vis* 8: 20, 1–19, 2008.

Cheng S, Sabes PN. Calibration of visually guided reaching is driven by error-corrective learning and internal dynamics. *J Neurophysiol* 97: 3057–3069, 2007.

Cohen HS, Bloomberg JJ, Mulavara AP. Obstacle avoidance in novel visual environments improved by variable practice training. *Percept Mot Skills* 101: 853, 2005.

Cunningham HA. Aiming error under transformed spatial maps suggest a structure for visual-motor maps. *J Exp Psychol* 15:3: 493–506, 1989.

Donchin O, Francis JT, Shadmehr R. Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J Neurosci* 23: 9032–9045, 2003.

Duncan CP. Description of learning to learn in human subjects. *Am J Psychol* 73: 108–114, 1960.

Fernandez-Ruiz J, Wong W, Armstrong IT, Flanagan JR. Relation between reaction time and reach errors during visuomotor adaptation. *Behav Brain Res* 219: 8–14, 2011.

Ghilardi MF, Moisello C, Silvestri G, Ghez C, Krakauer JW. Learning of a sequential motor skill comprises explicit and implicit components that consolidate differently. *J Neurophysiol* 101: 2218–2229, 2009.

Halford GS, Bain JD, Maybery MT, Andrews G. Induction of relational schemas: common processes in reasoning and complex learning. *Cogn Psychol* 35: 201–245, 1998.

Harlow HF. The formation of learning sets. *Psychol Rev* 56: 51–65, 1949.

Haruno M, Wolpert DM, Kawato M. Mosaic model for sensorimotor learning and control. *Neural Comput* 13: 2201–2220, 2001.

Howard IS, Ingram JN, Wolpert DM. A modular planar robotic manipulandum with end-point torque control. *J Neurosci Methods* 181: 199–211, 2009.

Huang VS, Haith A, Mazzoni P, Krakauer JW. Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. *Neuron* 70: 787- 801, 2011.

Huang VS, Shadmehr R. Persistence of motor memories reflects statistics of the learning event. *J Neurophysiol* 102: 931–940, 2009.

Imamizu H, Sugimoto N, Osu R, Tsutsui K, Sugiyama K, Wada Y, Kawato M. Explicit contextual information selectively contributes to predictive switching of internal models. *Exp Brain Res* 181: 395- 408, 2007.

Imamizu H, Uno Y, Kawato M. Internal representations of the motor apparatus - implications from generalization in visuomotor learning. *J Exp Psychol Hum Percept Perform* 21: 1174–1198, 1995.

Izawa J, Shadmehr R. Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput Biol* 7: e1002012, 2011.

Kemp C, Tenenbaum JB. Structured statistical models of inductive reasoning. *Psychol Rev* 116: 20–58, 2009.

Kojima Y, Iwamoto Y, Yoshida K. Memory of learning facilitates saccadic adaptation in the monkey. *J Neurosci* 24: 7531–7539, 2004.

Krakauer JW. Motor learning and consolidation: the case of visuomotor rotation. *Adv Exp Med Biol* 629: 405–421, 2009.

Krakauer JW, Ghilardi MF, Ghez C. Independent learning of internal models for kinematic and dynamic control of reaching. *Nat Neurosci* 2: 1026–1031, 1999.

Krakauer JW, Pine ZM, Ghilardi MF, Ghez C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci* 20: 8916–8924, 2000.

Langbein J, Siebert K, Nürnberg G, Manteuffel G. Learning to learn during visual discrimination in group housed dwarf goats (capra hircus). *J Comp Psychol* 121: 447, 2007.

Lee JY, Schweighofer N. Dual adaptation supports a parallel architecture of motor memory. *J Neurosci* 29: 10396–10404, 2009.

Mackintosh NJ, Little L. Intradimensional and extradimensional shift learning by pigeons. *Psychon Sci* 1969.

Magill RA, Hall KG. A review of the contextual interference effect in motor skill acquisition. *Hum Mov Sci* 9: 241–289, 1990.

Miall RC, Jenkinson N, Kulkarni K. Adaptation to rotated visual feedback: a re-examination of motor interference. *Exp Brain Res* 154: 201–210, 2004.

Mulavara AP, Cohen HS, Bloomberg JJ. Critical features of training that facilitate adaptive generalization of over ground locomotion. *Gait Posture* 29: 242–248, 2009.

Osu R, Hirai S, Yoshioka T, Kawato M. Random presentation enables subjects to adapt to two opposing forces on the hand. *Nat Neurosci* 7: 111–112, 2004.

Pekny SE, Criscimagna-Hemminger SE, Shadmehr R. Protection and expression of human motor memories. *J Neurosci* 31: 13829–13839, 2011.

Redding GM, Wallace B. Strategic calibration and spatial alignment: a model from prism adaptation. *J Mot Behav* 34: 126–138, 2002.

Redding GM, Wallace B. Generalization of prism adaptation. *J Exp Psychol Hum Percept Perform* 32: 1006–1022, 2006a.

Redding GM, Wallace B. Prism adaptation and unilateral neglect: review and analysis. *Neuropsychologia* 44: 1–20, 2006b.

Roller C, Cohen H, Kimball K, Bloomberg J. Variable practice with lenses improves visuo-motor plasticity. *Brain Res Cogn Brain Res* 12: 341–352, 2001.

Saunders JA, Knill DC. Humans use continuous visual feedback from the hand to control both the direction and distance of pointing movements. *Exp Brain Res* 162: 458–473, 2005.

Schrier AM. Learning how to learn: the significance and current status of learning set formation. *Primates* 25: 95–102, 1984.

Schweighofer N, Lee J, Goh H, Choi Y, Kim S, Stewart J, Lewthwaite R, Winstein C. Mechanisms of the contextual interference effect in individuals post-stroke. *J Neurophysiol* 2011.

Seidler RD. Multiple motor learning experiences enhance motor adaptability. *J Cogn Neurosci* 16: 65–73, 2004.

Seidler RD. Older adults can learn to learn new motor skills. *Behav Brain Res* 183: 118–122, 2007.

Shabbott BA, Sainburg RL. On-line corrections for visuomotor errors. *Exp Brain Res* 195: 59–72, 2009.

Shadmehr R, Brashers-Krug T. Functional stages in the formation of human long-term motor memory. *J Neurosci* 17: 409–419, 1997.

Shea JB, Morgan RL. Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *J Exp Psychol Hum Learn* 5: 179, 1979.

Sing GC, Joiner WM, Nanayakkara T, Brayanov JB, Smith MA. Primitives for motor adaptation reflect correlated neural tuning to position and velocity. *Neuron* 64: 575–589, 2009.

Sing GC, Smith MA. Reduction in learning rates associated with anterograde interference results from interactions between different timescales in motor adaptation. *PLoS Comput Biol* 6: e1000893, 2010.

Smeets JBJ, van den Dobbelsteen JJ, de Grave DDJ, van Beers RJ, Brenner E. Sensory integration does not lead to sensory calibration. *Proc Natl Acad Sci USA* 103: 18781, 2006.

Smith MA, Ghazizadeh A, Shadmehr R. Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol* 4: e179, 2006.

Tenenbaum JB, Griffiths TL. Generalization, similarity, and bayesian inference. *Behav Brain Sci* 24: 629–640, 2001.

Thoroughman KA, Shadmehr R. Learning of action through adaptive combination of motor primitives. *Nature* 407: 742–747, 2000.

Warren JM. The comparative psychology of learning. *Annu Rev Psychol* 16: 95–118, 1965.

Weiner MJ, Hallett M, Funkenstein HH. Adaptation to lateral displacement of vision in patients with lesions of the central nervous system. *Neurology* 33: 766–772, 1983.

Welch RB, Bridgeman B, Anand S, Browman KE. Alternating prism exposure causes dual adaptation and generalization to a novel displacement. *Percept Psychophys* 54: 195–204, 1993.

Wigmore V, Tong C, Flanagan JR. Visuomotor rotations of varying size and direction compete for a single internal model in motor working memory. *J Exp Psychol Hum Percept Perform* 28: 447–457, 2002.

Wolpert D, Kawato M. Multiple paired forward and inverse models for motor control. *Neural Netw* 11: 1317–1329, 1998.

Zarahn E, Weston GD, Liang J, Mazzoni P, Krakauer JW. Explaining savings for visuomotor adaptation: linear time-invariant state-space models are not sufficient. *J Neurophysiol* 100: 2537–2548, 2008.