

Evolution of Motor Memory During the Seconds After Observation of Motor Error

Vincent S. Huang and Reza Shadmehr

Laboratory for Computational Motor Control, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland

Submitted 6 December 2006; accepted in final form 4 April 2007

Huang VS, Shadmehr R. Evolution of motor memory during the seconds after observation of motor error. *J Neurophysiol* 97: 3976–3985, 2007; doi:10.1152/jn.01281.2006. When a movement results in error, the nervous system amends the motor commands that generate the subsequent movement. Here we show that this adaptation depends not just on error, but also on passage of time between the two movements. We observed that subjects learned a reaching task faster, i.e., with fewer trials, when the intertrial time intervals (ITIs) were lengthened. We hypothesized two computational mechanisms that could have accounted for this. First, learning could have been driven by a Bayesian process where the learner assumed that errors are the result of perturbations that have multiple timescales. In theory, longer ITIs can produce faster learning because passage of time might increase uncertainty, which in turn increases sensitivity to error. Second, error in a trial may result in a trace that decays with time. If the learner continued to sample from the trace during the ITI, then adaptation would increase with increased ITIs. The two models made separate predictions: The Bayesian model predicted that when movements are separated by random ITIs, the learner would learn most from a trial that followed a long time interval. In contrast, the trace model predicted that the learner would learn most from a trial that preceded a long time interval. We performed two experiments to test for these predictions and in both experiments found evidence for the trace model. We suggest that motor error produces an error memory trace that decays with a time constant of about 4 s, continuously promoting adaptation until the next movement.

INTRODUCTION

Learning depends not just on number of repeated exposures, but also on the temporal distribution of the exposures. Consider the distinction between *massed* and *spaced* training, first coined by Ebbinghaus (1964) on the study of memory. In massed training, the trials take place in close temporal proximity. In spaced training, trials are separated by periods of rest. Ebbinghaus found that spacing training sets over time was more effective in allowing him to memorize a list of nonsense words to criterion with less practice than massing them in a single set. Spaced training also improves rates of learning in other tasks (Aboukhalil et al. 2004; Commins et al. 2003; Han et al. 1998; Savion-Lemieux and Penhune 2005). For example, two recent studies examined this effect in the context of reach adaptation. Bock et al. (2005) trained subjects to point in a novel visual feedback environment and observed that the rate of adaptation was faster if the trials were separated by 5 to 40 s than by 1 s. Similarly, Francis (2005) noted that learning to control a novel tool (reaching while holding a robot arm in a

force field) was faster if the trials were separated by 5–20 than by 0.5 s. These results are not accounted for by current computational models of motor learning where adaptation is driven only by motor error (Donchin et al. 2003; Scheidt et al. 2001; Wainwright et al. 2005) because these models treat the spaced and massed training paradigms identically. Rather, the results suggest that trial-to-trial effect of motor error depends on the intertrial interval between the movements.

More recent computational models of motor learning suggest that motor error engages multiple adaptive processes of different timescales: one process strongly responds to error but has poor retention and another has poor sensitivity to error but has better retention (Smith et al. 2006). This model and its Bayesian variant (Kording 2007) emphasize that passage of time is an important variable that influences content of memory. Can such models help explain the massed versus spaced training effects?

Here we begin with an experiment that confirms the previous findings that reach adaptation in force fields is indeed faster (i.e., require fewer trials) when the intertrial interval (ITI) is increased from 4 to 14 s. We demonstrate that the multiple-timescale model under a Bayesian formulation can account for this result. Because time passage increases the Bayesian learner's uncertainty about his/her environment and increased uncertainty promotes the incentive to adapt, longer ITIs lead to fewer trials required for adaptation. The model predicts that the learner will learn more from a trial that immediately follows a long delay than one that immediately follows a short delay.

To test this prediction, we performed another experiment where ITIs were randomly distributed. However, we found results inconsistent with the predictions of the Bayesian model. People learned more from a trial that *preceded* a long delay, not a trial that *followed* a long delay. Therefore the improved adaptation rates in spaced trials were unlikely to be a result of increased uncertainty. Rather, our results suggest that movement errors produced a trace that continued to benefit the learner during the ITI. It appears that this trace has a time constant of about 4 s.

METHODS

We reconsidered a well-studied reach adaptation task (Shadmehr and Mussa-Ivaldi 1994) and asked whether spaced training benefited rates of adaptation. Subjects held the handle of a two-joint, planar manipulandum equipped with torque motors, rotary encoders, and force transducers and reached to visual targets. Subject's upper arm

Address for reprint requests and other correspondence: V. Huang, Johns Hopkins University School of Medicine, 720 Rutland Ave., 416 Traylor Building, Baltimore, MD 21205 (E-mail: vhuang@jhu.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

was supported by a sling restricting movements to the same horizontal plane as the manipulandum. During “field trials,” the torque motors of the manipulandum perturbed the movement of the subjects by a viscous curl-force field

$$F = V\dot{x} \quad V = \begin{bmatrix} 0 & 13 \\ -13 & 0 \end{bmatrix} \quad (1)$$

In Eq. 1, force F is in Newtons, and hand velocity \dot{x} is in meters per second. No forces were applied during “catch trials” and trials in the null training set.

Behavioral training

All procedures were approved by the Johns Hopkins Medicine Institutional Review Board. Subjects gave consent before their participation in the study. All subjects were healthy and right-handed. They were naive to the purpose of the experiment and had never participated in any experiment with our device before. Participants were seated in front of a flat-screen monitor situated at eye level (Fig. 1). With their right hand they grasped the handle at the end of the manipulandum to navigate a cursor (a white dot) on the screen. Subjects were trained to maintain the cursor at the center of the screen

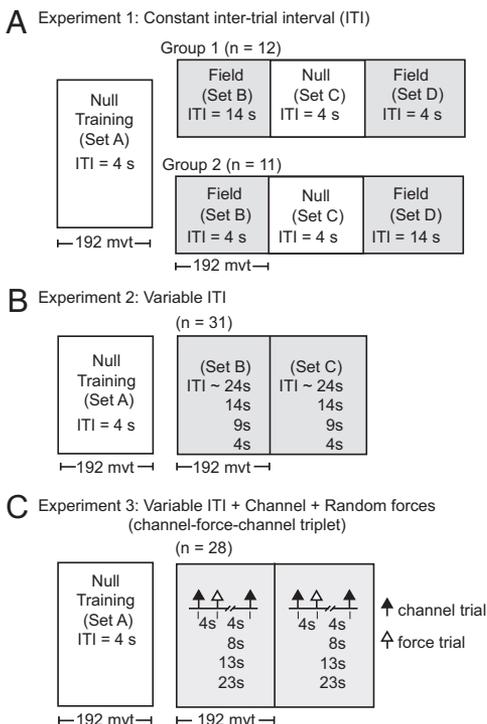


FIG. 1. Experiment design. *A*: experiment 1 protocol. Intertrial intervals (ITIs) were constant within a movement set (sets A, B, C, D). Shaded blocks indicate that the force field was on for all but catch movement trials in that set. *B*: experiment 2 protocol. ITIs were randomly varied during field training sets. *C*: experiment 3 protocol. Field trials (open triangle) were sandwiched between pairs of channel trials (solid triangle) and we varied the time between the field trial and the subsequent channel trial.

as indicated by a yellow crosshair until a 1-cm² green target box appeared, which also served as the “go” signal. They were also told to “aim for the target as it appears and try making a straight point-to-point movement to the target in a fast and smooth fashion.” The yellow crosshair vanished as soon as subject started to move.

The green target box could appear at any one of the eight locations spaced on an invisible circle of 10-cm radius centered at the crosshair. The sequence of the target locations was determined pseudorandomly with all eight possible target locations visited with equal probability. After completion of the movement, the green target box turned magenta when the tangential peak speed exceeded 0.55 m/s or cyan when the speed faltered to <0.20 m/s, respectively. If the movement duration surpassed 0.57 s or fell to <0.43 s, the box turned red or blue. If the movement profile met the above parameters, it was considered ideal and the target box exploded. Distinctive audio feedback was also given for magenta, blue, and exploding target boxes. The manipulandum subsequently returned the subject’s hand to the origin at the center of the screen.

Performance measure

Movement errors were measured as the signed perpendicular displacements (PDs) of the reach at peak speed with respect to a straight line to the target. We grouped 32 movements into one movement bin and averaged the errors in field and catch trials to arrive at \overline{PD}_{field} and \overline{PD}_{catch} . Next, we computed a learning index (LI) (Criscimagna-Hemming et al. 2003; Hwang et al. 2003)

$$LI = \frac{\overline{PD}_{catch}}{\overline{PD}_{catch} - \overline{PD}_{field}} \quad (2)$$

The denominator of this expression is a measure of limb compliance; smaller stiffness gives rise to larger differences in errors between catch and field trials. The numerator is a measure of change in motor output with respect to null trials. Therefore the ratio is a measure of learning, normalized with respect to limb compliance. As subjects learn to predict the forces, their movement errors in field trials decrease, whereas errors in catch trials increase. Thus we would expect the learning index to grow and plateau as the subjects adapt. Complete adaptation would result in a learning index of one. However, catch trials cause unlearning (Thoroughman and Shadmehr 2000) and prevent the index from reaching one. The ratio of field trials to catch trials (i.e., 5/6) dictates the theoretical limit of learning index, which is 0.83. In our experience, the highest actual learning performance is slightly below this number (Criscimagna-Hemming et al. 2003). Movements were excluded if they did not meet the following criteria: maximal tangential speed was between 0.20 and 0.55 m/s, movement duration was between 0.3 and 1.2 s, and the total movement trajectory length was <20 cm.

Experiment 1: constant ITI

In this experiment we sought to replicate the results of Francis (2005) and Bock et al. (2005). The experiment was divided into four sets (Fig. 1A), each containing 192 reach trials: set A (null baseline), set B (field), set C (null washout), and set D (field). In sets B and D, the field was randomly removed in one sixth of the trials (catch trials). Short breaks of 5 min were given between sets. Subjects ($n = 24$) were randomly assigned to two counterbalanced groups. One individual did not follow instructions and was excluded from data analysis. Both groups performed the same sequence of movements in each of the four sets with one difference: in set B of the first group, after the hand returned to the center location the target presentation was delayed by 10 s, whereas target presentation in all other sets was delayed by only 0.5 s (sets A, C, D). In the second group, target presentation was delayed by 10 s in set D and by 0.5 s in all other sets

(sets A, B, C). We defined ITI to be the time between the onsets of consecutive reaching trials and it included all delays and movement time. Thus the presentation delays resulted in mean ITI of either 4 or 14 s.

Because we found that longer ITI produced significantly faster rates of adaptation, we considered two general models that might account for the data.

MODEL 1 (BAYESIAN MULTIRATE MODEL): MOTOR ADAPTATION AS OPTIMAL INFERENCE AT MULTIPLE RATES. One way to account for time-dependent changes in motor performance is to envision that the learner is a Bayesian estimator that assumes that motor error arises from multiple causes: some perturbations go away quickly but tend to be highly variable (fast system; e.g., fatigue), whereas other perturbations tend to go away slowly and tend to be less variable (slow system; e.g., disease). We recently formalized this idea and demonstrated that a multiple-system model can account for a large body of data in saccade and reach adaptation (Kording 2007). The principal idea in this model is that movement error results in a credit assignment problem for the nervous system. To solve a credit assignment problem, we need to determine how to vary the contributions of the two systems to a common task: what is the timescale of perturbation that is most likely responsible for the current error? Is the perturbation likely to go away quickly or is it likely to be sustained? If it is likely to be sustained—as with spaced perturbations—the learner should increase his/her error sensitivity for slower timescales. To show how such a model will be affected by ITI, suppose that the learner assumes that the perturbations (e.g., force f imposed on the limb) are caused by a linear combination of two sources, each with its own states, timescale, and noise properties

$$\begin{aligned} f^{(n)} &= z_1^{(n)} + z_2^{(n)} + \epsilon_0 \quad \epsilon_0 \in N(0, \sigma_0^2) \\ z_1^{(n)} &= a_1 z_1^{(n-1)} + \epsilon_1 \quad \epsilon_1 \in N(0, \sigma_1^2) \\ z_2^{(n)} &= a_2 z_2^{(n-1)} + \epsilon_2 \quad \epsilon_2 \in N(0, \sigma_2^2) \end{aligned} \tag{3}$$

In the preceding system of equations at each iteration n , the learner’s state (represented by variable z) is a reflection of two underlying sources: one system (represented by state z_1) for the fast perturbations and another (represented by state z_2) for the slow perturbations. ϵ_0 is the noise in our sensors that measure the perturbation and ϵ_1 and ϵ_2 are noises associated with the fast and slow perturbations (we assume that $a_1 < a_2 < 1$ and $\sigma_1 > \sigma_2$). The learner’s “knowledge” on each trial is a sum of contributions of the fast and slow systems with associated noises. It is convenient to rewrite Eq. 3 in vector format as

$$\begin{aligned} f^{(n)} &= \mathbf{x}^T \mathbf{z}^{(n)} + \epsilon_0 \quad \epsilon_0 \in N(0, \sigma_0^2) \\ \mathbf{z}^{(n)} &= A \mathbf{z}^{(n-1)} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \in N(0, Q) \end{aligned} \tag{4}$$

In Eq. 4, $\mathbf{x}^T = [1 \ 1]$ and A and Q are diagonal matrices with components described in Eq. 3. On trial n , given that the learner has observed the last $n - 1$ trials, it will have a prior estimate $\hat{\mathbf{z}}^{(n|n-1)}$ and a predicted perturbation $\hat{f}^{(n)} = \mathbf{x}^T \hat{\mathbf{z}}^{(n|n-1)}$. The optimal way that it can distribute the error $f^{(n)} - \hat{f}^{(n)}$ to each of the two potential sources is described by the gain $\mathbf{k}^{(n)}$ (the Kalman gain). The mean of the posterior estimate will become

$$\begin{aligned} \hat{\mathbf{z}}^{(n|n)} &= \hat{\mathbf{z}}^{(n|n-1)} + \mathbf{k}^{(n)} [f^{(n)} - \mathbf{x}^T \hat{\mathbf{z}}^{(n|n-1)}] \\ \mathbf{k}^{(n)} &= \frac{P^{(n|n-1)} \mathbf{x}}{\mathbf{x}^T P^{(n|n-1)} \mathbf{x} + \sigma_0^2} \end{aligned} \tag{5}$$

In Eq. 5, $P^{(n|n-1)}$ is the prior uncertainty matrix, describing the variance covariance of each component of \mathbf{z} .

Now if during an iteration of the model, the learner makes a movement, it will make an observation, and therefore the posterior uncertainties will change

$$P^{(n|n)} = [I - \mathbf{k}^{(n)} \mathbf{x}^T] P^{(n|n-1)} \tag{6}$$

During model iterations where the learner is not allowed to make observations, the posterior estimate will not change

$$\begin{aligned} \hat{\mathbf{z}}^{(n|n)} &= \hat{\mathbf{z}}^{(n|n-1)} \\ P^{(n|n)} &= P^{(n|n-1)} \end{aligned} \tag{7}$$

However, regardless of whether the learner makes an observation, at the next iteration the prior uncertainty will change

$$P^{(n+1|n)} = A P^{(n|n)} A^T + Q \tag{8}$$

We see that uncertainty decreases when an observation is made (Eq. 6), but can potentially increase between iterations (Eq. 8) because of Q . We assume that the learner updates the uncertainty with each iteration. Assuming that each iteration takes a constant amount of time, the longer ITIs allow more iterations than shorter ITIs. Consequently, longer ITIs (i.e., after more iterations) can produce increased uncertainty. This in turn produces higher sensitivity to error (Eq. 5), which can result in an increased rate of trial-to-trial adaptation. As a consequence, this model predicts that longer ITIs have the potential to produce faster rates of adaptation.

MODEL 2 (ERROR TRACE MODEL): ADAPTATION AS CONTINUOUS INTEGRATION OF MOTOR ERROR. Model 1 assumes that adaptation in response to error is an instantaneous process that completes by the next iteration. A different way to view adaptation is to imagine that it is a process initiated with the experience of error, but continues as long as the error memory trace is available. Let us assume that the error trace is an exponentially decaying function of the form

$$\mathbf{y}^{(n)} \frac{1}{r} \exp(-\tau/r)$$

where $\mathbf{y}^{(n)}$ is the error experienced at trial n at time t_{n+1} , τ is the time variable, and r is the time constant for the temporal decay. Suppose that the learner continuously learns from the error trace that was initiated at the time of trial n until the new error trace at trial $n + 1$ interrupts this process, the learner’s state at t_{n+1} (immediately before experiencing the error in trail $n + 1$) becomes

$$\mathbf{z}(t_{n+1}) = \mathbf{z}(t_n) + \mathbf{y}^{(n)} \int_n^{n+1} \frac{1}{r} \exp\left(\frac{t_n - \tau}{r}\right) d\tau \tag{9}$$

The integral in Eq. 9 can be simplified to

$$\mathbf{z}(t_{n+1}) = \mathbf{z}(t_n) + \mathbf{y}^{(n)} \left[1 - \exp\left(\frac{t_n - t_{n+1}}{r}\right) \right] \tag{10}$$

In this model of adaptation, a movement is like a point process that resets the previous error trace and replaces it with the most recently acquired sample. As ITIs increase, the contribution of error, i.e., the exponential term in Eq. 9, increases. As a result, the system learns more from an error when the trials are spaced in time.

PREDICTIONS OF THE TWO MODELS. Models 1 and 2 both predict that longer ITIs will affect rates of adaptation, but their mechanisms are different. In Model 1, when a trial is followed by a long delay, parameter uncertainties can increase, which in turn increases the sensitivity to error in the trial that follows the long delay. In other words, longer ITIs signal the learner to pay attention. In Model 2, when a trial is followed by a long delay, the error that was experienced in the preceding trial is integrated over a longer time interval. In other words, longer ITIs allow more samplings from the error. This increases the sensitivity to error in the trial that preceded the long delay. Therefore the key experiment is one that measures the learner’s sensitivity to error in trial n , as a function of the time that either preceded or followed that trial.

Experiment 2: variable ITI

This experiment (Fig. 1B) consisted of three sets (each with 192 trials) separated by short breaks: set A (null), set B (field), and set C (field). During the null training set, presentations of targets were delayed by 0.5 s as in *experiment 1*. During the field sets, one sixth of the trials were catch trials. However, unlike *experiment 1*, the delay before target presentation was pseudorandomly selected to be 0.5, 5, 10, or 20 s. Equal numbers of different delays for a given movement were presented in each set. Because an ITI also included the time for movements (that remained relatively constant for each participant), imposing this “go” signal delay resulted in mean ITIs of 4, 9, 14, and 24 s.

We recruited a new group of naive subjects for this experiment ($n = 31$). To assess sensitivity to error and the influence of ITI, we fitted the trial-by-trial movement errors to a state-space model (Wainwright et al. 2005). The state-space model included a hidden state for each of the eight directions of movement. We estimated a trial-to-trial generalization function from the direction in which error was experienced to all possible directions. This generalization function was a measure of sensitivity to error. The models predicted that changing the ITI would affect this sensitivity. In particular, Model 1 predicted that the sensitivity to error experienced in trial n should increase as a function of the delay that preceded that trial. Model 2 predicted that the sensitivity to error experienced in trial n should increase as a function of the delay that followed that trial.

Experiment 3: variable ITI with channel trials and random forces

The results of *experiment 2* were consistent with predictions of Model 2 but not Model 1. To test the assumptions of Model 2 more directly, we performed a final experiment. In this experiment, we measured state of the motor system directly through the use of “channel trials” (Hwang et al. 2006a; Scheidt et al. 2000; Smith et al. 2006). In a channel trial, the robot restricts the hand’s motion along a straight line to the target. Although it prevents errors during the reach, it allows us to measure how much force the subject expected to experience for that trial. This expectation is equal to the force that the subject produces against the channel wall. Previous work found that during reach adaptation, these forces gradually approximate the force field that the robot produces during free movements (Hwang et al. 2006). Our idea here was to use channel trials to measure the change in the expected force as a function of error in the previous reach trial.

Experiment 3 (Fig. 1C) consisted of three sets (each 192 trials), separated by short breaks: set A (null baseline), set B (random field), and set C (random field). We recruited a new group of naive subjects ($n = 28$). Unlike Experiments 1 and 2, however, there were only two targets: either up or down with respect to the center position. In the field sets subject were given, in a pseudorandom order, a clockwise curl-force field (Eq. 1, $+V$, 3/8 of the movements), a counterclockwise curl-force field ($-V$, 3/8 of the movements), or a “channel trial” in which movements perpendicular to the target direction were prevented by a stiff one-dimensional spring/damper (2 kN/m, 45 Ns/m). The seamless production of the channel force was based on the hand position in the center start box and unperceivable to participants unless they purposely tried to move in the perpendicular direction. Such movements were not observed and participants reported that they were not aware of this force pattern at all.

We were interested in the change in the force exerted by the subject against the channel walls when two channel trials were separated by a trial in which there was a movement error. There were 34 such channel–force–channel trial triplets dispersed pseudorandomly in set B and 35 in set C. In the triplet, between the first and the second movements, the ITI was kept at a constant 4 s. Between the second and the third movements, we varied the ITI in a random fashion identical to the *experiment 2* design. Thus for each triplet we looked

at the difference in the force output in the first and third movements (both were channel trials) as a function of the time passed since the error experienced in the second movement. Model 2 predicts that when one experiences a reach error in trial n , on trial $n + 1$ one will produce a force against the channel walls that is proportional to this error and this sensitivity will grow as a function of the time between trial n and trial $n + 1$.

Bootstrap methods for estimating the confidence limits

We followed the procedures described in our previous publications (Donchin et al. 2003), which followed procedures laid out by Efton and Tibshirani (1993), to estimate the SEs (Fig. 4) in the model parameters in the analysis of *experiment 2*. We resampled data from the 31 subjects with replacement and estimated the parameters from the averaged sample. We iterated this procedure 200 times and the SD of the 200 estimations of the parameters yielded an estimation of the SE of the parameters.

RESULTS

Experiment 1: spaced training resulted in faster adaptation

In *experiment 1*, we sought to replicate the results of Francis (2005) and Bock et al. (2005). We quantified the effect of massed versus spaced training in a within-subject design (Fig. 1A). Subjects were randomly assigned to train with either a long (13.60 ± 0.13 s, mean \pm SD) or short delay (3.88 ± 0.14 s) before each trial. We found that they showed better adaptation in the longer ITI set—compensation in field trials was stronger and aftereffects in catch trials were larger (Fig. 2, A–C). We measured performance with a learning index in each of the six 32-movement bins in each set. Repeated-measures three-factor ANOVA analysis (ITI, set order, and movement bin) on the learning index showed no significant effect resulting from set order [$F(1,21) = 0.020$, $P = 0.89$]. We therefore combined data from the two groups and considered the effect of ITI on the learning (Fig. 2D). We found a significant interaction effect between ITI and movement bin [$F(5,105) = 4.07$, $P = 0.002$]. Post hoc analysis showed significantly better performance in bins 2, 3, and 4 within subjects (paired t -test, $P < 0.05$ for each bin) for the long ITI set. Therefore the adaptation rate was enhanced with the longer ITI.

We checked whether the gains in performance with longer ITI might have arisen from a fatiguelike process. In our task, forces that counter the perturbing field are about one third of the forces that move the arm toward the target (Bhushan and Shadmehr 1999). Therefore if the limb fatigues with short ITIs, forces that move the limb toward the target should show a positive correlation with respect to the time passed since the last trial. As a proxy for this force, we looked at the magnitude of peak velocity vectors parallel to the direction of the target—a measure highly correlated with fatigue (de Haan et al. 1989; Jaric et al. 1997). If the limb fatigues with smaller ITIs, movements should become slower with smaller ITIs. Peak velocity was not different in the two ITI groups (two-tailed t -test, $t = -0.22$, $df = 44$, $P = 0.83$). In addition to the velocity measure, we would expect the fatigued group to show a lower performance as a result of the inability to express learning. However, the two groups attained similar performance toward the end of the training session (Fig. 2D). Together these results suggested that differences in perfor-

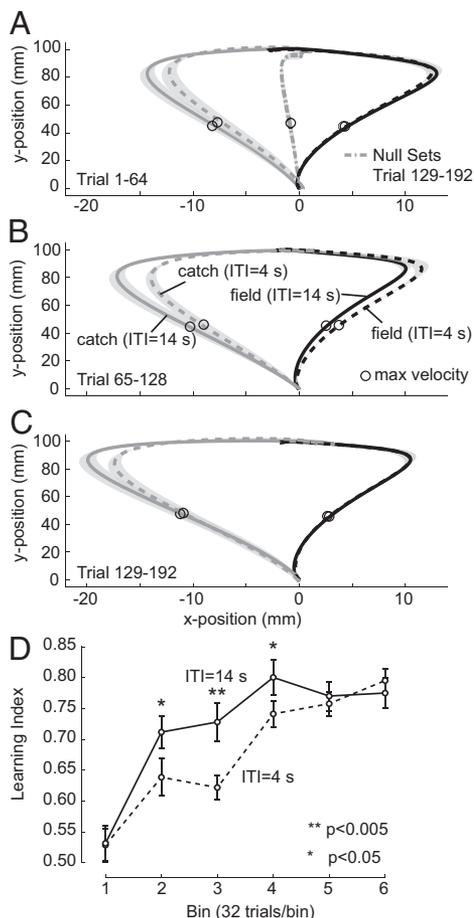


FIG. 2. Mean movement paths averaged across subjects, directions, and training sets in *experiment 1*. Shadings indicate the SE trajectory averaged across subjects. *A*: mean movement paths in the early stage of training (field and catch trials). *B*: mean movement paths in the middle stage. *C*: mean movement paths in the late stage. *D*: comparison of learning performance between short and long ITI movement sets. Each sequence of movement data were grouped into 32-movement bins and a learning index was computed for each bin. Learning indices were averaged across subjects and sets of the same ITI. Within-subject performances were significantly different for bins 2, 3, and 4 (paired *t*-test, $P < 0.05$ for each within-subject comparison). Presence of catch trials limits the theoretical ceiling of learning indices to 0.83. Error bars are SE.

mance were unlikely attributable to a potential for fatigue in the short ITI group.

Existing models and predictions

To account for the observation that performance during motor learning exhibited time-dependent changes such as savings and spontaneous recovery, our group previously proposed a deterministic multirate model that suggests that motor output is a sum of at least two systems: a fast adapting system that rapidly forgets and a slow adapting system that has good retention (Smith et al. 2006). In this model, the learner's state (represented by variable z) is a reflection of two underlying systems: one system (represented by state z_1) is highly sensitive to errors and changes rapidly, but has a limited capacity and will tend to quickly forget. Another system (represented by state z_2) has low sensitivity to error and changes slowly, but has large capacity and will tend to remember. The learner's "knowledge" on each trial is a sum of contribution of the fast

and slow systems: $z = z_1 + z_2$. In this model, after an error $y^{(n)}$ in trial n , the state of the learner $\{z_1, z_2\}$ changes deterministically as follows

$$z_1^{(n-1)} = a_1 z_1^{(n)} + b_1 y^{(n)} \quad (\text{fast system})$$

$$z_2^{(n+1)} = a_2 z_2^{(n)} + b_2 y^{(n)} \quad (\text{slow system}) \quad (11)$$

In *Eq. 11* we have $a_1 < a_2 < 1$ and $b_2 < b_1 < 1$. In this equation, passage of time affects the learner through the variables a_1 and a_2 . Because these variables are < 1 , passage of time always degrades memory. Therefore the model cannot explain the finding of improved rates of learning in spaced training. However, a variation of this model casts the time-scales in a Bayesian framework (*Eq. 3*). In this framework, the learner keeps a measure of uncertainty about its knowledge. Importantly, the passage of time between trials affects this uncertainty. As uncertainty escalates, the learner's incentive to learn increases. Figure 3*A* shows a simulation of this method of learning. We simulated 150 trials in two conditions: with a short ITI (two model iterations between observations) or a long ITI (20 model iterations between observations). The longer ITI produced faster rates of adaptation. To see the reason for this, it is instructive to examine the parameter uncertainties in the long ITI scenario (Fig. 3*B*). With each observation (i.e., trial), the learner acquires information and therefore the uncertainty declines. During the interval between observations, the uncertainties increase. Yet, the rate of this increase is different for the fast and slow states. The faster state has higher noise and so its uncertainty rapidly increases until it reaches an asymptote. The slower state has less noise and so its uncertainty slowly increases during the ITI. Therefore the longer ITI disproportionately affects the uncertainty of the slow state. Increased uncertainty means an increased sensitivity to error that the learner experiences in the subsequent trial (*Eq. 5*). As a result, the Bayesian learner adapts faster with a longer ITI.

The Bayesian model explains that the learner adapts faster in the longer ITI design because his/her uncertainty grows with passage of time. Therefore this model predicts that the sensitivity to error in trial n will become larger with increased time between trials $n - 1$ and n . *Experiment 2* was designed to test this prediction.

The link between trial-to-trial sensitivity to error, generalization, and overall learning rate

In the Bayesian model, *Eqs. 5–8* describe how the state of the learner changes on a trial-to-trial basis. We showed that in simulation, an increased trial-to-trial sensitivity to error sufficiently expedites overall learning rate (Fig. 3*A*). The trial-to-trial sensitivity to error, however, is not uniform across all movement directions. Gandolfo et al. (1996) demonstrated that the amount of trial-to-trial sensitivity depends on the angular disparity between movement directions. That is, errors in one direction affect the states of the learner in other directions—a phenomenon termed *generalization*. Therefore if increased ITIs increase trial-to-trial learning and in turn overall learning rate, the effect will produce a modulation of generalization as a function of ITIs.

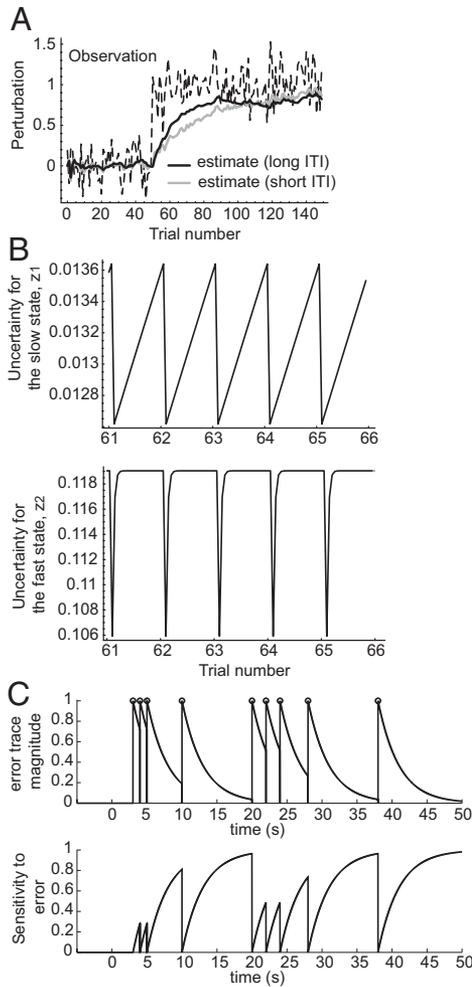


FIG. 3. Effect of ITI on the Bayesian and error trace models. *A*: simulation of the Bayesian multirate model. Thin black line represents the perturbation observed by the learner and the thick lines are the learner's estimate of the perturbation when observations are made with either a long (black) or a short ITI (gray). To produce these plots, Eqs. 7–9 were simulated iteratively with observation noise $\sigma_o^2 = 0.04$, state noises $\sigma_1^2 = 0.00008$ and $\sigma_2^2 = 0.1$, and forgetting rates $a_1 = 0.999$ and $a_2 = 0.4$. We assumed that each iteration took a constant amount of time and defined a "trial" as an observation. For the short ITI, the learner made an observation on every other iteration of the simulation. For the long ITI, the learner made an observation on every 20th iteration of the simulation. Longer ITI produced a faster rate of adaptation. *B*: evolution of parameter uncertainty in the long ITI scenario for 6 consecutive trials. Learner estimates the perturbation as a sum of 2 parameters in Eq. 4, z_1 (slow state) and z_2 (fast state). There is an uncertainty (or variance) associated with each parameter. Increased uncertainty arising from a longer ITI before an observation (i.e., a "trial") will result in increased sensitivity to error in the following observation and thus faster rate of adaptation. When the learner makes an observation, the uncertainties decrease sharply. Between trials, the learner cannot make observations but estimation of uncertainties change. Uncertainty for the fast state rapidly increases and saturates. In contrast, the uncertainty for the slow state gradually increases. *C*: illustration of the error trace model. For simplicity, each observation/trial (circles) generates an exponentially decaying trace. An old error trace is replaced as soon as a new movement is made. A long interval after an observation allows more time to integrate the error trace and therefore increases the sensitivity to the preceding error.

Experiment 2: sensitivity to error in trial n increased with time between trials n and $n + 1$

In *experiment 2*, the delay between movements was drawn from a multinomial random variable. The ITIs were approximately 4, 9, 14, and 24 s [respectively, 4.14 ± 0.17 , $8.63 \pm$

0.14 , 13.63 ± 0.16 , or 23.70 ± 0.12 s (mean \pm SD)]. We were interested in measuring the learner's sensitivity to error in each trial as a function of the ITI. To estimate this sensitivity, we used a state-space approach (Donchin et al. 2003). For each subject we measured the error $\mathbf{y}^{*(n)}$ in each trial n , force $\mathbf{f}^{(n)}$ produced by the robot at that velocity, and target direction $L^{(n)}$. Because the movement directions were identical between subjects, we averaged the movement errors and forces across subjects and arrived at a single sequence.

To determine how the error in each trial affected the movement that the subject made in the subsequent trial, we fitted this sequence to a hidden state-space model (Donchin et al. 2003; Waincott et al. 2005). The hidden states represented the knowledge of the learner about the perturbation in each direction of movements. On experience of an error in a given direction, we estimated how this error was generalized to other directions. We also estimated how this generalization was modulated by the time spent between the trials. The sequence of movement errors were fitted to the following dynamical system

$$\begin{aligned} \mathbf{y}^{(n)} &= D\mathbf{f}^{(n)} - L^{(n)}\mathbf{z}^{(n)} \\ \mathbf{z}^{(n+1)} &= a(t_{n+1} - t_n)\mathbf{z}^{(n)} + BL^{(n)T}\mathbf{y}^{(n)}k(t_n - t_{n-1}) \end{aligned} \quad (12)$$

In this model, \mathbf{z} represents a vector of hidden states (learner's knowledge about the perturbations for each of the eight possible target directions). By fitting the observed variables [$\mathbf{y}^{*(n)}$, $\mathbf{f}^{(n)}$, $L^{(n)}$] to Eq. 12, we estimated the unknown parameters D (the arm's compliance), B (the generalization function), a (a time-dependent function that describes the deterioration of the state during the time between trials), and k (a time-dependent function that modulates the generalization function as a function of time between trials). The Bayesian model predicted that k would be a monotonically increasing function of Δ , where $\Delta = t_n - t_{n-1}$.

Using nonlinear optimization (the *lsqnonlin* function in Matlab with default settings), we fitted Eq. 12 to the measured data [$F(35,349) = 33.95$, $r^2 = 0.7730$, $P < 0.0001$]. The resulting generalization function B (with SEs of the mean estimated through a bootstrap procedure) is plotted in Fig. 4A. This function had its peak at 0° and decreased with angular distance in a pattern similar to those recorded in other studies where ITIs were kept constant (Donchin et al. 2003; Waincott et al. 2005). We found that the function a remained extremely close to 1 (Fig. 4B), suggesting that there was little or no forgetting during the seconds that passed between trials. Arm compliance D was consistent with previous measurements of limb compliance (Fig. 4C) (Mussa-Ivaldi et al. 1985). However, contrary to the predictions of the Bayesian model, we found that k did not monotonically increase with the time that preceded the trial (Fig. 4D, gray line).

The error trace model

Because the results of our fit were inconsistent with the Bayesian model, we considered an alternative model. It is possible that in spaced training, adaptation rate is faster because errors produce a memory trace that decays with time but that the learner continues to benefit from the trace during the ITI period (Fig. 3C), effectively continuing to learn from the error trace. In this model, the time that is of importance is the

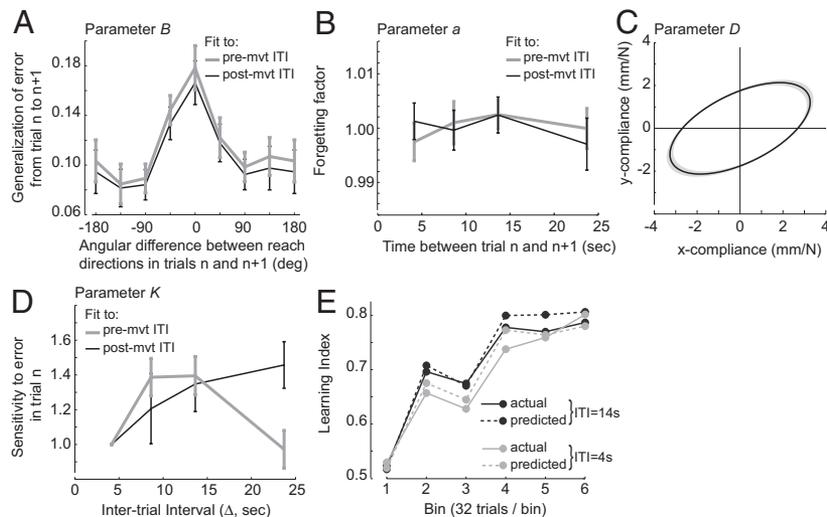


FIG. 4. Sensitivity to error as estimated from *experiment 2*. Data from *experiment 2* were fitted to *Eqs. 12* and *13*. For *Eq. 12*, it is the premovement ITI that affects the sensitivity to error. For *Eq. 13*, it is the postmovement ITI that affects the sensitivity to error. Bayesian model predicts that sensitivity to error should increase with premovement ITI. Trace model predicts that sensitivity should increase with postmovement ITI. Error bars are bootstrap estimates of SE. *A*: generalization function *B*, which quantifies the fraction of error experienced in trial *n* that affected motor output in trial *n* + 1. Function is parameterized with respect to the angular distance between trials *n* and *n* + 1. Error bars are bootstrap estimates of the SE (200 resamples). There were no significant differences in estimates of *B* for the 2 models. *B*: parameter *a*, which estimates the effect of time passage on the decay of memory. A value of 1 indicates no decay. *C*: parameter *D*, which estimates limb compliance. Estimates for the 2 models were nearly identical. *D*: parameter *k*, which estimates the sensitivity to error in trial *n*. For the premovement ITI model, the plot shows *k*(Δ), where $\Delta = t_n - t_{n-1}$, or $\Delta = t_{n+1} - t_n$. To estimate *k*(Δ), we arbitrarily set *k* = 1 for $\Delta = 4$ s and found the remaining values with respect to this value. *E*: cross-validation of the error trace model. We fitted *Eq. 13* to the data in *experiment 2*, resulting in the parameters shown in *A–D*, and then used these parameters to predict performance in *experiment 1*. Figure has the same format as Fig. 2*D*.

period that follows a trial, not the time that precedes it (*Eq. 10*). The error trace model predicts that sensitivity to error in trial *n* should increase as a function of the time between trials *n* and *n* + 1. (In contrast, the Bayesian model predicted that the sensitivity to error in trial *n* should increase a function of time between trials *n* – 1 and *n*.) To test for this prediction, we slightly modified *Eq. 12* to represent this idea and then fitted *Eq. 13* to the same data observed in *experiment 2*

$$\mathbf{y}^{(n)} = D\mathbf{f}^{(n)} - L^{(n)}\mathbf{z}^{(n)}$$

$$\mathbf{z}^{(n+1)} = a(t_{n+1} - t_n)\mathbf{z}^{(n)} + BL^{(n)T}\mathbf{y}^{(n)}k(t_{n+1} - t_n) \quad (13)$$

The results produced a highly significant fit [$r^2 = 0.7733$, $F(35,349) = 34.01$, $P < 0.0001$]. Although there were few or no changes in the model parameters *B*, *a*, or *D* (Fig. 4, *A–C*), the important parameter *k* changed significantly: it now became a function that monotonically increased with ITI (Fig. 4*D*). That is, results of *experiment 2* suggested that the sensitivity to error in trial *n* increased as a function of postmovement ITI, not premovement ITI. This result was consistent with the error trace model but not the Bayesian model.

Was the sensitivity to error significantly greater at longer ITIs? To check for this, we asked whether the slope of the function *k*(Δ) in *Eq. 13* was significantly greater than zero. We first computed *k*(Δ) from each of the bootstrapped groups and then estimated the slope by fitting a straight line. The *P* value was estimated by counting the number of nonpositive slopes and then dividing by the total number of bootstrapped samples. We found a significantly positive slope ($P < 0.005$). On average, the subjects' generalization was 45% higher during those trials with the longest ITI with respect to those with the shortest ITI.

What was the time constant of the error trace? To estimate this time constant, we replaced the term *k*(Δ) in *Eq. 13* with the

exponential function in *Eq. 10* and refitted the system of equations to the measured data. We again found a highly significant fit [$r^2 = 0.77$, $F(29,355) = 40.8$, $P < 0.0001$]. The fit estimated an error trace time constant of about 4 s ($r = 3.77 \pm 0.6$ s).

In summary, we found that improved performance in the spaced trials was attributed to increased error sensitivity as a function of the period that followed the movement, as would be predicted by the error trace model, and not the period that preceded the movement, as would be predicted by the Bayesian model.

Cross-validation of the error trace model

To validate the error trace model, we performed two tests. First, we asked whether the specific parameter values found in the random ITI of *experiment 2* could explain the performances in the constant ITI of *experiment 1*. Second, we performed *experiment 3* to specifically measure the change in motor output as a function of ITI.

In our first test, we asked whether the model of *Eq. 13* could predict the specific shape of the learning function that we had measured in the short- and long-delay conditions of Fig. 2*D*. We ran the dynamic system of *Eq. 13* on the target sequence $L^{(n)}$ and force sequence $\mathbf{f}^{(n)}$ of *experiment 1*, generating a sequence of errors in field and catch trials. We then computed a learning index on this sequence of movement errors in the same way that we had computed the performance of our subjects (*Eq. 2*). For simplicity, predicted and actual performances were computed on a single sequence of movements by first averaging data across subjects; the results were similar to those shown in Fig. 2*D*. Figure 4*E* shows performance of the predicted and the actual data from *experiment 1*. There was an

excellent correspondence between the predicted performance and the measured data.

Experiment 3: predictions of the error trace model

A crucial prediction of the error trace model is that when one experiences a motor error, the longer one waits, the larger will be the effect of this error on the motor output in the next trial. To test this prediction, we used force channels, first introduced by Scheidt et al. (2000). In a “channel” trial, the hand is guided and restricted along a straight line to the target. The motor output that is relevant to the task is the force that the subject produces against the walls of this channel. In a triplet of “channel–field–channel” trials, one can measure the change in motor output between the two channel trials as a function of the error in the intervening field trial. The prediction of the error trace model is that this change should monotonically increase with the time period between the field trial (when error was experienced) and the second channel trial.

In a channel trial n , let us label the force perpendicular to the direction of motion produced at maximum velocity as $u^{(n)}$. Assume $u^{(n)} = \hat{V}^{(n)}\dot{x}^{(n)}$, where $\hat{V}^{(n)}$ is the subject’s best guess at the constant V in Eq. 1, and that $\hat{V}^{(n+1)} = \hat{V}^{(n)}$ because the subject did not experience an error in trial n (as it was a channel trial). Now in trial $n + 1$ (field trial), the subject experienced a movement with error $y^{(n+1)}$, and this results in adaptation

$$\hat{V}^{(n+2)} = \hat{V}^{(n)} + s(\Delta)y^{(n+1)} \quad (14)$$

In Eq. 14, the learner’s trial-to-trial sensitivity to error is labeled with variable s . This sensitivity is a function of time between trial $n + 1$ and trial $n + 2$, that is: $\Delta = t_{n+2} - t_{n+1}$. Thus for any channel–force–channel triplet, we have

$$\begin{aligned} u^{(n+2)} &= \hat{V}^{(n+2)}\dot{x}^{(n+2)} \\ &= u^{(n)}\frac{\dot{x}^{(n+2)}}{\dot{x}^{(n)}} + s(\Delta)y^{(n+1)}\dot{x}^{(n+2)} \end{aligned} \quad (15)$$

We fitted this equation to each triplet in *experiment 3* and estimated sensitivity to the error experienced in movement $n + 1$ as a function of the time interval between movements $n + 1$ and $n + 2$ (Fig. 5). (We kept the time interval between movements n and $n + 1$ constant in this experiment.) The ITIs

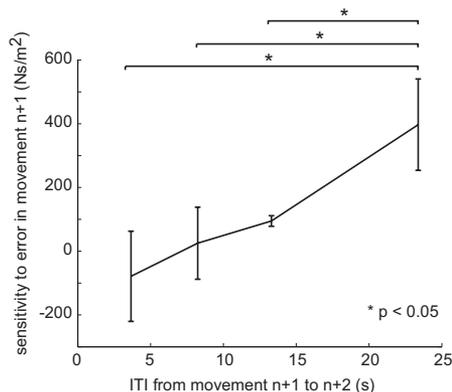


FIG. 5. Results of *experiment 3*. Sensitivity to random errors in trial $n + 1$ as a function of time between trial $n + 1$ and $n + 2$. (Time interval between trials n and $n + 1$ was kept constant at 4 s.) Parameter for sensitivity s was estimated by the change in the motor output u in channel trials n and $n + 2$ (Eq. 15). To arrive at this figure, Eq. 15 was fitted to the data for each subject, resulting in an estimate of $s(\Delta)$ for each subject. Values shown are the means \pm SE.

between movement $n + 1$ and $n + 2$ were 4, 8, 13, and 23 s [or, more precisely, 3.66 ± 0.05 , 8.23 ± 0.06 , 13.30 ± 0.05 , and 23.37 ± 0.07 s (mean \pm SD)]. There was a significant effect of the time interval between movements $n + 1$ and $n + 2$ [one-way ANOVA, $F(3,81) = 3.156$, $P = 0.029$] and one-tailed t -test revealed that the sensitivity was significantly higher for longer ITIs than for shorter ones (Fig. 5). Therefore consistent with the predictions of the error trace model, we observed that when one experienced an error, one learned more from that error if one waited longer before the next trial.

DISCUSSION

Two previous reports demonstrated that reach adaptation required fewer training trials when trials were spaced in time (Bock et al. 2005; Francis 2005)—a finding that we reproduced in *experiment 1*. (In terms of absolute time, long ITI sessions took longer to complete.) These data demonstrate that the motor system is affected by not only motor error, but also time. How does time affect the way the brain learns from motor error?

We considered two models of motor adaptation that are sensitive to passage of time: a Bayesian multirate model (Eq. 3) and an error trace model (Eq. 10). Our previous work suggested that motor errors result in an adaptive response in at least two “systems”: a fast system that rapidly learns but has poor retention and a slow system that is less sensitive to error but hardly forgets (Smith et al. 2006). Unfortunately, this model could not explain the results of *experiment 1*. A Bayesian variant of this model recasts it in a probabilistic framework (Kording 2007). It hypothesizes that spaced training leads to improved rates of adaptation because during the time between trials $n - 1$ and n , the brain becomes uncertain about its internal model. The increased uncertainty results in increased sensitivity to subsequent motor errors. The model predicts that if one could measure error sensitivity on each trial, one would find that the sensitivity to error in trial n increases as a function of the time between trials $n - 1$ and n . In *experiment 2*, we tested and found results that were inconsistent with this prediction: the sensitivity to error in trial n monotonically increased with the time period that *followed* that trial, not the period that *preceded* that trial. Therefore spaced training improved rates of performance not because time delay made the learner more sensitive to the error in the next movement, but because it made the learner learn more from the error in the last movement.

To account for this result, we proposed that the error might be represented by a trace that exponentially declined with time, effectively allowing the nervous system to learn from the trace for as long as it was available. Results of *experiment 2* suggested that the error trace had a time constant of 4 s. To test the model more directly, we performed a final experiment where movements experienced random errors, but were sandwiched between “channel” trials from which we could measure change in motor output from error. Consistent with the trace model, we found that the sensitivity to error experienced in trial n increased with the delay between trials n and $n + 1$.

We assumed that the error trace was “reset” by the next trial. We also considered the possibility that the error trace lingered beyond the immediate next trial. If the time constant of such a “lingering” trace is more than a few seconds, one expects that

the generalization function will be independent of the ITI. This was inconsistent with our results in the second experiment.

We believe cognitive strategies had minimal effect in our data. We examined the effect of conscious policy versus an implicit memory in a recent report and found that there was a high probability that subjects would become conscious of the force field pattern when the number of targets was small (three) and the forces were consistent (Hwang et al. 2006b). Furthermore, in that experiment we did find that subjects who became conscious of the force pattern had a small but significant boost on learning versus those who did not. In the current experiment, we designed our procedures to minimize this effect. First, we had eight targets rather than three. Second, we performed a control experiment where forces were random. Importantly, we observed that the effect of ITI in the random experiment was consistent with the ITI effect in the constant force field experiment. Furthermore, we asked subjects what they were thinking during the intertrial intervals in postexperimental questionnaires (*experiment 3*). No subject was aware of the force pattern (because it was random) and only one subject in 28 answered that it was an attempt to come up with a cognitive strategy. Because participants were required to manually keep the cursor centered during the wait, the vast majority answered that they were trying to focus on centering the cursor.

The idea that adaptation might take place during the time between trials is a common theme among computational models of learning where events produce an eligibility trace for synaptic plasticity. For example, Sutton and Barto (1981) suggested that a stimulus or error signal that excites a neuron may produce an eligibility trace on the neuron's synapses that acts as a low-pass filter of that input. When the input is removed, the trace declines exponentially in time. As long as the error and stimulus traces are available, their coincidence results in modification of the synapse associated with the stimulus. If each new error or stimulus cancels the trace of the previous input, such models predict that the effect of a given error should grow with the ITI between trials. The results of *experiments 2* and *3* are consistent with this framework.

For the reaching task considered here, a candidate area where such computations may be performed is the cerebellum (Smith and Shadmehr 2005). An eligibility trace (or in our terms, sensitivity to error) may be represented by the concentrations of second-order messenger chemicals. Several studies suggest that parallel fiber activity is responsible for the gradual rise of this trace (Kettner et al. 1997; Raymond and Lisberger 1998). Similar proposals were previously suggested for timed learning in delayed conditioning of eye blinks (Fiala et al. 1996); parallel fiber activity leads to increased phosphorylation of receptors over time and, in turn, reduces Purkinje cell firing during the interval between the sustaining conditioned stimulus (CS) and the onset of unconditioned stimulus (US). Interestingly, it was previously observed that delayed conditioned response can be learned if CS precedes the US by ≤ 4 s (Gormezano 1966), suggesting that such persistent phosphorylation of Purkinje cell receptors would have similar time course put forth here in the error trace model.

From a neurobiological perspective, synaptic changes that are produced by spaced training produce memories that are dependent on protein synthesis (Comas et al. 2004; Josselyn et al. 2001; Locatelli et al. 2002; Maldonado et al. 1997; Scharf

et al. 2002; Tully et al. 1994). Indeed, cellular response in animal models to spaced stimuli may prime additional memory traces and give rise to resistance to memory interference in different temporal phases (Isabel et al. 2004). However, such cellular processes generally occur on a much longer timescale than what we examined here. The structure of our model might imply that the error trace is kept in some kind of buffer that continues to benefit the learner. Any mechanism with which the influence of the error on the next trial can grow as a function of ITI will produce the same result. What might be the neural basis of such a mechanism?

Neurons that are stimulated with longer ITIs produce larger long-term potentiation (Scharf et al. 2002) and are more resistant to depotentiation. Staddon et al. (Staddon and Higa 1996; Staddon et al. 2002) and Fusi et al. (2005) proposed a cascade model of synaptic plasticity that can account for this. For example, in the model of Fusi et al. (2005), a given synaptic strength is supported by a synaptic state that may be shallow or deep in its cascade. The probability of transition in the synaptic state depends on the depth of that state: the deeper the state, the more resistant it is to change. If we imagine that it takes time for the internal state of the synapse to transition from one depth in its cascade to another and that the time needed increases with the depth of the state, then events that can cause synaptic change are more effective when they come spaced in time. At short ITI, only those synapses change that have a shallow internal state. With increased ITI, one engages not only the shallow state synapses, but also has a higher likelihood of causing state change in the deeper state synapses. Neural models of adaptation that rely on such synapses should exhibit the ITI-dependent patterns of generalization that we found here.

There are a number of limitations to our error trace model of motor adaptation. By itself, the model cannot account for much of the rich body of data that was recently highlighted in Smith et al. (2006). For example, if there is an error trace, at this point we do not know how that trace affects the fast and slow systems that were inferred from that study. To account for those data, one idea is to combine the error trace model with the Bayesian model (Kording 2007) so that the effect of an observed error is a memory trace that decays in time. It was previously suggested that uncertainty of a task variable is encoded in the lateral intraparietal area (Platt and Glimcher 1997; Schall and Thompson 1999) and acetylcholine and norepinephrine were suggested to play crucial roles in forming the context-dependent priors during learning (Dayan and Yu 2006; Yu and Dayan 2005). Our data do suggest that this uncertainty does not change during the delay period between trials.

ACKNOWLEDGMENTS

The authors acknowledge J. T. Francis for discussion in the preliminary experiments; O. Donchin, J. Diedrichsen, and M. Smith for discussion of the modeling; and H. Chen in preparing the manuscript.

GRANTS

This work was supported by National Institute of Neurological Disorders and Stroke Grant NS-37422 and the Human Frontier Research Program.

REFERENCES

Aboukhalil A, Shelhamer M, Clendaniel R. Acquisition of context-specific adaptation is enhanced with rest intervals between changes in context state,

- suggesting a new form of motor consolidation. *Neurosci Lett* 369: 162–167, 2004.
- Bhushan N, Shadmehr R.** Computational nature of human adaptive control during learning of reaching movements in force fields. *Biol Cybern* 81: 39–60, 1999.
- Bock O, Thomas M, Grigorova V.** The effect of rest breaks on human sensorimotor adaptation. *Exp Brain Res* 163: 258–260, 2005.
- Comas D, Petit F, Preat T.** Drosophila long-term memory formation involves regulation of cathepsin activity. *Nature* 430: 460–463, 2004.
- Commins S, Cunningham L, Harvey D, Walsh D.** Massed but not spaced training impairs spatial memory. *Behav Brain Res* 139: 215–223, 2003.
- Crisicimagna-Hemminger SE, Donchin O, Gazzaniga MS, Shadmehr R.** Learned dynamics of reaching movements generalize from dominant to nondominant arm. *J Neurophysiol* 89: 168–176, 2003.
- Dayan P, Yu AJ.** Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network* 17: 335–350, 2006.
- de Haan A, Jones DA, Sargeant AJ.** Changes in velocity of shortening, power output and relaxation rate during fatigue of rat medial gastrocnemius muscle. *Pfluegers Arch* 413: 422–428, 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.
- Ebbinghaus H.** *Memory: A Contribution to Experimental Psychology*. New York: Dover, 1964.
- Efron B, Tibshirani RJ.** *An Introduction to the Bootstrap Monographs on Statistics and Applied Probability*. New York: Chapman & Hall, 1993.
- Fiala JC, Grossberg S, Bullock D.** Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *J Neurosci* 16: 3760–3774, 1996.
- Francis JT.** Influence of the inter-reach-interval on motor learning. *Exp Brain Res* 167: 128–131, 2005.
- Fusi S, Drew PJ, Abbott LF.** Cascade models of synaptically stored memories. *Neuron* 45: 599–611, 2005.
- Gandolfo F, Mussa-Ivaldi FA, Bizzi E.** Motor learning by field approximation. *Proc Natl Acad Sci USA* 93: 3843–3846, 1996.
- Gormezano I.** Classical conditioning. In: *Experimental Methods and Instrumentation in Psychology*, edited by Sidowski JB. New York: McGraw-Hill, 1966.
- Han JS, Gallagher M, Holland P.** Hippocampal lesions enhance configural learning by reducing proactive interference. *Hippocampus* 8: 138–146, 1998.
- Hwang EJ, Donchin O, Smith MA, Shadmehr R.** A gain-field encoding of limb position and velocity in the internal model of arm dynamics. *PLoS Biol* 1: e25, 2003.
- Hwang EJ, Smith MA, Shadmehr R.** Adaptation and generalization in acceleration-dependent force fields. *Exp Brain Res* 169: 496–506, 2006a.
- Hwang EJ, Smith MA, Shadmehr R.** Dissociable effects of the implicit and explicit memory systems on learning control of reaching. *Exp Brain Res* 173: 425–437, 2006b.
- Isabel G, Pascual A, Preat T.** Exclusive consolidated memory phases in Drosophila. *Science* 304: 1024–1027, 2004.
- Jaric S, Radovanovic S, Milanovic S, Ljubisavljevic M, Anastasijevic R.** A comparison of the effects of agonist and antagonist muscle fatigue on performance of rapid movements. *Eur J Appl Physiol Occup Physiol* 76: 41–47, 1997.
- Josselyn SA, Shi C, Carlezon WA Jr, Neve RL, Nestler EJ, Davis M.** Long-term memory is facilitated by cAMP response element-binding protein overexpression in the amygdala. *J Neurosci* 21: 2404–2412, 2001.
- Kettner RE, Mahamud S, Leung HC, Sitkoff N, Houk JC, Peterson BW, Barto AG.** Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement. *J Neurophysiol* 77: 2115–2130, 1997.
- Kording KP, Tenenbaum T, Shadmehr R.** The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat Neurosci* In press.
- Locatelli F, Maldonado H, Romano A.** Two critical periods for cAMP-dependent protein kinase activity during long-term memory consolidation in the crab Chasmagnathus. *Neurobiol Learn Mem* 77: 234–249, 2002.
- Maldonado H, Romano A, Tomsic D.** Long-term habituation (LTH) in the crab Chasmagnathus: a model for behavioral and mechanistic studies of memory. *Braz J Med Biol Res* 30: 813–826, 1997.
- Mussa-Ivaldi FA, Hogan N, Bizzi E.** Neural, mechanical, and geometric factors subserving arm posture in humans. *J Neurosci* 5: 2732–2743, 1985.
- Platt ML, Glimcher PW.** Responses of intraparietal neurons to saccadic targets and visual distractors. *J Neurophysiol* 78: 1574–1589, 1997.
- Raymond JL, Lisberger SG.** Neural learning rules for the vestibulo-ocular reflex. *J Neurosci* 18: 9112–9129, 1998.
- Savion-Lemieux T, Penhune VB.** The effects of practice and delay on motor skill learning and retention. *Exp Brain Res* 161: 423–431, 2005.
- Schall JD, Thompson KG.** Neural selection and control of visually guided eye movements. *Annu Rev Neurosci* 22: 241–259, 1999.
- Scharf MT, Woo NH, Lattal KM, Young JZ, Nguyen PV, Abel T.** Protein synthesis is required for the enhancement of long-term potentiation and long-term memory by spaced training. *J Neurophysiol* 87: 2770–2777, 2002.
- Scheidt RA, Dingwell JB, Mussa-Ivaldi FA.** Learning to move amid uncertainty. *J Neurophysiol* 86: 971–985, 2001.
- Scheidt RA, Reinkensmeyer DJ, Conditt MA, Rymer WZ, Mussa-Ivaldi FA.** Persistence of motor adaptation during constrained, multi-joint, arm movements. *J Neurophysiol* 84: 853–862, 2000.
- Shadmehr R, Mussa-Ivaldi FA.** Adaptive representation of dynamics during learning of a motor task. *J Neurosci* 14: 3208–3224, 1994.
- Smith MA, Ghazizadeh A, Shadmehr R.** Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol* 4: e179, 2006.
- Smith MA, Shadmehr R.** Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J Neurophysiol* 93: 2809–2821, 2005.
- Staddon JE, Chelaru IM, Higa JJ.** Habituation, memory and the brain: the dynamics of interval timing. *Behav Process* 57: 71–88, 2002.
- Staddon JE, Higa JJ.** Multiple time scales in simple habituation. *Psychol Rev* 103: 720–733, 1996.
- Sutton RS, Barto AG.** Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev* 88: 135–170, 1981.
- Thoroughman KA, Shadmehr R.** Learning of action through adaptive combination of motor primitives. *Nature* 407: 742–747, 2000.
- Tully T, Preat T, Boynton SC, Del Vecchio M.** Genetic dissection of consolidated memory in Drosophila. *Cell* 79: 35–47, 1994.
- Wainwright SK, Donchin O, Shadmehr R.** Internal models and contextual cues: encoding serial order and direction of movement. *J Neurophysiol* 93: 786–800, 2005.
- Yu AJ, Dayan P.** Uncertainty, neuromodulation, and attention. *Neuron* 46: 681–692, 2005.