

The origins of anterograde interference in visuomotor adaptation

Gonzalo Lerner¹, Scott Albert², Pedro A. Caffaro¹, Jorge I. Villalta¹, Florencia Jacobacci¹,
Reza Shadmehr², and Valeria Della-Maggiore¹

Gonzalo Lerner and Scott Albert have contributed equally to this work.

¹ Universidad de Buenos Aires. Facultad de Medicina. Departamento de fisiología y biofísica. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Instituto de Fisiología y Biofísica Houssay, Buenos Aires, Argentina.

² Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, USA.

Corresponding author: V Della-Maggiore. IFIBIO Houssay, Department of Physiology and Biophysics, Paraguay 2155, 7th floor, School of Medicine, University of Buenos Aires, Buenos Aires, 1121, ARGENTINA. Tel & fax: 5411 5950 9500 x 2132, email: vdellamaggiore@fmed.uba.ar

1

2 **ABSTRACT**

3 Anterograde interference refers to the negative impact of prior learning on the propensity
4 for future learning. Previous work has shown that subsequent adaptation to two
5 perturbations of opposing sign, *A* and *B*, impairs performance in *B*. Here, we aimed to
6 unveil the mechanism at the basis of anterograde interference by tracking its impact as a
7 function of time through a 24h period. We found that the memory of *A* biased performance
8 in *B* for all time intervals. Conversely, learning from error was hindered up to 1h following
9 acquisition of *A*, with release from interference occurring at 6h. These findings suggest
10 that poor performance induced by prior learning is driven by two distinct mechanisms: a
11 long-lasting bias that acts as a prior and hinders the initial level of performance, and a
12 short-lasting learning impairment that originates from a reduction in error-sensitivity. Our
13 work provides insight into the timeline of memory stabilization in visuomotor adaptation.

14

15

16

17 INTRODUCTION

18 We gain robustness through adaptation: in the face of environmental and/or internal
19 perturbations, adaptation maintains the precise control of elementary movements like
20 reaching and saccades. Like other types of learning, adaptation may lead to interference
21 or facilitation depending on the level of congruency of sequentially learned materials.
22 Facilitation of learning is commonly referred to as savings, a process by which subsequent
23 exposure to the same perturbation results in faster learning (Krakauer, 2009). In contrast,
24 successive adaptation to opposing perturbations (e.g., rotation *A* followed by rotation *B*)
25 may lead to a deficit in the learning of *B*. This phenomenon, known as anterograde
26 interference, has been reported in visuomotor and force-field adaptation paradigms when
27 successively adapting to conflicting perturbations within the same reaching task (Brashers-
28 Krug, Shadmehr, & Bizzi, 1996; Sing & Smith, 2010; Tong & Flanagan, 2003; Wigmore,
29 Tong, & Flanagan, 2002). Yet, there is currently no consensus on whether anterograde
30 interference is transient or long lasting. In fact, whereas some studies suggest that
31 anterograde effects may last less than a few hours (e.g., Brashers-Krug et al., 1996;
32 Thoroughman & Shadmehr, 1999), others appear to point to a long lasting impact in the
33 time scale of days (Caithness et al., 2004; Miall, Jenkinson, & Kulkarni, 2004). It has even
34 been suggested that anterograde interference may be stronger than retrograde
35 interference (Caithness et al., 2004; Miall et al., 2004; Sing & Smith, 2010), masking the
36 effect of interest in retrograde protocols aimed at unveiling the time course of memory
37 consolidation (Miall et al., 2004).

38 This lack of consensus may be partly due to the manner in which anterograde
39 interference is quantified (Sing, Joiner, Nanayakkara, Braynov, & Smith, 2009). Previous
40 studies estimated the amount of interference predominantly based on the initial level of
41 performance, computed by averaging through the first trials of the learning curve. This
42 empirical measure does not discriminate between changes in learning rate and retention.

43 That is, initial performance in *B* is a mixture of how much the subject has retained what
44 they learned in *A*, and how much they can learn from errors experienced in *B*. If
45 anterograde interference arises from impairment in the ability to learn, one would expect
46 that prior exposure to *A* would reduce the learning rate in *B*. Yet, with the exception of Sing
47 & Smith (2010) no study that we are aware of has focused on the rate of learning as the
48 fundamental measure of anterograde interference.

49 Here, we aimed to unveil the origins of anterograde interference by varying the time
50 interval elapsed between adaptation to opposing rotations through a 24 h period. Our
51 experimental approach allowed us to estimate the contribution of a prior memory
52 independently from the ability to learn. The recruitment of a large number of subjects ($n =$
53 93) allowed us to analyze and contrast individual measures of learning during adaptation
54 to *A* and *B*, when the two events were separated by 5 min, 1 h, 6 h and 24 h. In addition,
55 we analyzed cycle-by-cycle learning using a state-space model (Albert & Shadmehr, 2018;
56 Cheng & Sabes, 2006; Donchin, Francis, & Shadmehr, 2003; Ethier, Zee, & Shadmehr,
57 2008; Smith, Ghazizadeh, & Shadmehr, 2006). This allowed us to identify the impact of
58 anterograde interference on three variables: initial state, error-sensitivity, and retention.

59 We found that poor performance observed when *A* and *B* are learned successively
60 is driven by two distinct phenomena operating on different time scales: the influence of a
61 prior memory that biases initial behavior, and an impairment in the ability to learn from
62 errors in the new context. Whereas the former appears to impact behavior on a scale of
63 days, the latter resolves within a 6 h period.

64

65 **MATERIALS AND METHODS**

66 ***Participants***

67 Ninety-three healthy volunteers (33 males; ages: mean \pm std. dev. 24 ± 4 years old) with no
68 known history of neurological or psychiatric disorders were recruited from the School of

69 Medicine of the University of Buenos Aires. All subjects were right handed as assessed by
70 the Edinburgh handedness inventory (Oldfield, 1971). The experimental procedure was
71 approved by the local Ethics Committee and carried out according to the Declaration of
72 Helsinki.

73

74 ***Experimental Paradigm***

75 Subjects were seated in a comfortable chair and performed a center-out-back task using a
76 joystick operated with the thumb and index fingers of their right hand. Visual information
77 was presented on a computer screen. The right elbow laid comfortably on an armrest and
78 the wrist laid on a structure that fixed the joystick over a desktop. Subjects were told to
79 maintain the same wrist posture across experimental sessions. Vision of the hand was
80 occluded throughout the study.

81 At the beginning of each trial, we displayed one of 8 potential targets (0.4 cm
82 diameter, placed 2 cm from the start point and concentrically located 45° from each other)
83 on a computer screen. Joystick position was represented on the screen with a grey cursor
84 of the same size as the target. The gain of the joystick was set to discourage subjects from
85 correcting their movements online. Specifically, a displacement of 1.44 cm of the tip of the
86 joystick moved the cursor on the screen by 2 cm. On average, movement time for correct
87 trials was 125.5 ± 26.6 ms (mean \pm 1 std. dev.), providing little or no opportunity for within-
88 movement corrections based on visual feedback. Participants were instructed to make a
89 shooting movement through the target, as fast as possible, starting at target onset. There
90 were 8 trials per cycle (one for each target) and 11 cycles per block. The order of target
91 presentation was randomized within each cycle.

92 Two types of trials were presented throughout the experimental session (Fig. 1A).
93 During *null* trials, participants performed shooting movements in the absence of a
94 perturbation. During *perturbed* trials, a counterclockwise (CCW, labeled as perturbation A),

95 or a clockwise (CW, labeled as perturbation *B*) visual rotation of 30° was applied to alter
96 the trajectory of the cursor.

97 Feedback about the subject's movement was provided on each trial via the color of
98 the cursor, which varied along a gradient between red (miss) and green (hit). Furthermore,
99 subjects had a limited amount of time to complete the movement after the appearance of
100 the target. If the elapsed time exceeded 900 ms, the trial was aborted and the cursor was
101 turned red until the next trial. Target hits with error < 2.5° were rewarded with a simulated
102 sound of an animated explosion. The total score (hit percentage) was displayed on the
103 screen at the end of each block. Subjects were instructed to try to maximize this score
104 throughout the experiment. The task was programmed using MATLAB's Psychophysics
105 Toolbox, Version 3 (Brainard, 1997).

106

107 ***Experimental Procedure***

108 Figure 1A illustrates the experimental design. Participants were randomly assigned to one
109 of four experimental groups or a control group. The experimental groups (Fig. 1A)
110 performed one block (11 cycles) of null trials followed by six blocks (66 cycles) of CCW
111 perturbed trials (perturbation *A*). After a variable time interval, they performed six blocks
112 (66 cycles) of the CW perturbation (perturbation *B*). The four experimental groups were
113 distinguished by the amount of time that separated the two rotations: 5 min ($n = 16$), 1 h (n
114 $= 20$), 6 h ($n = 19$), and 24 h ($n = 18$). This variation in the period between perturbations *A*
115 and *B* allowed us to assess how the passage of time impacted on the initial level of
116 performance in *B* (first cycle), as well as on each subject's ability to adapt to *B*.

117 A group of subjects ($n = 20$) experienced only the *B* perturbation. This control group
118 served two purposes. First, it was critical for our analysis of anterograde interference,
119 serving as our benchmark for performance in *B* without any potential influence of learning
120 in *A*. Second, given that subjects always learned *A* before *B*, this group was key in ruling

121 out an order effect. Control subjects performed 1 block (11 cycles) of null trials followed by
122 6 blocks (66 cycles) of *B*.

123

124 ***Data post-processing***

125 For each trial, the pointing angle was computed based on the angle of motion of the
126 joystick relative to the line segment connecting the start and target positions. Trials in
127 which pointing angles exceeded 120° or deviated by more than 45° from the median of the
128 trials for each cycle were excluded from further analysis (1.6% of all trials). After this
129 processing, the trial-by-trial data were converted to cycle-by-cycle time series by
130 calculating the median pointing angle in each 8-trial cycle for each subject. Unless
131 otherwise noted, the cycle-by-cycle data were used for each analysis reported in this work.

132

133 ***Model-free data analysis***

134 We empirically quantified each subject's learning rate in *A* and *B* by fitting a single
135 exponential function (Eq. 1) to the pointing angle corresponding to the *A* and *B* periods.

$$136 \quad y(t) = \alpha \exp(-\beta t) + c \quad (\text{Eq. 1})$$

137 Here $y(t)$ represents the pointing angle on cycle t . The first cycle of the rotation was
138 represented by $t = 0$. The exponential fits included three parameters. The parameters α
139 and c determine the initial bias and the asymptote of the exponential, respectively. The
140 parameter β represents the learning rate of the subject. We constrained the relationship
141 between α and c to force the exponential fit to intersect subject behavior at time step $t = 0$.
142 Therefore, the exponential function had only two free parameters; the third was fixed by
143 the initial level of subject performance. We fit one exponential function to the 66 cycles of
144 the *A* rotation and another one to the 66 cycles of the *B* rotation (Figure 1). Each period

145 was fit using the *fmincon* function (MATLAB 2018a) to minimize the squared error between
146 subject behavior and the exponential fit.

147 Although the exponential function closely approximates the decay of motor error
148 during adaptation to a single perturbation, its learning rate parameter reflects a mixture of
149 cycle-by-cycle forgetting and error-based learning. This potentially confounds our analysis
150 of interference because during the *A* perturbation, the direction of forgetting (always
151 towards baseline performance) opposes the direction of error-based learning. However,
152 during the *B* perturbation, an initial bias in the performance of the experimental groups
153 towards *A* causes forgetting and error-based learning to act in the same direction. This
154 relationship switches once subjects pass the “zero point” of baseline performance: here
155 retention and error-based learning again oppose one another. These considerations
156 illustrate the difficulties inherent in using exponential fits to disambiguate the differential
157 effects learning and forgetting may have on behavior.

158

159 ***State-Space Model***

160 To better quantify subject performance in *A* and *B*, we used a state-space model that
161 dissociates the effect of cycle-to-cycle learning from forgetting while appreciating initial
162 biases in learning.

163 When people perform a movement that produces an unexpected result, they learn
164 from their movement error and retain part of this learning over time. In other words,
165 behavior during sensorimotor adaptation can be described as a process of error-based
166 learning and trial-by-trial forgetting (Donchin et al., 2003; Smith et al., 2006; Thoroughman
167 & Shadmehr, 2000). State-space models of learning consider how the behavior of a
168 learner changes due to trial-by-trial error-based learning, and decay of memory due to the
169 passage of time (i.e., trials). To examine the anterograde interference of *A* on *B*, we fit a
170 single module state-space model to the empirical data. This allowed us to ascribe any

171 differences in performance during the B period to meaningful quantities: sensitivity to error,
172 forgetting rate, and initial state.

173 We imagined that the state of the learner (the internal estimate of the visuomotor
174 rotation) changed from one cycle to the next, due to error-based learning and partial
175 forgetting, according to Eq. (2).

$$176 \quad \mathbf{x}^{(t+1)} = a\mathbf{x}^{(t)} + b\mathbf{e}^{(t)} + \varepsilon_x^{(t)} \quad (\text{Eq. 2})$$

177 Here $\mathbf{x}^{(t)}$ represents the state of the learner on cycle t . The parameter a is a retention
178 factor that encapsulates how well the subject retains a memory of the perturbation from
179 one cycle to the next. The parameter b represents sensitivity to error and determines the
180 rate at which each subject learns from error. The error sensitivity is multiplied by the visual
181 error $\mathbf{e}^{(t)}$ between the pointing angle and target. The change in state from one cycle to the
182 next is corrupted by state noise $\varepsilon_x^{(t)}$ which we assumed to be Gaussian with mean zero
183 and variance equal to σ_x^2 .

184 The internal state of the subject is not a measurable quantity. Rather, on each
185 cycle, the motor output of the subject is measured. We imagine that the motor output
186 directly reflects the internal state but is corrupted by motor execution noise according to
187 Eq. (3).

$$188 \quad \mathbf{y}^{(t)} = \mathbf{x}^{(t)} + \varepsilon_y^{(t)} \quad (\text{Eq. 3})$$

189 As with our exponential fit of Eq. (1), here $\mathbf{y}^{(t)}$ represents the subject's pointing angle on
190 cycle t . We assumed that the motor execution noise $\varepsilon_y^{(t)}$ corrupting the reaching movement
191 was Gaussian with mean zero and variance equal to σ_y^2 .

192 We fit the state-space model to cycle-by-cycle single subject behavior using the
193 Expectation Maximization (EM) algorithm (Albert & Shadmehr, 2018). The algorithm

194 identified the parameter set that maximized the likelihood of observing each sequence of
195 subject pointing angles given the parameters and structure of our state-space model. This
196 parameter set contained 6 parameters: the retention factor a , error sensitivity b , state
197 noise variance σ_x^2 , motor noise variance σ_y^2 , and two parameters describing the initial
198 state of the learner. We modeled the initial state of the learner as a normally distributed
199 random variable with mean x_1 and variance σ_1^2 . The parameter x_1 served as our estimate
200 of the initial bias of the learner.

201 To fit the model, we started the EM algorithm from 5 different initial parameter sets,
202 performed 100 iterations of the algorithm (Albert & Shadmehr, 2018), and selected the
203 parameter set with the greatest likelihood. We fit our state-space model to single subject
204 behavior separately for the A and B periods. For the A period, we fit the 77 cycles
205 encompassing the first 11 null cycles and the following 66 CCW rotation cycles (Fig. 1).
206 We fit the initial null trials along with the perturbation trials to increase confidence in the
207 model parameters. For the B period, we fit the 66 cycles encompassing the CW rotation
208 (Fig. 1).

209

210 *Validation of the single state-space model*

211 Our primary analysis assumed that learning could be represented using a single adaptive
212 state. For a single state system, impairment in the learning rate in B requires that the
213 learning system (*i.e.*, the model parameters) has changed from the A to the B period. Prior
214 accounts of anterograde interference considered how an impairment of learning might
215 arise from the emergent properties of a two-state system (Sing & Smith, 2010). A two-state
216 system might show a change in learning rate during the B period even if the system has
217 not changed, simply due to differing initial biases in the underlying adaptive states of the
218 system.

219 To validate the choice of a single state model, in a set of mathematical control
220 studies, we tested if our findings were also consistent with the predictions of a two-state
221 framework. For this analysis, we fit a two-state model of learning to the *A* and *B*
222 sequences of subject pointing angles. The two-state model of learning is equivalent to the
223 single state model of learning, with the exception that learning is described as the
224 combined output of two parallel adaptive states, a fast learning state and a slow learning
225 state. The states evolve over trials according to Eq. (4).

$$\begin{aligned}x_s^{(t+1)} &= a_s x_s^{(t)} + b_s e^{(t)} + \varepsilon_x^{(t)} \\x_f^{(t+1)} &= a_f x_f^{(t)} + b_f e^{(t)} + \varepsilon_x^{(t)}\end{aligned}\tag{Eq. 4}$$

227 Here, the slow and the fast states are represented by the quantities x_s and x_f ,
228 respectively. As with the single state model (Eq. 2) each state changes due to forgetting
229 (described by its retention factor a) and error-based learning (described by its error
230 sensitivity b). These internal estimates of the perturbation are additively combined to
231 determine motor behavior according to Eq. 5.

$$y^{(t)} = x_s^{(t)} + x_f^{(t)} + \varepsilon_y^{(t)}\tag{Eq. 5}$$

233 We fit this two-state model of learning to subject behavior during the *A* and *B* periods using
234 the EM algorithm (Albert & Shadmehr, 2018). The algorithm identified the parameter set
235 that maximized the likelihood of observing each sequence of subject pointing angles. We
236 fit the model to the same cycles in *A* and *B* described for the single state model fits. To fit
237 the model, we started the EM algorithm from 20 different initial parameter sets, performed
238 250 iterations of the algorithm, and selected the parameter set with the greatest likelihood.
239 The model parameter set consisted of 9 variables: slow and fast retention factors a_s and a_f
240 , slow and fast error sensitivities b_s and b_f , the variances of state evolution and motor
241 execution, and three parameters for the initial state of the learner. We modeled the initial

242 fast and slow states as normally distributed random variables with mean $x_s^{(1)}$ and $x_f^{(1)}$, and
243 variance σ_1^2 . Each model was fit under the linear constraints $b_f > b_s$ and $a_s > a_f$. These
244 constraints enforce that the slow state learns more slowly than the fast state, but also
245 retains its memory better from one trial to the next (Smith et al., 2006).

246 We compared the single state model and two-state model in their abilities to
247 describe subject behavior. To compare these models, we computed the Bayesian
248 Information Criterion (BIC) according to Eq. 6.

$$249 \quad \quad \quad BIC = k \log(n) - 2 \log(L_{\max}) \quad \quad \quad (\text{Eq. 6})$$

250 Here k represents the number of model parameters (6 for the single state model, 9
251 for the two-state model), n represents the number of data points, and L_{\max} represents the
252 maximum likelihood for the model fit obtained using the EM algorithm. To obtain a single
253 estimate of BIC for each subject, we averaged the BIC over the A and B periods. To
254 quantify the evidence for each model, we compared the BIC distributions for the single
255 state and two-state models for all subjects in the experimental groups using a paired t -test.

256 Finally, to test the hypothesis (Sing & Smith, 2010) that a learning impairment in B
257 could trivially arise from the interplay between fast and slow state dynamics (rather than an
258 actual impairment in the learning system) we simulated behavior during the B period using
259 a two-state model, and compared this behavior to actual subject behavior. We reasoned
260 that if the learning system did not differ across the A and B periods, the B period behavior
261 should be predicted by the model parameters fit to the A period. To test this idea, we used
262 the model parameters fit to the A period to simulate individual subject behavior during the
263 B period, focusing specifically on the 5 min group that demonstrated the greatest amount
264 of interference.

265 To simulate behavior in the B period, we calculated the expected value of subject
266 performance by removing the noise terms in *Eqs.* 4 and 5. Due to the passage of time,
267 behavior at the start of the B period exhibited decay relative to the performance at the end
268 of the A period. We accounted for this forgetting using the initial biases in the fast and slow
269 states. To do this, we calculated the amount of forgetting that occurred from the last cycle
270 in A to the first cycle in B , as a percentage. We then estimated the final levels of the slow
271 and fast states in A by simulating behavior in A using *Eqs.* 4 and 5, without any noise
272 terms. For our simulation of the B period, we set the initial fast and slow state levels to the
273 final levels in A , scaled down by the forgetting percentage.

274 We calculated the rate of learning using our exponential model (*Eq.* 1) of behavior
275 both for the actual behavior recorded in B and the behavior simulated using the two-state
276 model. We compared these rates using a paired t -test to determine how well the two-state
277 parameters from the A period accounted for the interference observed in the B period.

278

279 **Statistical assessment**

280 Statistical differences were assessed at the 95% level of confidence. Prior to statistical
281 testing, outlying parameter values were detected and removed based on a threshold of
282 three median absolute deviations from the group median. For cases where our variables of
283 interest did not fail tests for normality and equality of variance, we used a one-way ANOVA
284 for our statistical testing. In cases where the statistical distributions failed tests for both
285 equal variance across groups (Bartlett's test) and normality (Shapiro-Wilk test) we used
286 the Kruskal-Wallis test to detect non-parametric differences across experimental groups. In
287 cases where our statistical tests indicated a significant effect of group ($p < 0.05$), we used
288 either Tukey's test (following one-way ANOVA) or Dunn's test (following Kruskal-Wallis) for
289 post-hoc testing. For the latter, pairwise tests of all experimental groups were conducted
290 against the control group and Bonferroni corrected. In cases where one-way ANOVA was

291 used for statistical testing, complementary figures depict the mean statistical quantity for
292 each group as well as the standard error of the mean, calculated assuming a normal
293 distribution. In cases where Kruskal-Wallis was used for statistical testing, complementary
294 figures depict the median statistical quantity for each group as well as the standard error of
295 the median (estimated with bootstrapping). When comparing mean values against zero, a
296 one-sample t-test test was used followed by the Bonferroni correction for multiple
297 comparisons.

298

299 **RESULTS**

300 When people adapt to perturbation *A*, and then switch to the opposite perturbation *B*,
301 performance in *B* appears impaired (Brashers-Krug et al., 1996; Braun, Aertsen, Wolpert,
302 & Mehring, 2009; Caithness et al., 2004; Shadmehr & Brashers-Krug, 1997; Tong &
303 Flanagan, 2003). Pinpointing the origin of this behavioral deficit is difficult because
304 performance in *B* may reflect two different processes: the level of retention of the memory
305 of *A*, and the ability to learn *B*. In addition, these factors may vary independently as a
306 function of time. Our study aimed to dissociate between these two factors by varying the
307 time interval elapsed between *A* and *B* as subjects adapted to conflicting visuomotor
308 rotations.

309 On each trial, subjects moved a joystick to displace a cursor to one of 8 targets. On
310 average, movement time for correct trials was 125.5 ± 26.6 ms (mean \pm 1 std. dev.),
311 providing little or no opportunity for within-movement corrections. All groups initially trained
312 in a baseline period of null trials (no perturbation) followed by adaptation to perturbation *A*
313 (Figure 1A). After completion of training in *A*, subjects waited for a specific amount of time
314 (5 min, 1 h, 6 h or 24 h), and then were exposed to perturbation *B*. Figure 1B shows the
315 pointing angle during null trials (cycles 1 to 11), learning of *A* (cycles 12 to 77) and learning
316 of *B* (cycles 78 to 143) for each of the experimental groups (black curves) and the control

317 group (red curve). Pointing angle refers to angle of motion of the joystick relative to the line
318 segment connecting the start and target positions. As expected, the pointing angle during
319 null trials was close to zero. During exposure to perturbation *A*, subjects shifted their
320 pointing angle gradually, asymptotically approaching -30° (Shmuelof et al., 2015). After
321 adapting to *A* and waiting the assigned time, subjects returned and were exposed to
322 perturbation *B*.

323 How did learning of *A* impact performance in *B*? We quantified the initial level of
324 performance in *B* as the mean pointing angle during the first cycle of adaptation for each
325 group (Fig. 2A). Given that little or no learning is expected to take place in one cycle (1
326 cycle = one trial per target), this measure allowed us to estimate the recall of *A*. The initial
327 level of performance in *B* was biased towards *A*, and decayed as a function of time (Fig.
328 2A, one-way ANOVA, $F_{(69,3)} = 3.37$, $p = 0.029$; Tukey's test, 5 min marginally different from
329 6 h with $p = 0.073$, 5 min different from 24 h with $p = 0.029$, all other comparisons have $p >$
330 0.358). Notably, even at 24 h the memory of *A* remained strong, exhibiting nearly 50%
331 retention (one-sample t-test against zero with Bonferroni correction: $p < 0.001$ for all
332 groups). This observation is consistent with the presence of a lingering memory of *A*
333 (Shadmehr & Brashers-Krug, 1997; Thoroughman & Shadmehr, 1999).

334 In summary, during initial performance in *B* the movements were strongly
335 influenced by the presence of a memory of *A*. This memory decayed with time, but was
336 still present at 24 h.

337

338 ***Model-free analysis***

339 In order to assess if having learned *A* also impaired the ability to learn *B*, we fit the motor
340 output for each subject and each group during adaptation to *A* and *B* with an exponential
341 function (Eq. 1). We found that during the *A* period there was no difference in the learning

342 rates across the four experimental groups (Kruskal-Wallis, $X^2(62) = 4.75$, $p = 0.19$).
343 Therefore, the various groups were indistinguishable during learning of *A*.

344 Did prior exposure to *A* interfere with the rate at which *B* was acquired? To quantify
345 the ability to learn *B*, we statistically compared the rate of change in performance for each
346 experimental group during adaptation to *B* with that of the control group (Figure 2B). Non-
347 parametric testing revealed a significant effect of group on the ability to learn *B* (Fig. 2B;
348 Kruskal-Wallis, $X^2(80) = 10.84$, $p = 0.029$). Post-hoc comparison between each
349 experimental group and the control group identified a significant difference at 5 min and 1
350 h (Dunn's test with Bonferroni correction, 5 min different from control with $p = 0.044$, 1 h
351 different from control with $p = 0.024$), that disappeared by 6 h (6 h and 24 h not different
352 from control with $p > 0.952$). This temporal pattern in the impairment of motor learning is
353 consistent with anterograde interference (Krakauer, 2009).

354 To visualize differences in the learning of *B*, we artificially aligned the performance
355 of the control group to each experimental group by shifting the control learning curve along
356 the time axis (Figure 3). This procedure makes use of a fundamental property of
357 exponential functions:

358
$$\alpha_1 \exp(-\beta t) = \alpha_2 \exp(-\beta(t - t_0)), \text{ where } t_0 = \ln\left(\frac{\alpha_1}{\alpha_2}\right) / \beta \quad (\text{Eq. 7})$$

359

360 If two exponentials start at different points (α_1 and α_2) but share the same decay rate β ,
361 then shifting one in time by t_0 will result in complete overlap of the two functions.

362 To optimally align the behavior of the control with that of the experimental groups,
363 we fit an exponential function (Eq. 1) to the median behavior of each group. We next
364 calculated the cycle on which the predicted behavior would intersect a pointing angle of 0°
365 for each group, and shifted the behavior of the control to match that of each experimental

366 group at $Y = 0$ (Fig 3). This temporal displacement points towards an impairment in the
367 learning rate of the 5 min and the 1 h groups.

368 Finally, to rule out the possibility that our results may be explained by an order
369 effect (subjects always learned the CCW rotation before the CW rotation), we statistically
370 compared the rate of learning of the control group with those of the experimental groups
371 during learning in *A*. No differences were found between the learning rates of *A* and *B*
372 control (Kruskal-Wallis, $X^2(78) = 5.53$, $p = 0.237$). Therefore, the control condition rules out
373 the possibility that our results are explained by the order in which the perturbations were
374 learned.

375 In summary, we conclude that while the lingering memory of *A* caused the starting
376 point of the learning process to be strongly biased in all experimental groups, the learning
377 process itself was impaired at 5 min and 1 h only. Release from interference, determined
378 as the time point at which the rate of learning resembled that of the control group, took
379 place about 6 h post adaptation.

380

381 ***State-space model***

382 The exponential model we employed for our empirical analysis implicitly assumed that the
383 rate of learning remained constant across trials. For the *B* period, this assumption is
384 unlikely to be true because initially, learning from the errors induced by *B* is aided by
385 forgetting of the memory of *A*. That is, as the *B* period starts, performance falls toward
386 baseline, and the rate of this fall is due to two processes: forgetting of *A*, and learning from
387 error in *B*. During this period, forgetting and learning act in the same direction. However,
388 once the performance crosses baseline levels, the influence of memory decay on behavior
389 is in the opposite direction to learning from error. State-space models of learning
390 disentangle these processes of forgetting and learning. For this reason, we fit a state-

391 space model to the data from individual subjects separately during the *A* and *B* periods
392 (Eq. 2 and 3).

393 The state-space model assumes that learning is governed by two processes: a
394 process that learns from error, and a process that retains a fraction of that memory from
395 one trial to the next. The state-space model closely tracked the observed behavior (Fig. 4).
396 To quantify the model's goodness-of-fit, we computed the fraction of each subject's
397 behavioral variance accounted for by our model fit (R^2). To measure this coefficient of
398 determination, we computed the expected value of the behavior predicted by our
399 stochastic model (Eqs. 2 and 3) and compared this model prediction to each individual
400 subject's data. We found that across subjects, our model accounted for approximately 81.4
401 $\pm 8.4\%$ (mean ± 1 std. dev.) of the variance in subject behavior. We repeated this analysis
402 at the group level, where noise in the process of learning (Eq. 2) and production of a
403 movement (Eq. 3) is smoothed over subjects. For each group, we computed the median
404 behavior (Fig. 4A, black curves for experimental groups, red curve for control), the median
405 behavior predicted by our model (Fig. 4A, blue curves for experimental groups, green for
406 control), and then the coefficient of determination for these two time-courses. At the group
407 level, the model accounted for 96.0 to 98.2% of the variance in median subject behavior.

408 After validating our model, we next considered how anterograde interference could
409 be quantified at the level of three different processes that could affect performance in *B*:
410 (1) memory of *A*, (2) cycle-to-cycle forgetting rate in *B*, and (3) learning from error in *B*.
411 These three processes are represented separately by three specific model parameters: (1)
412 the initial state of the learner in *B*, (2) the retention factor, and (3) the error-sensitivity.

413 Unsurprisingly, the initial state of the learner in *B* (Fig. 4B) closely followed our
414 empirical estimate of the initial level of performance in *B* (Fig. 2A). As the interval between
415 *A* and *B* increased, the initial state of the learner in *B*, *i.e.*, the amount of the *A* memory
416 retained over time, decreased (one-way ANOVA, $F_{(69,3)} = 3.89$, $p = 0.013$; Tukey's test, 5

417 min different than 24 h with $p = 0.020$, marginal differences between 5 min and 6 h as well
418 as 1 h and 24 h with $p = 0.097$, all other comparisons have $p > 0.347$). However, despite
419 this temporal decay, all groups retained at least 50% of the *A* memory (one-sample t-test
420 with Bonferroni correction, all groups $p < 0.001$). Therefore, impairment of performance in
421 *B* was in part caused by a lingering memory of *A* that did not fully decay even after 24 h.

422 To what extent was the impairment in *B* driven by changes in the rate of cycle-by-
423 cycle memory retention and learning from error? Similar to our empirical analysis, we first
424 confirmed that the experimental groups did not differ in performance during the *A* period.
425 That is, there was no difference in the retention factor (Kruskal-Wallis, $X^2(66) = 0.53$, $p =$
426 0.912) or error sensitivity (Kruskal-Wallis, $X^2(65) = 1.16$, $p = 0.763$) across the
427 experimental groups during adaptation to *A*. Furthermore, we found no difference in the
428 retention factor during learning in *B* for any of the experimental groups, including the
429 control group (Fig. 4C; Kruskal-Wallis, $X^2(79) = 5.66$, $p = 0.226$). In contrast, error
430 sensitivity was affected by prior learning in a manner consistent with anterograde
431 interference (Fig. 4D; Kruskal-Wallis, $X^2(83) = 14.47$, $p = 0.006$). Post-hoc tests against the
432 control group unveiled a significant reduction in error sensitivity at 5 min and 1 h but not at
433 longer time intervals (Dunn's test with Bonferroni correction, 5 min different from control
434 with $p = 0.008$, 1 h different from control with $p = 0.004$, 6 h and 24 h not different from
435 control with $p > 0.132$).

436 In summary, our state-space model pointed to a similar conclusion drawn from our
437 empirical findings. Prior exposure to *A* resulted in a bias in the initial state of *B* that
438 persisted through 24 h. Moreover, prior exposure produced a reduction in error sensitivity,
439 but this effect was short lasting: we found no evidence for it when the time between *A* and
440 *B* was 6 h or more. Therefore, differences in performance in *B* for any timescale greater
441 than 1 h were likely related to a prior memory of *A* and not to a deficit in learning.

442

443 ***Validation of the computational approach***

444 An earlier account of anterograde interference (Sing & Smith, 2010) investigated how the
445 emergent properties of a learning system composed of two parallel adaptive states (Smith
446 et al., 2006) could demonstrate impaired learning in *B* after the experience of *A*, even if
447 there was no change in the learning rate of either state. Two-state models of learning posit
448 that adaptation is supported by two parallel learning processes, a slow process (Fig. 5A,
449 red) that learns little from error but exhibits strong retention over trials, and a fast process
450 (Fig. 5A, green) that learns greatly from error but has poor ability to retain its memory over
451 trials. Sing and Smith (2010), demonstrated that if a two-state learning system is exposed
452 to the *A* perturbation (+30°), followed by the opposing *B* perturbation (-30°), learning of *B*
453 would appear to be slowed (Fig. 5A, compare *B* curve with the naïve *A* curve shown in
454 blue) because the slow state is heavily biased towards *A* at the start of *B*. That is, a two-
455 state system can exhibit anterograde interference (a slowing of the overall adaptation
456 rate), despite the fact that the individual learning rates have not changed from *A* to *B*.
457 Could the reduction in error sensitivity we report in our analysis of anterograde
458 interference be explained by a two-state system whose constitutive parameters do not
459 differ across the *A* and *B* periods?

460 To answer this question, we first mathematically compared the likelihood that our
461 data was better described by a two-state system rather than a single state system. We fit a
462 two-state model of learning (Albert & Shadmehr, 2018) separately for the *A* and *B* periods
463 and compared the likelihood of this model with that of the single state model explored in
464 our primary analysis using the Bayesian Information Criterion (BIC). At the level of
465 individual subjects, we found that a two-state model of learning was justified in only 5 of
466 the 73 subjects in the experimental groups (Fig. 5B, red lines). Therefore, in this task, the
467 measured behavior was better described by a single state model of learning (Fig. 5C,
468 lower BIC for single state model, paired t-test, $t(92) = 16.133$, $p < 0.001$).

469 We next asked if a two-state model fit to the *A* period of behavior would produce
470 the pattern of interference we measured in *B*, as described by Sing and Smith (2010). For
471 this analysis, we focused on the 5 min group that exhibited the most significant amount of
472 anterograde interference (Fig. 2B). Using the two-state model parameters fit to the *A*
473 period of behavior, we simulated subject performance during the *B* period. We found that
474 the dynamics of learning predicted by a two-state model whose learning rates were
475 unchanged from *A* to *B* (Fig. 5D, red), did not visually resemble the pattern of interference
476 we measured in the 5 min experimental group (Fig. 5D, black). Indeed, the exponential
477 rate of improvement in *B* exhibited by the two-state model simulation was a rather poor
478 predictor of the actually measured behavior as exhibited by the 5 min group (Fig. 5E,
479 paired t-test, $t(15) = 4.235$, $p < 0.001$). That is, our subjects learned much slower in the *B*
480 perturbation than predicted by a two-state model in which parameters remain constant in *A*
481 and *B*. This result gives us some assurance that the reduction in error sensitivity we
482 observed during the *B* period was not due to a trivial property of a two-state learning
483 system. Rather, it appears that following adaptation to *A*, the motor system was
484 fundamentally impaired in its ability to learn *B*.

485

486 **DISCUSSION**

487 How do motor memories influence one another? In this work, we studied the expression of
488 anterograde interference in visuomotor adaptation by varying the time elapsed between
489 learning opposing perturbations. We examined the impact of prior learning on the initial
490 level of performance as well as the rate of learning over the time course of 5 min through
491 24 h. We found that these two parameters behaved very differently as a function of time.
492 On one hand, adaptation in *A* biased the initial level of performance in *B*. Although the
493 magnitude of this effect decreased with time, it remained strong at 24 h. On the other
494 hand, prior adaptation hindered the ability to learn from error when perturbations were

495 separated by 5 min and 1 h but not at 6 h and beyond, suggesting that actual learning is
496 impaired on a shorter time scale, subsiding within a 6 h window.

497 There has been no general agreement in the literature of sensorimotor adaptation
498 regarding how to define and, therefore, quantify anterograde interference. With the
499 exception of Sing and Smith (2010), who measured the relative change in learning rate,
500 most previous studies estimated anterograde interference based on the initial level of
501 performance, by averaging across the first trials/cycles/blocks (Brashers-Krug et al., 1996;
502 Krakauer, Ghez, & Ghilardi, 2005; Lee & Schweighofer, 2009; Shadmehr & Brashers-Krug,
503 1997; Tong & Flanagan, 2003). For example, Tong and Flanagan (2003) reported
504 interference at 5 min based on the average of the second and third cycles. Likewise, Miall
505 and collaborators (2004) reported interference at 15 min based on the initial state obtained
506 from fitting a power function, while noting that the rate of learning was not affected. Yet,
507 there is evidence suggesting that using the initial level of performance as a proxy for
508 anterograde interference may confound the actual ability of learning in *B* with the bias of a
509 lingering memory of *A*. For example, Thoroughman and Shadmehr (1999) have shown
510 that the preferred direction of the biceps and triceps during exposure to the second
511 opposing force field is appropriate to solve the first force field. Moreover, Sing and Smith
512 (2010) have demonstrated that the magnitude of initial errors does not impact on the ability
513 to learn. In fact, high initial error may be associated either with faster (Sing et al., 2009) or
514 slower learning. Therefore, assessing anterograde interference based on the initial level of
515 performance may overestimate its magnitude. To rule out this possibility here we
516 compared the bias imposed by the memory of *A* with the deficit observed in the rate of
517 learning. We reasoned that if, as suggested by previous work, the initial level of
518 performance reflects the level of anterograde interference then the two measures should
519 behave similarly as a function of time. In contrast, we found that initial performance was
520 profoundly hindered throughout the 24 h of testing, whereas the ability to learn returned to

521 control levels by 6 h. This impairment in the ability to learn that gradually subsides with the
522 passage of time is consistent with anterograde interference.

523 Our work sheds light on a long-standing debate regarding the failure of retrograde
524 protocols at unveiling the time course of memory consolidation. Over the past two
525 decades, several laboratories have attempted to uncover the time course of memory
526 stabilization in sensorimotor adaptation using behavioral protocols based on retrograde
527 interference (e.g. Brashers-Krug et al., 1996; Caithness et al., 2004; Krakauer et al.,
528 2005). In these studies, subjects usually adapt to opposing perturbations A (A_1) and B
529 separated by a time interval that varies between minutes to 24 h. Next, they wait for a
530 further period of time (usually 24 h) and are again exposed to A (A_2) to assess the integrity
531 of the motor memory. Consolidation of the memory of A should be reflected as the
532 presence of savings (a faster rate of learning) in A_2 . Although this approach has proved
533 successful in declarative (Lechner, Squire, & Byrne, 1999; Tulving, 1969) and some kinds
534 of motor skill learning tasks (Korman et al., 2007; Walker, Brakefield, Hobson, & Stickgold,
535 2003), it has led to inconclusive results in sensorimotor adaptation. In fact, with the
536 exception of three force-field studies reporting release from interference at around 6 h
537 (Brashers-Krug et al., 1996; Shadmehr & Brashers-Krug, 1997) or later (Overduin,
538 Richardson, Lane, Bizzi, & Press, 2006), other experiments have shown complete lack of
539 savings even if 24 h are interposed between A_1 and B (Caithness et al., 2004; Goedert &
540 Willingham, 2002; Krakauer et al., 2005). Miall and collaborators (2004) have claimed that
541 naïve performance at recall (A_2) reported in retrograde protocols (Caithness et al., 2004;
542 Goedert & Willingham, 2002; Krakauer et al., 2005) actually reflects a mixture of
543 anterograde interference from B and the integrity of the memory of A , and not catastrophic
544 retrograde interference. It is important to note, however, that these authors measured
545 anterograde interference based on the initial level of performance. In light of our findings,
546 the interpretation of these studies may need to be revised. Because release from

547 interference occurs at 6 h anterograde interference is not likely to cause naïve
548 performance in A_2 . Future work using retrograde protocols should in fact track the integrity
549 of the memory based on the speed of learning in experimental designs in which B and A_2
550 are separated by at least 6 h.

551 The temporal dissociation we observed between the initial level of performance and
552 the rate of learning likely reflects two distinct processes at play: the persistence of a prior
553 memory and competition for the same neural resources. The formation of memory involves
554 learning-dependent synaptic plasticity as part of a process known as long-term
555 potentiation (LTP). Given that biological substrates underlying synaptic plasticity are
556 limited by nature, cellular modifications induced by learning temporarily constrain the
557 capacity for further LTP induction. This phenomenon known as occlusion, which reflects
558 competition from the same neural resources, has been implemented experimentally to
559 investigate LTP maintenance and memory stabilization (Ling et al., 2002). Using this
560 approach, it has been shown that motor skill learning in rats and humans is associated
561 with LTP (Cantarero, Tang, O'Malley, Salas, & Celnik, 2013; Rioult-Pedotti, Friedman,
562 Hess, & Donoghue, 1998; Rioult-Pedotti, Friedman, & Donoghue, 2000). Cantarero and
563 collaborators showed that in fact, in humans, occlusion fades around 6 h after motor skill
564 learning. In this light, we may speculate that adaptation in A may have partially occluded
565 the capacity for further synaptic plasticity, thereby hindering adaptation in B . The timing of
566 release from interference described herein (6 h) coincides with the peak in functional
567 connectivity of a visuomotor adaptation network that includes the primary motor cortex
568 (M1), the posterior parietal cortex (PPC) and the cerebellum (Della-Maggiore, Villalta,
569 Kovacevic, & McIntosh, 2015). These regions have been linked to memory formation in
570 this paradigm (Della-Maggiore, Malfait, Ostry, & Paus, 2004; Hadipour-Niktarash, Lee,
571 Desmond, & Shadmehr, 2007; Landi, Baguear, & Della-Maggiore, 2011; Richardson et al.,
572 2006). Whether this timing reflects the process of motor memory consolidation is now a

573 hypothesis amenable for testing.

574 Using a state-space model allowed us to identify which aspect of learning was
575 affected by anterograde interference. In other words, was the decrease in the rate of
576 learning observed in *B* caused by a deficit in the ability to learn from error or in the ability
577 to retain information cycle-by-cycle? Error sensitivity refers to how the brain responds to
578 perceived error from one movement to the next and, as such, has been widely used to
579 study changes across learning sessions in the context of savings and meta-learning
580 (Herzfeld, Vaswani, Marko, & Shadmehr, 2014; Leow, de Rugy, Marinovic, Riek, & Carroll,
581 2016). The retention factor, on the other hand, refers to the degree of decay that occurs
582 from one movement to another, reflecting the ability to retain information. It has been
583 previously shown that the retention factor and error sensitivity can be independently
584 affected by different manipulations (Galea, Mallia, Rothwell, & Diedrichsen, 2015) such as
585 punishment (increasing only the sensitivity to error) and reward (enhancing retention).
586 Here we found that prior learning of a contrasting perturbation had no effect on the cycle-
587 by-cycle retention of subsequent learning.

588 Humans also have the capacity to change their error sensitivity depending on their
589 prior experience with errors (Gonzalez Castro, Hadjiosif, Hemphill, & Smith, 2014). For
590 example, Herzfeld and colleagues (2014) compared the sensitivity to error in participants
591 that were exposed to alternating force fields, and found that individuals who experienced a
592 slow switching environment showed a greater sensitivity to error than those who were
593 exposed to a fast switching environment. These results indicate that the brain relies on a
594 history of past errors to learn and, thus, when errors are not predictable learning is
595 attenuated. Likewise, here we found that the temporal distance separating two opposing
596 perturbations also hinders sensitivity to error. We may speculate that when perturbations
597 are presented in close proximity (as in our 5 min and 1 h groups) competition for neural
598 resources may hinder the ability to maintain a history of errors, resulting in a reduced error

599 sensitivity. This phenomenon may revert at longer intervals as neural competition
600 subsides.

601 In conclusion, we have examined the strength and duration of anterograde
602 interference in visuomotor adaptation by tracking its impact on behavior when learning
603 opposing perturbations separated from 5 min through 24 h. We found that prior learning
604 dramatically hindered the initial state at all time intervals. This was likely due to a bias
605 imposed by a lingering memory associated with the previous perturbation. Prior learning
606 also impaired the ability to learn from error for at least 1 h, with release from interference
607 detected about 6 h post training. The occurrence of release from interference within this
608 time interval is consistent with a process of memory stabilization for this type of learning.
609 Our findings suggest that poor performance observed when opposing rotations are learned
610 subsequently is driven by two distinct phenomena operating on different time scales (days
611 vs hours): a long-lasting influence of a memory that acts as a prior which negatively
612 influences the initial level of performance, and a shorter-lasting impairment of learning.

613

614 **ACKNOWLEDGMENTS**

615 This work was supported by the Argentinian Ministry of Defense (PIDDEF), the Argentinian
616 Agency for the promotion of Science and Technology (FONCyT, ANPCyT), and the
617 University of Buenos Aires (UBACyT). Reza Shadmehr was supported by grants from the
618 National Institutes of Health (R01-NS078311, R01-NS095706). Scott Albert was supported
619 by a pre-doctoral fellowship from the National Institutes of Health (NS-095706).

620

621

622

623

FIGURES

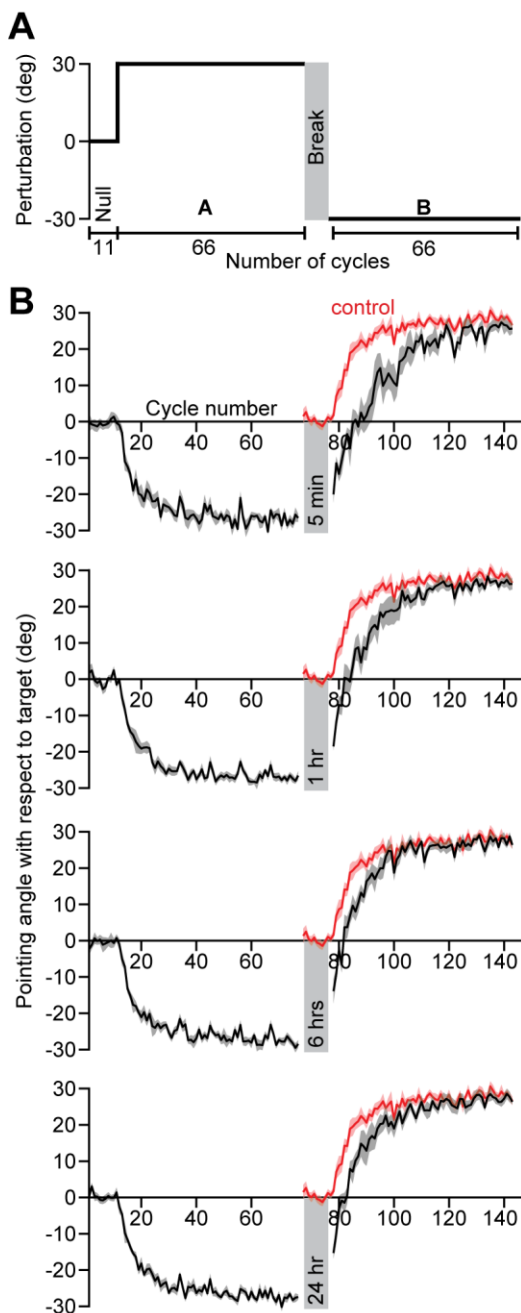


Figure 1. Experimental paradigm and learning curves. **A.** Paradigm. Subjects held a joystick and made pointing movements towards one of eight visual targets shown on a display. The experiment began with 11 cycles of null trials (Null) after which a 30° counterclockwise rotation was applied to the cursor for 66 cycles (A). Next, each experimental group waited a different length of time ranging from 5 min to 24 h. After this break, subjects were immediately exposed to a -30° clockwise rotation (B) for 66 cycles. **B.** Behavior. Subject pointing angles on each trial were collapsed into cycles by identifying the median pointing angle across each cycle of 8 trials. Each inset shows the median behavior of 1 of the 4 experimental groups. The shaded region indicates ± 1 standard error of the median. Each group differs in the amount of time that elapsed between the exposure to the A and B periods (from top to bottom: 5 min, 1 h, 6 h, and 24 h). The behavior for each group (black) is compared with that of a control group (red) that was exposed to 11 cycles of null and then 66 cycles of the B rotation.

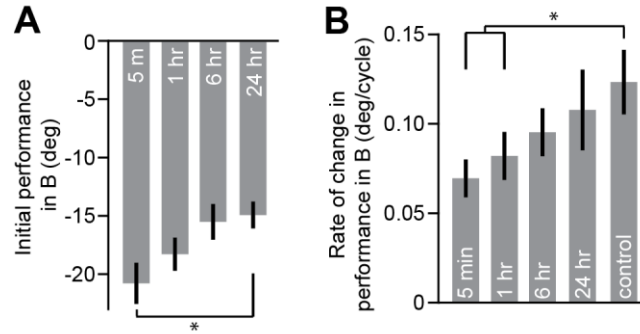


Figure 2. Effect of prior learning on different behavioral parameters. **A.** The initial level of performance in *B*, estimated from the mean pointing angle on the first cycle is displayed here for all groups. Given that learning within one cycle is minimal, this measure reflects the retention of the memory of *A*. **B.** The median rate of improvement (*i.e.*, the empirical learning rate) in *B* for all experimental groups and the control group is shown. In both **A** and **B**, experimental groups are shown from left to right in order of increasing temporal separation between the *A* and *B* periods. The control group is shown at far right where appropriate. In **A** and **B**, error bars indicate ± 1 standard error of the mean and median, respectively. Asterisks indicate a level of significance of $p < 0.05$.

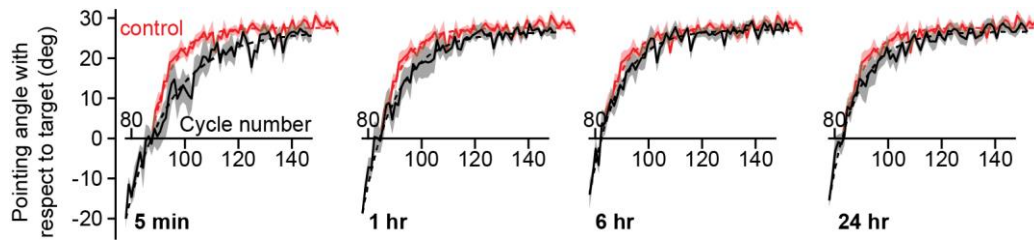


Figure 3. Anterograde interference and performance in *B*. The median pointing angle of each experimental group during learning the *B* rotation (solid black) and the control group (solid red) is shown. The shaded error region indicates ± 1 standard error of the median. To highlight any difference in the rate of learning of the *B* rotation, the behavior of the control group was shifted in all plots, so that it roughly intersected the behavior of each experimental group at a pointing angle of approximately 0° . The magnitude of the shift was determined by fitting exponential curves (dashed lines) to the behavior of the experimental group (black) and control group (red) and horizontally shifting the control group so that the exponential fits intersected at 0° . From left to right, experimental groups are ordered in terms of increasing separation between the *A* and *B* perturbations.

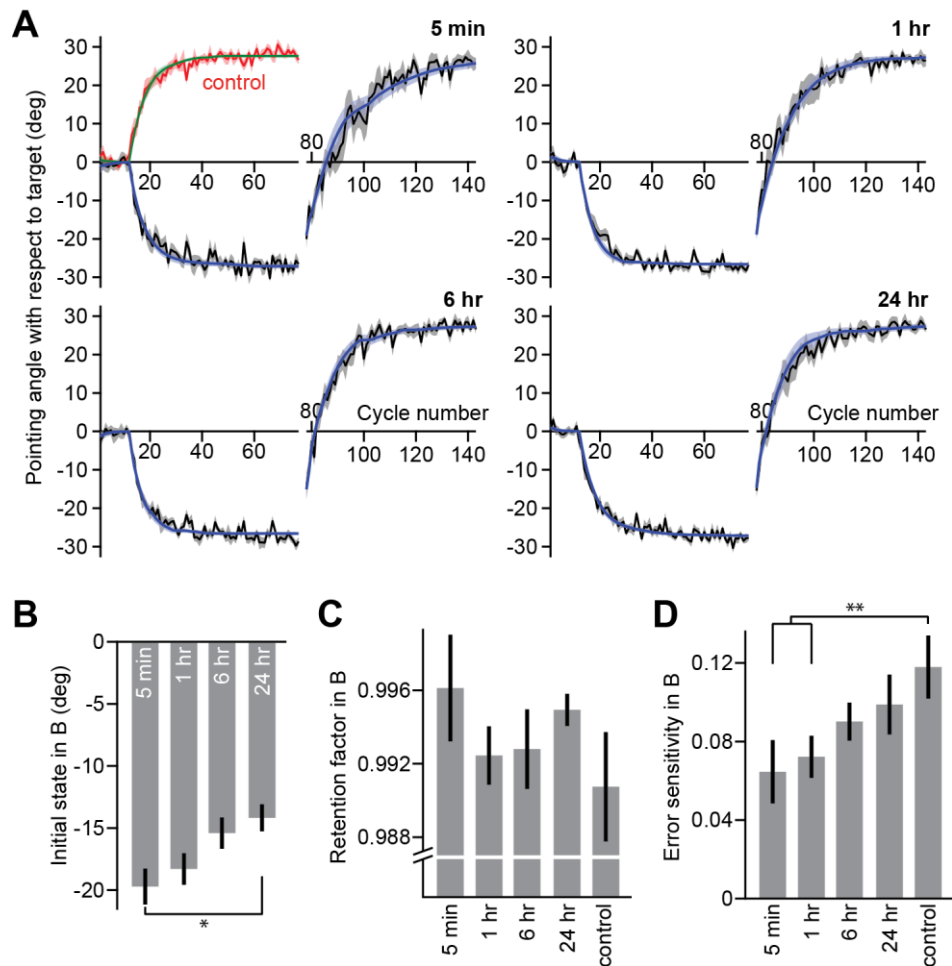


Figure 4. State-space model fit to behavior. **A.** We fit individual subject behavior using a single module state-space model of learning that accounted for cycle-by-cycle forgetting, error-based learning, and initial bias. We fit behavior separately for the *A* (cycles 1 through 77) and *B* (cycles 78 through 143) periods. Each plot depicts the median pointing angle for 1 of the 4 experimental groups (black lines) as well as the median pointing angle predicted from simulating the state-space model without noise (blue lines) using the maximum likelihood model parameter sets identified for each subject. Corresponding behavior (red) and state-space predictions (green) are provided for the control group in the top-left plot. The shaded error region indicates ± 1 standard error of the median. **B.** Initial state. Here we report the initial state of the learner at the start of the *B* period. **C.** Here we report the retention factor during the *B* period for the experimental and control groups. **D.** We report the error sensitivity during the *B* period for the experimental and control groups. The height of each bar denotes the mean (**B**) or median (**C** and **D**) parameter value for each experimental group. Error bars indicate ± 1 standard error of the mean or median. In all plots, experimental groups are shown from left to right in order of increasing temporal separation between the *A* and *B* periods. The control group is shown at far right where appropriate. Asterisks indicate a level of significance of $p < 0.05$ (*) or $p < 0.01$ (**).

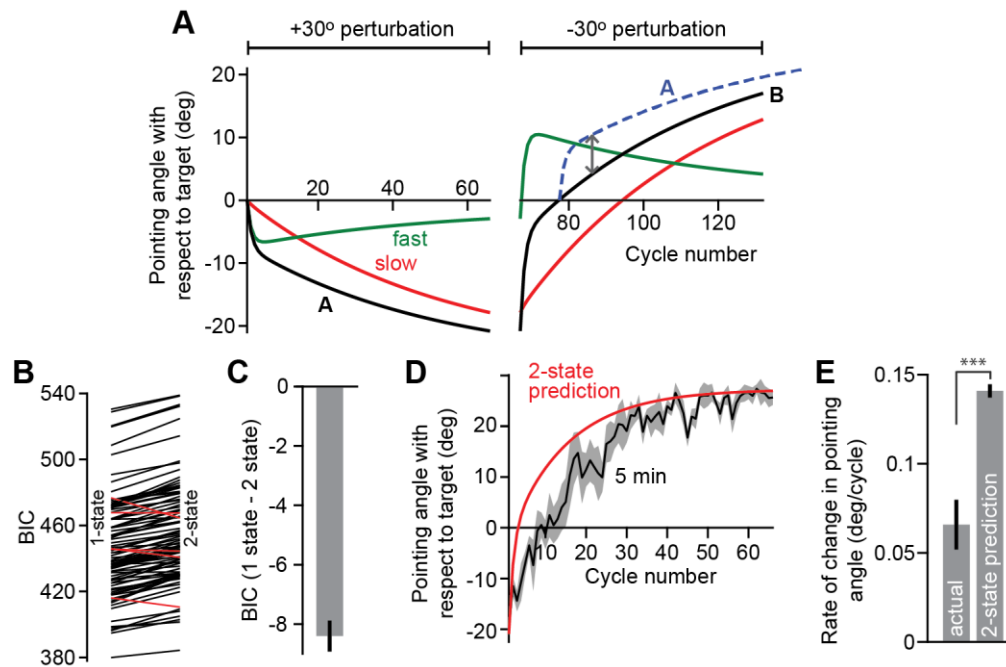


Figure 5. Validation of the state-space model. Prior accounts of anterograde interference have shown that a two-state model of learning can show an impairment in learning of the *B* memory without any change in the constitutive parameters of the two-state system from *A* to *B*. **A.** The two-state model (parameters obtained from Sing and Smith, 2010) posits that behavior (black) can be decomposed into parallel contributions from a slow (red) and fast (green) state. After learning the *A* perturbation (left half of the figure, +30° perturbation) both the fast and slow states are biased towards a memory of *A*. At the start of the *B* period (right half of the figure, -30° perturbation) the initial bias of the slow state towards *A* can slow the apparent rate of learning of the *B* perturbation, as compared to initial learning of *A* (blue trace shows the pointing angle during *A* with opposite sign and shifted to the 0° pointing angle of *B*). **B.** We asked if a two-state system could better account for subject behavior than the single state system considered in the primary analysis. We calculated the Bayesian Information Criterion (BIC) associated with single-state and two-state model fits to individual subject behavior. The endpoint of each line shows the average BIC for the *A* and *B* periods (left, single state model; right, two-state model). Each line depicts the result for a single subject. Red lines indicate subjects for which the two-state model was superior to the single state model. Black lines indicate subjects for which the single state model was superior to the two-state model. **C.** We calculated the difference in BIC for the single state and two-state models. Negative values indicate higher evidence for the single-state model. The height of the bar indicates the mean BIC, and error bars indicate ± 1 standard error of the mean. **D.** We simulated behavior of each subject during the *B* perturbation (red) using the two-state model parameters fit to the *A* period. The initial bias of the slow state towards a memory of *A* did not produce learning in *B* that resembled the impaired rate of learning exhibited in the 5 min experimental group (black). Solid lines indicate the median prediction or behavior across subjects. Error shade indicates ± 1 standard error of the median. **E.** To quantify any discrepancy between the rate of learning in *B* and the rate of learning predicted by the two-state model we fit our empirical exponential learning curve to the actual and simulated *B* behaviors. The height of each bar indicates the median empirical learning rate. Error bars indicate ± 1 standard error of the median. Asterisks indicate a level of significance of $p < 0.001$.

624 **REFERENCES**

- 625 Albert, S. T., & Shadmehr, R. (2018). Estimating properties of the fast and slow adaptive
626 processes during sensorimotor adaptation. *Journal of Neurophysiology*, *119*(4),
627 1367–1393. <https://doi.org/10.1152/jn.00197.2017>
- 628 Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor
629 memory. *Nature*, *382*(6588), 252–255. <https://doi.org/10.1038/382252a0>
- 630 Braun, D. A., Aertsen, A., Wolpert, D. M., & Mehring, C. (2009). Motor Task Variation
631 Induces Structural Learning. *Current Biology*, *19*(4), 352–357.
632 <https://doi.org/10.1016/j.cub.2009.01.036>
- 633 Caithness, G., Osu, R., Bays, P., Chase, H., Klassen, J., Kawato, M., ... Flanagan, J.
634 (2004). Failure to Consolidate the Consolidation Theory of Learning for Sensorimotor
635 Adaptation Tasks. *Journal of Neuroscience*, *24*(40), 8662–8671.
636 <https://doi.org/10.1523/JNEUROSCI.2214-04.2004>
- 637 Cantarero, G., Tang, B., O'Malley, R., Salas, R., & Celnik, P. (2013). Motor Learning
638 Interference Is Proportional to Occlusion of LTP-Like Plasticity. *Journal of*
639 *Neuroscience*, *33*(11), 4634–4641. [https://doi.org/10.1523/JNEUROSCI.4706-](https://doi.org/10.1523/JNEUROSCI.4706-12.2013)
640 [12.2013](https://doi.org/10.1523/JNEUROSCI.4706-12.2013)
- 641 Cheng, S., & Sabes, P. N. (2006). Modeling Sensorimotor Learning with Linear Dynamical
642 Systems. *Neural Computation*, *18*(4), 760–793.
643 <https://doi.org/10.1162/neco.2006.18.4.760>
- 644 Della-Maggiore, V., Malfait, N., Ostry, D. J., & Paus, T. (2004). Stimulation of the posterior
645 parietal cortex interferes with arm trajectory adjustments during the learning of new
646 dynamics. *Journal of Neuroscience*, *24*(44), 9971–9976.
647 <https://doi.org/10.1523/JNEUROSCI.2833-04.2004>
- 648 Della-Maggiore, V., Villalta, J. I., Kovacevic, N., & McIntosh, A. R. (2015). Functional
649 Evidence for Memory Stabilization in Sensorimotor Adaptation: A 24-h Resting-State

- 650 fMRI Study. *Cerebral Cortex*, 27(3), bhv289. <https://doi.org/10.1093/cercor/bhv289>
- 651 Donchin, O., Francis, J. T., & Shadmehr, R. (2003). Quantifying Generalization from Trial-
652 by-Trial Behavior of Adaptive Systems that Learn with Basis Functions: Theory and
653 Experiments in Human Motor Control. *The Journal of Neuroscience*, 23(27), 9032–
654 9045. <https://doi.org/10.1523/JNEUROSCI.23-27-09032.2003>
- 655 Ethier, V., Zee, D. S., & Shadmehr, R. (2008). Spontaneous Recovery of Motor Memory
656 During Saccade Adaptation. *Journal of Neurophysiology*, 99(5), 2577–2583.
657 <https://doi.org/10.1152/jn.00015.2008>
- 658 Galea, J. M., Mallia, E., Rothwell, J., & Diedrichsen, J. (2015). The dissociable effects of
659 punishment and reward on motor learning. *Nature Neuroscience*, 18(4), 597–602.
660 <https://doi.org/10.1038/nn.3956>
- 661 Goedert, K. M., & Willingham, D. B. (2002). Patterns of interference in sequence learning
662 and prism adaptation inconsistent with the consolidation hypothesis. *Learning &
663 Memory*, 9(5), 279–292. <https://doi.org/10.1101/lm.50102>
- 664 Gonzalez Castro, L. N., Hadjiosif, A. M., Hemphill, M. A., & Smith, M. A. (2014).
665 Environmental Consistency Determines the Rate of Motor Adaptation. *Current
666 Biology*, 24(10), 1050–1061. <https://doi.org/10.1016/j.cub.2014.03.049>
- 667 Hadipour-Niktarash, A., Lee, C. K., Desmond, J. E., & Shadmehr, R. (2007). Impairment of
668 Retention But Not Acquisition of a Visuomotor Skill Through Time-Dependent
669 Disruption of Primary Motor Cortex. *Journal of Neuroscience*, 27(49), 13413–13419.
670 <https://doi.org/10.1523/JNEUROSCI.2570-07.2007>
- 671 Herzfeld, D. J., Vaswani, P. A., Marko, M. K., & Shadmehr, R. (2014). A memory of errors
672 in sensorimotor learning. *Science*, 345(6202), 1349–1353.
673 <https://doi.org/10.1126/science.1253138>
- 674 Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime
675 sleep condenses the time course of motor memory consolidation. *Nature*

- 676 *Neuroscience*, 10(9), 1206–1213. <https://doi.org/10.1038/nn1959>
- 677 Krakauer, J. W. (2009). Motor Learning and Consolidation: The Case of Visuomotor
678 Rotation (pp. 405–421). https://doi.org/10.1007/978-0-387-77064-2_21
- 679 Krakauer, J. W., Ghez, C., & Ghilardi, M. F. (2005). Adaptation to visuomotor
680 transformations: consolidation, interference, and forgetting. *Journal of Neuroscience*,
681 25(2), 473–478. <https://doi.org/10.1523/JNEUROSCI.4218-04.2005>
- 682 Landi, S. M., Baguear, F., & Della-Maggiore, V. (2011). One Week of Motor Adaptation
683 Induces Structural Changes in Primary Motor Cortex That Predict Long-Term Memory
684 One Year Later. *Journal of Neuroscience*, 31(33), 11808–11813.
685 <https://doi.org/10.1523/JNEUROSCI.2253-11.2011>
- 686 Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 Years of Consolidation—
687 Remembering Müller and Pilzecker. *Learning & Memory*, 6, 77–87.
688 <https://doi.org/10.1101/lm.6.2.77>
- 689 Lee, J.-Y., & Schweighofer, N. (2009). Dual Adaptation Supports a Parallel Architecture of
690 Motor Memory. *Journal of Neuroscience*, 29(33), 10396–10404.
691 <https://doi.org/10.1523/JNEUROSCI.1294-09.2009>
- 692 Leow, L.-A., de Rugy, A., Marinovic, W., Riek, S., & Carroll, T. J. (2016). Savings for
693 visuomotor adaptation require prior history of error, not prior repetition of successful
694 actions. *Journal of Neurophysiology*, 116(4), 1603–1614.
695 <https://doi.org/10.1152/jn.01055.2015>
- 696 Ling, D. S. F., Benardo, L. S., Serrano, P. A., Blace, N., Kelly, M. T., Crary, J. F., &
697 Sacktor, T. C. (2002). Protein kinase M ζ is necessary and sufficient for LTP
698 maintenance. *Nature Neuroscience*, 5(4), 295–296. <https://doi.org/10.1038/nn829>
- 699 Miall, R. C., Jenkinson, N., & Kulkarni, K. (2004). Adaptation to rotated visual feedback: a
700 re-examination of motor interference. *Experimental Brain Research*, 154(2), 201–210.
701 <https://doi.org/10.1007/s00221-003-1630-2>

- 702 Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh
703 inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-](https://doi.org/10.1016/0028-3932(71)90067-4)
704 3932(71)90067-4
- 705 Overduin, S. A., Richardson, A. G., Lane, C. E., Bizzi, E., & Press, D. Z. (2006).
706 Intermittent Practice Facilitates Stable Motor Memories. *Journal of Neuroscience*,
707 26(46), 11888–11892. <https://doi.org/10.1523/JNEUROSCI.1320-06.2006>
- 708 Richardson, A. G., Overduin, S. A., Valero-Cabre, A., Padoa-Schioppa, C., Pascual-
709 Leone, A., Bizzi, E., & Press, D. Z. (2006). Disruption of Primary Motor Cortex before
710 Learning Impairs Memory of Movement Dynamics. *Journal of Neuroscience*, 26(48),
711 12466–12470. <https://doi.org/10.1523/JNEUROSCI.1139-06.2006>
- 712 Rioult-Pedotti, M.-S., Friedman, D., Hess, G., & Donoghue, J. P. (1998). Strengthening of
713 horizontal cortical connections following skill learning. *Nature Neuroscience*, 1(3),
714 230–234. <https://doi.org/10.1038/678>
- 715 Rioult-Pedotti, M. S., Friedman, D., & Donoghue, J. P. (2000). Learning-induced LTP in
716 neocortex. *Science (New York, N.Y.)*, 290(5491), 533–536.
717 <https://doi.org/10.1126/science.290.5491.533>
- 718 Shadmehr, R., & Brashers-Krug, T. (1997). Functional Stages in the Formation of Human
719 Long-Term Motor Memory. *The Journal of Neuroscience*, 17(1), 409–419.
720 <https://doi.org/10.1523/JNEUROSCI.17-01-00409.1997>
- 721 Shmuelof, L., Haith, A. M., Mazzoni, P., Huang, V. S., Krakauer, J. W., Vaswani, P. A., ...
722 Delnicki, R. J. (2015). Persistent Residual Errors in Motor Adaptation Tasks:
723 Reversion to Baseline and Exploratory Escape. *Journal of Neuroscience*, 35(17),
724 6969–6977. <https://doi.org/10.1523/jneurosci.2656-14.2015>
- 725 Sing, G. C., Joiner, W. M., Nanayakkara, T., Brayanov, J. B., & Smith, M. A. (2009).
726 Primitives for Motor Adaptation Reflect Correlated Neural Tuning to Position and
727 Velocity. *Neuron*, 64(4), 575–589. <https://doi.org/10.1016/j.neuron.2009.10.001>

- 728 Sing, G. C., & Smith, M. A. (2010). Reduction in Learning Rates Associated with
729 Anterograde Interference Results from Interactions between Different Timescales in
730 Motor Adaptation. *PLoS Computational Biology*, 6(8), e1000893.
731 <https://doi.org/10.1371/journal.pcbi.1000893>
- 732 Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting Adaptive Processes with
733 Different Timescales Underlie Short-Term Motor Learning. *PLoS Biology*, 4(6), e179.
734 <https://doi.org/10.1371/journal.pbio.0040179>
- 735 Thoroughman, K. A., & Shadmehr, R. (1999). Electromyographic Correlates of Learning
736 an Internal Model of Reaching Movements. *The Journal of Neuroscience*, 19(19),
737 8573–8588. <https://doi.org/10.1203/01.pdr.0000176834.47305.26>
- 738 Thoroughman, K. A., & Shadmehr, R. (2000). Learning of action through adaptive
739 combination of motor primitives. *Nature*, 407(6805), 742–747.
740 <https://doi.org/10.1038/35037588>
- 741 Tong, C., & Flanagan, J. R. (2003). Task-Specific Internal Models for Kinematic
742 Transformations. *Journal of Neurophysiology*, 90(2), 578–585.
743 <https://doi.org/10.1152/jn.01087.2002>
- 744 Tulving, E. (1969). Retrograde Amnesia in Free Recall. *Science*, 164(3875), 88–90.
745 <https://doi.org/10.1126/science.164.3875.88>
- 746 Walker, M. P., Brakefield, T., Allan Hobson, J., & Stickgold, R. (2003). Dissociable stages
747 of human memory consolidation and reconsolidation. *Nature*, 425(6958), 616–620.
748 <https://doi.org/10.1038/nature01930>
- 749 Wigmore, V., Tong, C., & Flanagan, J. R. (2002). Visuomotor rotations of varying size and
750 direction compete for a single internal model in a motor working memory. *Journal of*
751 *Experimental Psychology: Human Perception and Performance*, 28(2), 447–457.
752 <https://doi.org/10.1037/0096-1523.28.2.447>
- 753