

Active Learning: Learning a Motor Skill Without a Coach

Vincent S. Huang,¹ Reza Shadmehr,¹ and Jörn Diedrichsen²

¹Laboratory for Computational Motor Control, Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland; and ²School of Psychology, Bangor University, United Kingdom

Submitted 3 October 2007; accepted in final form 25 May 2008

Huang VS, Shadmehr R, Diedrichsen J. Active learning: learning a motor skill without a coach. *J Neurophysiol* 100: 879–887, 2008. First published May 28, 2008; doi:10.1152/jn.01095.2007. When we learn a new skill (e.g., golf) without a coach, we are “active learners”: we have to choose the specific components of the task on which to train (e.g., iron, driver, putter, etc.). What guides our selection of the training sequence? How do choices that people make compare with choices made by machine learning algorithms that attempt to optimize performance? We asked subjects to learn the novel dynamics of a robotic tool while moving it in four directions. They were instructed to choose their practice directions to maximize their performance in subsequent tests. We found that their choices were strongly influenced by motor errors: subjects tended to immediately repeat an action if that action had produced a large error. This strategy was correlated with better performance on test trials. However, even when participants performed perfectly on a movement, they did not avoid repeating that movement. The probability of repeating an action did not drop below chance even when no errors were observed. This behavior led to suboptimal performance. It also violated a strong prediction of current machine learning algorithms, which solve the active learning problem by choosing a training sequence that will maximally reduce the learner’s uncertainty about the task. While we show that these algorithms do not provide an adequate description of human behavior, our results suggest ways to improve human motor learning by helping people choose an optimal training sequence.

INTRODUCTION

A game like golf involves learning a number of specific subskills. Each golf club has a different weight and length and therefore varying degree of difficulty and accuracy. At the driving range a novice golfer gets a bucket of golf balls and practices with each golf club (i.e., skill components). A successful player needs to develop a high proficiency with all clubs. While we may have a coach that closely supervises our training and tells us which club to practice, most of our training time is spent alone: we must autonomously choose the skill component to practice. What factors affect our choices during training?

In the field of machine learning, this problem is termed “active learning.” The learner has access to a list of training examples; each example focuses on a different component of the task (e.g., a golf swing using the driver). There is no preset curriculum nor is there an instructor; therefore the learner picks out its own curriculum, one practice example at a time. Examples can be repeated if necessary, but the ultimate goal is to improve in the overall task performance with a minimum number of examples. How should one pick the training examples?

There are a number of potential criteria for such a decision. For example, one could use random exploration to minimize statistical bias, or select examples to minimize the element of surprise (Einhauser et al. 2007). In general, we are often faced with decision whether to exploit known alternatives or explore unknown ones. How the nervous system makes decisions in these situations remains an open question (Cohen et al. 2007). However, Cohn et al. (1996) suggested that in the case of active learning (i.e., choosing training examples), an optimal solution can be obtained if the goal is to minimize the uncertainty of the learner. In the current study, we will apply this approach in a human motor learning task and test whether behavior is well described by such a selection criterion.

Active learning

Mathematically, we can summarize active learning of a motor task as follows. The student has to learn a task consisting of P different components or behaviors, indicated by the multinomial variable x between 1 and P . A task component can be defined as a subtask in which mastery is required for the attainment of the overall task goal. The correct or optimal motor output is an observable but currently unknown function of the task component, $y(x)$. The learned behavior for each task component can be described by a set of parameters. For simplicity and without loss of generality, we assume here that each task component is described by a single parameter. Thus in our simple case, the learner has a set of parameters $[w^{(1)} \dots w^{(P)}]$ representing the knowledge of the movement required for each task component. For example, in our task, the subject was asked to compensate for a force perturbation when making reaching movements toward a number of different targets. Here $w^{(1)}, \dots, w^{(P)}$ represent the estimated force that the subject needs to produce for each movement direction to counteract the force perturbation. In other words, the parameters constitute the learner’s model of the task.

On trial n with component x_n , the learner produces the movement $\hat{y}_n = w_n^{(x_n)} + \varepsilon_n$, the estimated force plus random motor error ε_n , and observes the correct answer y_n . A general class of learning rules is gradient descent: The learner updates his own knowledge by the difference between actual and desired output, using the learning gain K_n (Donchin et al. 2003; Huang and Shadmehr 2007; Thoroughman and Shadmehr 2000)

$$w_{n+1}^{(x_n)} = w_n^{(x_n)} + K_n(y_n - \hat{y}_n) \quad (1)$$

The problem of active learning in this context is to pick the next training example x_n such that overall performance in the

Address for reprint requests and other correspondence: V. Huang, 710 W. 168th St., Rm. 13-12, Motor Performance Laboratory, Columbia University Neurological Institute, New York, NY 10033 (E-mail: vh2181@columbia.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.

task space can improve. In coached (i.e., passive) learning and most motor learning experiments, this problem does not arise because x_n is determined by the teacher or experimenter. Now what would constitute a good criterion to choose x_n ? Cohn and colleagues (1996) suggest that a learning system should attempt to reduce the expected squared error on the trial *after* the learning trial n . That is, the learner should try to pick a training example on trial n so that after he or she has learned from this training example following Eq. 1, the performance averaged across all skill components will be maximally improved.

For an unbiased learner (a learner who does not show systematic constant errors), the expected squared error on trial $n + 1$ is equal to the uncertainty the learner has about his learned output \hat{y} . Therefore good training examples will maximally reduce the uncertainty on the parameters $w^{(1)}, \dots, w^{(P)}$ after the learning. As we show in APPENDIX 1, this translates into a selection rule in which the learner should pick the task component in which the uncertainty prior to the learning trial is highest. This task component then will maximally benefit from learning on trial n , therefore also maximally reducing overall uncertainty. This decision rule is optimum for a number of different learning rules, including gradient descent with a constant learning rate, as well as learning using a Kalman filter (APPENDIX 1). Furthermore the decision rule also holds for a learner who forgets from trial to trial (APPENDIX 2), as long as the rate of forgetting is tuned to the rate of change in the environment (Koerding et al. 2007).

In all of these situations, learning in general leads to a reduction of uncertainty for the practiced task component. Therefore a learner who follows such a rule would not pick the same training example twice in a row; after practicing task component x on trial n , a different skill component will have a higher uncertainty. We ask here whether active learning in people follows this prediction and whether it can be described using an uncertainty-based decision rule.

Furthermore, we ask whether the choices during active learning are influenced by the error experienced during the last movement. When both the parameter and the motor noise ϵ_n are assumed to have Gaussian distribution, as is the case in the standard Kalman filter, the uncertainty of the model parameters will always be reduced for the component of the task that was practiced independent of the error observed. This would lead to the counterintuitive prediction that errors during active learning would not affect action selection.

However, there are versions of the Kalman filter that would increase the parameter uncertainty when a big error is observed. One example is a system where the output noise does not have a Gaussian distribution but is drawn from a mixture of Gaussians of identically zero means and different variances. In such case, the observation of a large error would lead to a reduced rate of learning (Kording and Wolpert 2004), and an increase in model uncertainty. Under the uncertainty-based selection rule, this would lead to repetitions of training examples when a large error is observed. However, any versions of such model would still predict that a learner should decrease its uncertainty when a small error is observed. Therefore any uncertainty-based selection model predicts that the learner is biased to not repeat an action when an error close to zero is observed.

To summarize, the uncertainty-based selection rule (Cohn et al. 1996) predicts that an optimal learner should choose the

task component with the highest uncertainty. After that choice is made, the learner should be biased to not immediately repeat the same action on the next trial. Here we have taken the first steps in quantifying the factors that affect human choices in active learning of a motor skill. Can their choices be understood in the framework of uncertainty estimation? Do motor errors affect our choice of training? In the present study, we used a force-adaptation task in which subjects learned the task of compensating for force perturbation when making reaching movements. In *experiment 1*, we observed how human participants select their own sequence of actions to practice and how motor errors influence these choices. In *experiment 2*, we further explored the idea that the variance of observation might influence the subjective estimation of its reliability. We tested the hypothesis that the consistency of observation was a contributing factor to human participants' strategies by varying the variances of the perturbation applied to each of the four task components.

METHODS

Hitting game

We used a hitting game to examine how the statistical properties of the training experience influenced the subjects' strategy in active learning (Fig. 1, A–D). The goal of the game was to become as proficient as possible at hitting a small target in one of four directions with a rapid center-out strike using a robotic manipulandum (Fig. 1A) (Huang and Shadmehr 2007; Hwang et al. 2003). Subjects' proficiency was rewarded with points in randomly interspersed test trials. During test trials, the computer chose the target for the subject. The closer the hand cursor came to the target, the greater the number of points. The score and the financial incentive depended solely on the performance on these test trials.

There were four possible targets, arranged on an invisible circle of a 10-cm radius. The targets were positioned at $-15, 75, 165,$ and 255° such that they were perpendicular to each other to minimize generalization of motor adaptation (Donchin et al. 2003). Hand position was displayed at all times as a 0.5×0.5 -cm white square cursor. At the beginning of each trial, the robot brought the hand to the center mark—a stationary 0.5×0.5 -cm white square. Targets were then displayed as red squares. After a short, variable delay, the targets turned white and the center mark vanished—this was the “go” signal for the center-out strike. As the movement crossed the invisible 10-cm radius circle, a yellow dot appeared at the crossing point to emphasize the distance between the strike and the goal as a measure of reach error. If the movement duration was too long (>0.23 s), a blue dot appeared instead. Beyond the invisible circle an elastic force field acted as a “pillow” to absorb the strike. On some trials the robot perturbed the movement with a velocity-dependent force field (see following text), which deflected the movement in clockwise or counter clockwise direction.

Because there were multiple targets present during active training trials, we needed to ascertain toward which target the subject was aiming. Therefore after each strike when the hand hit the pillow, subjects brought their hand back to the center of their intended target. At this point the center mark reappeared and the robot brought the hand back to the center.

Active training, passive training, and test trials

Participants were tested on randomly interspersed test trials in which one direction was chosen at random for them. In between test trials, participants trained for directions of their choice (active learning) or a direction chosen randomly for them (passive learning) in either an active

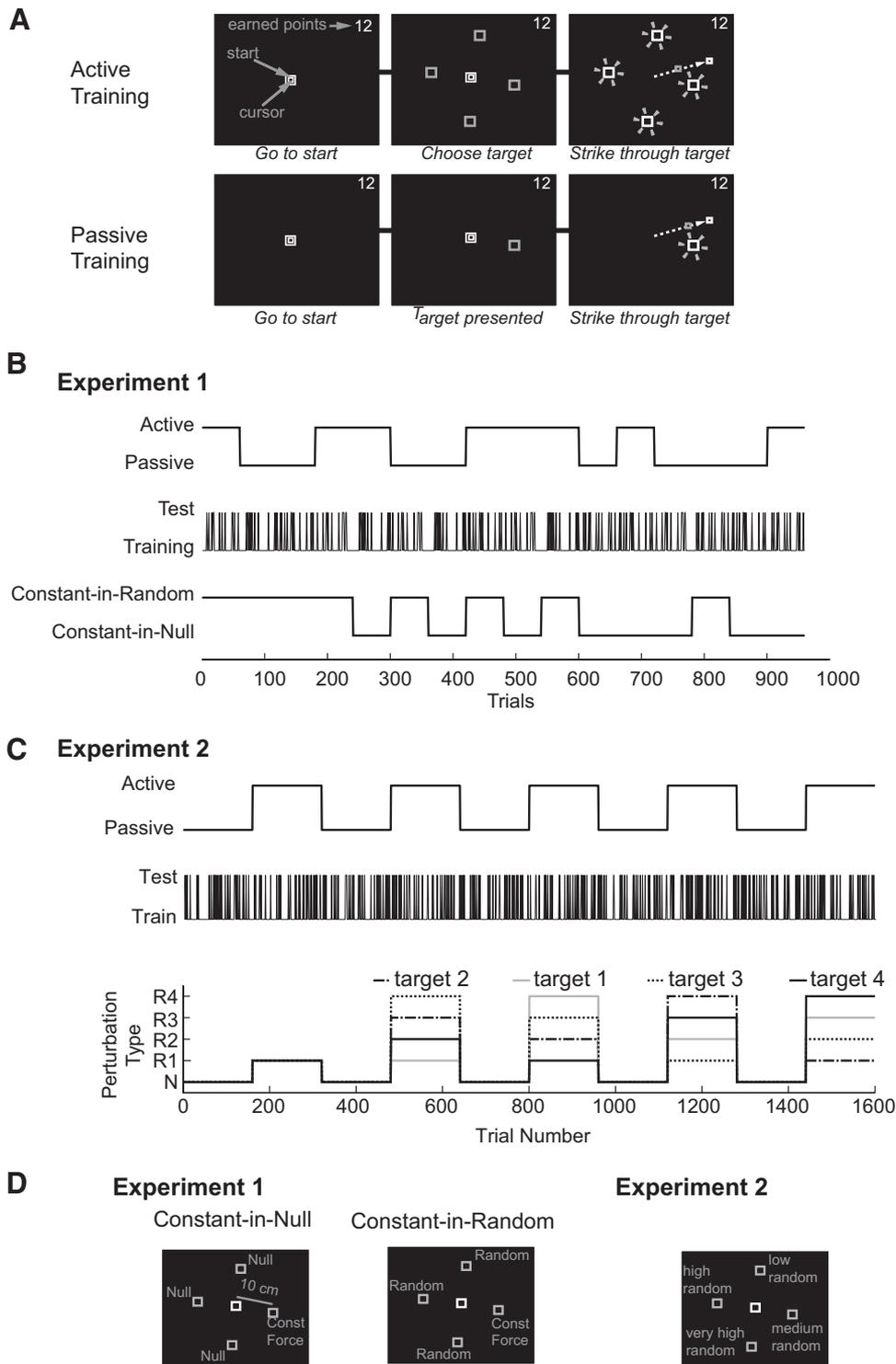


FIG. 1. *A*: experimental design. During active training sets, subjects chose the target of their movement. During passive training sets, the computer chose the target. Participants were tested for their performance at a random direction chosen by the computer during test trials. *B*: an example of *experiment 1* protocol for 1 subject. Each subject performed 16 sets of movements with pseudorandomly ordered conditions. *C*: an example of *experiment 2* protocol for 1 subject. *D*: force-target association patterns. Unknown to subjects, each of the 4 targets was associated with a type of perturbation during a block. In constant-in-null pattern, 1 of the target directions was associated with a viscous curl-force perturbation while the others were not associated with any force. In constant-in-random pattern, 1 target was associated with a constant viscous curl-force perturbation, while the other 3 were associated with perturbations in random direction and magnitude.

learning or passive learning block (Fig. 1, *A* and *B*). Comparison between the test trials in each type of blocks allowed us to assess the efficacy of the learning strategies. Participants were instructed to pick their movement directions in training trials so that they would maximize their performance in the test trials. We awarded participants a monetary reward dependent on the amounts of points earned in the test trials. It was made very clear that the score only reflected their performance in the test trials and not their performance on the training trials. This was important as we did not want to contaminate subjects' strategies during training with a greedy element of selecting an easy target for monetary return.

Test trials were randomly interspersed between the training trials (1 of 5) and clearly announced with the word "test" on the screen immediately before the trial started. The computer pseudorandomly picked the target among the four available to test the participants' performance. Performance was measured as the angular distance to the target at the point where the cursor crossed an invisible circle of 10 cm. Four accuracy levels were established: 5.16, 4.49, 3.61, and 2.48°. For each additional accuracy level achieved, the movement was award one additional point in that trial for up to a maximum of 4 points.

Four of five trials were training trials. The experiments were divided into blocks of 60 trials (*experiment 1*) or 160 trials (*experiment 2*); each block was either active training or passive training. The schedule of active and passive training blocks is shown in Fig. 1, *B* (*experiment 1*) and *C* (*experiment 2*). In training trials of *active training blocks*, all four targets were available and the subject chose their target direction to aim. In training trials of *passive training blocks*, the computer pseudorandomly picked the target among the four available.

Experiment 1

Our objective was to determine whether errors that subjects experienced during active learning affected the subsequent choices that they made. To that aim, we considered two kinds of errors: errors that were due to a consistent perturbation and errors that were due to an inconsistent perturbation. To induce errors, we applied a velocity-dependent force (viscosity of 10 Ns/m) that pushed the hand perpendicular to the hand movements toward some target directions.

The perturbations of each block of trials followed one of two patterns (Fig. 1*D*). In the *constant-in-null* pattern, movements toward three of the four targets were unperturbed (null or “N”), whereas movements toward one target were perturbed with a consistent curl force field (constant or “C”). The C target was assigned pseudorandomly for each block. The C target had a clockwise field for the first 30 trials of the block, and then the field switched to a counter clockwise field. A good active learning policy would have been to find the C target and continue to train mostly on that target.

It is possible that subjects would choose the C target because it was the only target that had any perturbation. In the *constant-in-random* pattern, again one target was picked to have a constant perturbation (C) that switched after 30 trials. In contrast to a constant-in-null block, however, a curl field was also presented during movements to the three remaining targets. These curl fields switched randomly between clockwise or counterclockwise fields (random or “R”). In these movements, the field had a random viscosity with a uniform distribution from -10 to $+10$ Ns/m to further mask the presence of the C target.

Experiment 2

To further explore the idea that the variance of the perturbations—i.e., the reliability of errors—influenced choices during active learning, we conducted a second experiment where the mean of the perturbations associated to the various targets were identical, but their variance differed (Fig. 1, *C* and *D*). Once again, four targets were available. At the start of each block (now 160 trials long), each target was assigned a curl field with a viscosity that had a mean of 10 Ns/m but a variance that was low (R1 target, viscosity uniformly drawn from 6 to 14 Ns/m), medium (R2 target, viscosity uniformly drawn from 2 to 18 Ns/m), high (R3 target, viscosity uniformly drawn from -2 to 22 Ns/m), and very high (R4 target, viscosity uniformly drawn from -6 to 26 Ns/m). Therefore observations at the R1 target should be the most consistent through out the block. Similar to *experiment 1*, participants earned points only during the sparsely distributed test trials in which the computer randomly tested participants' performance in one of the four targets. During these test trials, the viscosity was always 10 Ns/m (C targets).

Softmax regression procedures

We modeled the probability of choosing a target using a generalized linear model. We used a multinomial extension of logistic regression—softmax regression. The probability vector, \mathbf{p} , of selecting target x_n ($x_n = 1 \dots P$) depends on the vector \mathbf{v}_n , which in turn was a linear function of three factors: the 4×1 vector Θ_{bias} , with the mean constrained to zero (i.e., 3 free parameters) modeled the preference of

participants for one of the four particular targets, θ_{repeat} modeled the preference of each subject to repeat the last movement direction x_n , and finally θ_{error} modeled the increase in probability to repeat the last movement direction, as the absolute size of the last error $|y_n|$ increased. By writing the last choice x_n as the vector of indicator variables \mathbf{x}_n , the full model can be written as

$$\mathbf{v}_n = \Theta_{\text{bias}} + \theta_{\text{repeat}}\mathbf{x}_{n-1} + \theta_{\text{error}}\mathbf{x}_{n-1}|y_{n-1}| \quad (2)$$

$$\mathbf{p}_n = \begin{bmatrix} p(x_n = 1|\mathbf{v}_n) \\ \vdots \\ p(x_n = P|\mathbf{v}_n) \end{bmatrix} = \text{softmax}(\mathbf{v}_n) = \frac{\begin{bmatrix} \exp(v_n^{(1)}) \\ \vdots \\ \exp(v_n^{(P)}) \end{bmatrix}}{\sum_{j=1}^P \exp(v_n^{(j)})} \quad (3)$$

We fitted the parameters Θ_{bias} , θ_{repeat} , and θ_{error} by maximizing the log-likelihood of the data given our model using numerical methods (Matlab `fminsearch`).

Participants

Sixteen subjects participated in *experiment 1* and another 16 in *experiment 2*. For *experiment 1*, the experiment was counter-balanced across subjects for the order of the perturbations (constant-in-null and constant-in-random) and training conditions (active and passive). Subjects were healthy, right-handed, and naïve to the purpose of the experiment. Procedures and protocols were approved by the Johns Hopkins Medicine Institutional Review Board and participants gave their written consent prior to the experiments.

RESULTS

Errors in the last movement influence action selection

If active learners estimate the uncertainty about the desired output and then choose to train on a component on which they are most unsure (APPENDIX 1, Eq. A9), they should have a tendency *not* to practice in the same direction as the last trial. Contrary to this prediction of uncertainty-based models, our participants repeated the last movement direction with a probability of $35 \pm 12\%$ (Fig. 2*A*), significantly higher than just choosing a direction at random [2-tail *t*-test, $t(15) = 3.46$, $P < 0.01$]. When the participants decided to switch, they picked each of the other three directions with equal probability.

The probability of repeating a direction was also strongly modulated by the amount of error that the subject experienced in the previous trial. In Fig. 2*B*, we plotted the probability of repeating the last direction as a function of the absolute size of the error on the last trial. We found that probability of repeating a direction was an increasing function of the error size [1-factor ANOVA, $F(9,159) = 5.51$, $P \ll 0.001$]. Therefore a larger error in trial n led to an increased likelihood of repeating the same direction in trial $n + 1$. This was the case for both blocks in which movement to the remaining targets were unperturbed (constant-in-null pattern) or perturbed by a random force field (constant-in-random pattern).

Participants repeated even well-learned task components

When there were little or no errors in a trial, the probability of repeating the target approached 25%, the rate of random selection. Under any uncertainty-based models, the observation of a zero error should have decreased the uncertainty about the

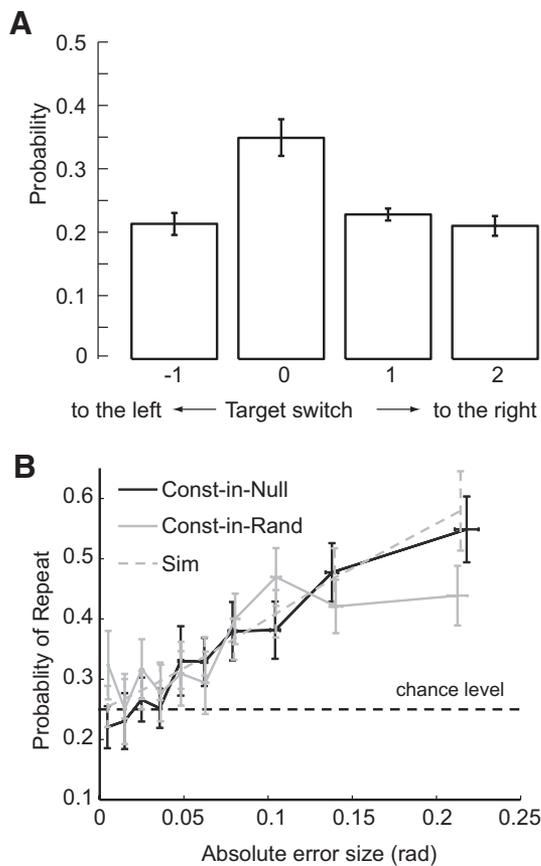


FIG. 2. *A*: probability of switching training directions from trial n to trial $n + 1$. The abscissa indicates the direction and the magnitude of the switch; 0 target switch means that the target of trial n was repeated in trial $n + 1$. White bars indicate the average probabilities of each switch options. Error bars indicate SE across participants. *B*: probability of repeating a target as a function of the absolute size of the recently experienced error, separated by the type of force-target patterns. ---, the average fit using the softmax regression model. Error bars indicate SE across participants.

corresponding task component. This would have then lowered the learner's probability of selecting this direction again below the probabilities of the other directions. While our data showed that error was a robust factor in encouraging repetition of a previously selected direction, a trial with a small error did not reduce the probability of re-selection of the same movement direction below chance as the uncertainty model had suggested. Participants, therefore clearly violated a fundamental prediction of uncertainty-based active learning models. Subjects completed a postexperimental questionnaire. While two subjects reported that they were repeating large error movements, cognitive responses were inconsistent across subjects.

To quantify these observations we estimated the contribution of error and the tendency to repeat a direction even in the absence of an error using softmax regression (a multinomial extension of logistic regression, see METHODS). The regression included a term to capture biases toward specific targets, a term that determined the probability of repeating a direction in the absence of error (θ_{repeat}), and a term that captured how much the probability of repeating increased with the absolute size of the last error (θ_{error}). From the estimated parameters, we were able to reproduce the sequence and trends of participants' choices (dashed line, Fig. 2*B*). While we did not find any significant bias toward any of the four targets [1-factor

ANOVA, $F(3,60) = 1.04$, $P = 0.38$], nearly all participants showed a positive θ_{error} , indicating that they were more likely to repeat an action when a large error was encountered [2-tailed t -test, $t(1,15) = 3.6$, $P < 0.01$]. Once we accounted for the size of the error, θ_{repeat} was not significantly different from zero, [$t(1,15) = 1.0$, $P = 0.32$]. That is, participants chose the well-learned movement direction just as likely as the other directions even when no error was encountered. This clearly violates the prediction of uncertainty-based selection models, as any variant of this model would have predicted a bias away from a just-practiced skill component when no error was encountered.

Relationship between selection strategy and performance

How did the participants' active learning strategies affect their performance? In general the average absolute errors in test trials were slightly, but not significantly bigger after active learning compared with passive learning blocks [paired t -test, $t(15) = -0.238$, $P = 0.82$]. We postulated that subjects might have used a combination of good (e.g., error-dependent repetition) and bad strategies (e.g., blind repetition). We looked at the correlation between strategy and performance on test trials after active learning. Because performance was determined largely by the overall proficiency of the participants, for each participant, we subtracted the average error during test trials after active learning from the average error after passive learning. The difference was then correlated with individual parameter estimates (θ_{repeat} and θ_{error}) from the softmax regression. A positive correlation of the parameter with the difference in errors indicated that this strategy facilitated learning, while a negative correlation indicated that this strategy hurt learning.

There was a positive correlation between error sensitivity and later active test performance (1-tail Spearman's correlation, $r = 0.48$, $P < 0.05$). Participants who sought to train in directions where their errors were big performed better in subsequent testing (Fig. 3, θ_{error}). Importantly, two participants that displayed error avoidance (negative values for θ_{error}) showed poorer performance relative to their own performance

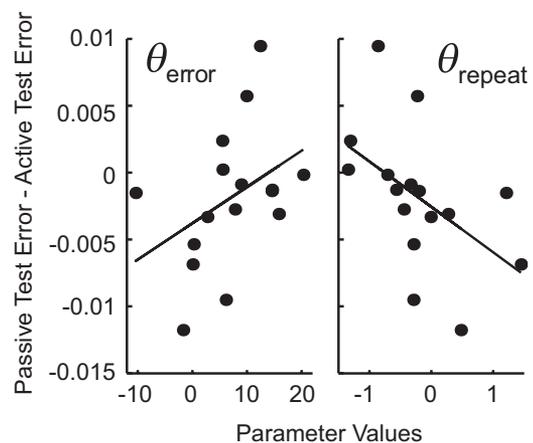


FIG. 3. Scatter plots of the values of model parameters (x axis) against the difference of the sizes of test trial errors between active training blocks and passive training blocks (y axis). Positive values on the y axis indicate better performance after active compared with passive learning. The tendency to repeat a movement direction when a large error was observed (θ_{error} , left) correlated positively with performance on test trials. The tendency to repeat a target even in the absence of an error (θ_{repeat} , right) led to worse outcomes.

after passive learning. Furthermore, we found a strong negative correlation between θ_{repeat} and subsequent test performance (Fig. 3). The more likely participants repeated a direction (after the influence of the error size has been accounted for) the worse was their test trial performance (1-tailed Spearman's correlation, $r = -0.60$, $P < 0.01$). Thus the violation of the optimal active learning strategy indeed hurt the performance of the participants in the active compared with the passive learning condition. These analyses show that the individual's active learning strategy influenced later performance on test trials. Furthermore, repetition of targets in the absence of errors led, as predicted by uncertainty-based models (Cohn et al. 1996), indeed to poorer learning outcomes.

Variance of the error signal

One reason to re-select the last direction despite no errors may be that one is trying to estimate the consistency of the observation. For example, if a participant found that for one direction the perturbation changed randomly from trial to trial, a good active learning strategy would be to ignore this direction because training here could not lead to further improvement. Did the variance of the perturbations affect the choices?

To test this idea, we introduced a condition—constant-in-random—in which one direction was perturbed with a constant force field (“C” target), whereas the other three were perturbed with a random force field (“R” targets). To assess the influence of consistency, we attempted to match the absolute sizes of the errors of the movements toward all the directions. Because participants would adapt to the constant force field, we introduced a stronger field in the constant target than in the random ones and flipped the perturbation direction after 30 trials (Fig. 4A). As a result the errors for the constant target were large immediately after the onset of the block and after the switch. However, the errors in the “C” target became smaller than the “R” targets by the end of each phase.

To account for these remaining differences in error, we selected trials that had similar performance in C and R targets (dotted trials in Fig. 4A; paired t -test for each trial and each participant, $P > 0.15$). For these trials, we found that the probability of choosing a “C” target was not different from choosing an “R” target [Fig. 4B, 2-tailed t -test, $t(30) = -0.43$,

$P = 0.67$]. Therefore variance of perturbations did not appear to influence choice.

While we attempted to match the absolute error size in *experiment 1*, the average size of the force field was different. Furthermore, the number of trials in a block (30) might have been too small to allow participants to estimate variance for the different skill components. To address these concerns and to test for the influence of error variance explicitly, we designed a second experiment in which the averages of the force perturbations were matched, and participants made substantially longer sequences of movements (160 per block). The perturbations associated with each target were drawn from a distribution that had identical mean (10 Ns/m). Each target in a block had a perturbation variance that was low, medium, high, or very high (Figs. 1D and Fig. 5B, abscissa). We labeled the corresponding targets R1, R2, R3, and R4 (Fig. 1C). Participants adapted to the mean force field for all four variance levels as seen in the decrease of mean error over many trials. However, trial-to-trial variance of errors remained high for movements with highly variable perturbation [1-way ANOVA, $F(3,63) = 30.45$, $P \ll 0.001$; Fig. 5, A and B, abscissa).

To determine whether the variance of the error influenced the choices during active learning, we first needed to account for the influence of the mean absolute error on choice because it increased with perturbation variance. We therefore used the same softmax regression approach as in *experiment 1*. As in *experiment 1*, we found that the error size of the last trial posed a significant influence on the target selection of next trial [2-tailed t -test, $t(15) = 2.54$, $P < 0.05$]. Participants also showed a slight tendency to repeat the last target even in the absence of error [2-tail t -test, $t(15) = 1.77$, $P = 0.09$], again violating the uncertainty-based models. Using these parameters, we then predicted the probability to practice on targets of each variance level assuming that participants did not have a bias toward either low- or high-variance targets (Fig. 5B). The observed probabilities were not significantly different from these predictions [2-factor ANOVA, $F(3,93) = 0.443$, $P = 0.72$]. For cross-validation purposes, we also used the parameters fitted using *experiment 1* data, and again, the observed probabilities were not significantly different [$F(3,93) = 1.123$, $P = 0.334$].

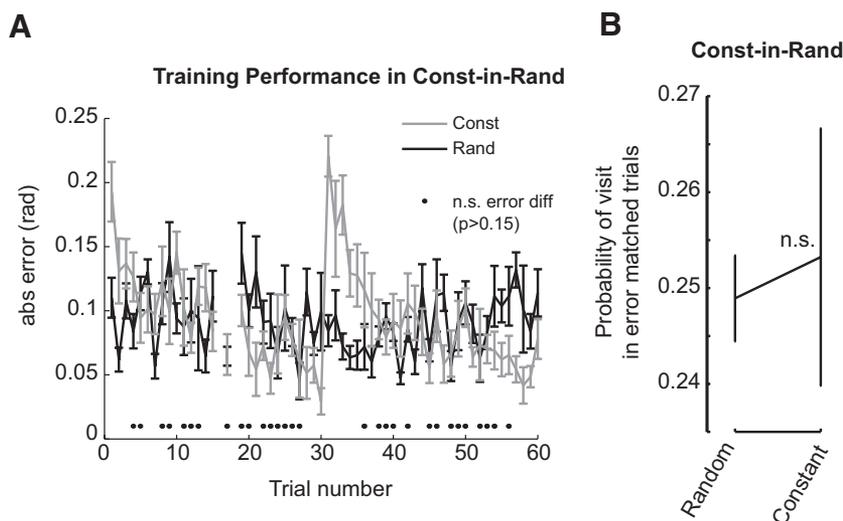


FIG. 4. *A*: average error for the constant and random targets as a function of trial number in active training. Graph shows initial high errors for the constant target with subsequent learning. ●, the trials that were picked in which the average absolute error size were matched for constant and random targets. The vertical bars indicate the SE across participants. *B*: overall probability of visiting a constant-force target vs. a random-force target for trials of matched absolute error sizes (● in *A*).

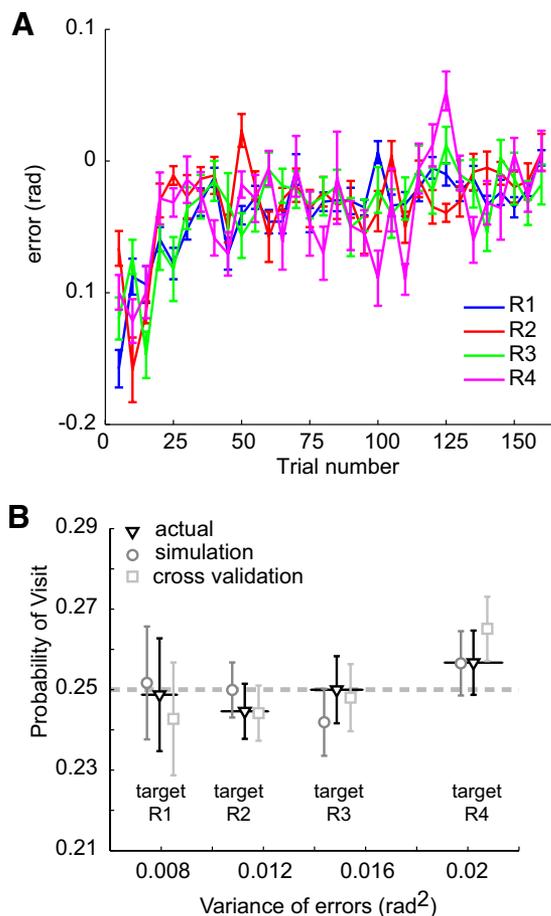


FIG. 5. *A*: average movement error separated by the variance of the perturbation (target R1 = smallest, R4 = highest variance). The vertical bars indicate the SE across participants. *B*: overall probability of selecting a target direction with small (R1), medium (R2), high (R3), and very high (R4) perturbation variance. Plotted are the actual probabilities, simulated probabilities using parameters obtained from logistic regression of *experiment 2* and simulated probabilities using parameters from *experiment 1* for cross-validation. ---, the predicted probability for an unbiased learner.

Finally, it is possible that participants repeated even high-variance targets because they attempted to reduce their motor variance strategically through an increase in stiffness for these movement directions (Burdet et al. 2001). To test for this possibility, we estimated the stiffness of the arm for each movement direction and variance level. While the stiffness varied systematically with the movement direction, reflecting the natural anisotropy of the arm (Mussa-Ivaldi et al. 1985), our estimates were not influenced by the variance of the target, ($F(3,58) = 1.80, P = 0.157$).

In summary, the results of *experiment 2* demonstrated that participant's choices were influenced by the absolute size of the error of the last movement but not by the variance of these errors, at least as measured over 160 trials.

DISCUSSION

The presented study, to our knowledge, is the first to investigate active learning strategies in human motor control. We used a task that had multiple components (movement directions). The participants' goal was to choose their own training schedule so that they would become proficient in all compo-

nents of the task. We found that the choices made by the learners were dominated by two main factors.

First, when subjects encountered a task component that resulted in large performance errors, they repeated that movement. It is intuitively clear that this strategy should lead to better learning as compared with random selection of task components: big errors can indicate a mismatch between the current estimate of the force perturbation and the correct value and therefore indicate the need to learn. Participants who sought out movement directions with large errors were more successful in subsequent test trials; participants who avoided errors were comparatively less successful.

Our second finding was that after performing a perfect movement (i.e., no errors), participants did not avoid that task component. Current algorithms in machine learning show the opposite tendency: making an observation close to an action component reduces the model uncertainty in the neighborhood of this observation and therefore reduces the probability of re-selecting this component in the next training trial. People did not follow this strategy during active learning. When no error was observed in a movement direction—a situation in which the estimated uncertainty of the output should have been reduced, they did not avoid this movement direction. This behavior was suboptimal as it was correlated with poorer test performance after active learning. Thus reducing the tendency to repeat well-executed task components may help people improve overall task performance.

Why do people repeat task components even when the last error was very small? We tested the hypothesis that this might reflect a strategy to test the consistency or variance of the task component. Such knowledge could then be used to avoid task components in which large errors arise from high variance of the environment rather than from a large mismatch between average required and average produced motor behavior. We found that while participant's choices were dependent on the absolute size of the last error, they were insensitive to the cross-trial variance of these errors. The results can imply one of two things. First, participants may not estimate the variance of the error signal over multiple trials. This is congruent with recent results that showed that variance of reward values does not influence decisions (Daw et al. 2006). Alternatively, it may imply that participants were trying to reduce the variance of the errors for movement directions with high trial-to-trial variance through a strategic increase in stiffness (Burdet et al. 2001). While this remains a possibility, our analysis suggests that they were not successful in doing so. As a result, we concluded that the strategy of re-selecting the target was suboptimal. Indeed for most participants who showed this behavior performance after active learning was slightly worse than after passive learning in which trials were picked at random.

Thus participants repeated already learned skills rather than explore new, untrained task components. Indeed, performance during training trials was better during active learning than during passive learning likely due to the larger number of target repetitions during active learning. While this strategy led to poorer performance in the short term, it may have increased the motivation during the task. Recent studies indicate that positive, motivating feedback may increase retention of learned motor skills in the long-term (Chiviacowsky and Wulf 2007). The optimality of the active machine-learning algorithm only reflects the minimization of cost terms asso-

ciated with the explicit task goal. Therefore it is possible that repeating a task stems from its benefits over a long-term period, a component that we did not assay in our protocol. In addition, it is possible that as the targeted skill involve more variables, other principles may determine the optimal active learning strategy (Wulf and Shea 2002). Indeed there is evidence that the sequence of learning examples affects retention properties of the acquired skill. In a task where participants were asked to learn three different punch styles, people who trained with a random schedule—practice trials on all three styles were conducted intermittently—retained their performance better after 10 min and after 10 days, when compared with people who trained one style at a time (Shea and Morgan 1979). Similar results emphasizing the benefits of concurrent and intermixed training of several subskills as a whole were found in basketball shooting (Memmert 2006), pistol shooting (Keller et al. 2006), surgery training (Brydges et al. 2007), and three-dimensional spatial orienting (Shebilske et al. 2006). For example, in the basketball study, it was found that people had better acquisition when shooting positions were blocked but better retention of the skills when shooting positions were randomized.

These results raise the possibility that choices made during training can have different effects on short-term versus long-term measures of performance. Based on our study we can only make inferences about the short-term effects of these choices. However, because most evidence suggest an improvement of long-term retention with intermixing of training examples, we think it is likely that these results would generalize to a longer time scale. Our results highlight that humans do not always choose the optimal learning strategy when given the chance to select their own training sequence, possibly preferring immediate positive feedback to the chance of exploring new, unlearned task components. We showed that we can separate aspects of learning strategy that improved overall performance from aspects that impaired performance. Our findings imply that it should be possible to design adaptive algorithms, i.e., an artificial coach, that would lead to better short-term gains than random training and, in particular, better than what the students are likely to do on their own. Specifically, the present results predict that an artificial coach could be designed that produces better performance simply by instructing the student to repeat task components only when the last error in that component was large. Such adaptive training algorithms may play a useful role in sports training, as well as robot-based rehabilitation training after stroke or developmental disabilities.

APPENDIX 1

Let us assume a task consists of P different skill components or behaviors ($x = 1 \dots P$), and that the learner's current skill level can be expressed by a set of corresponding parameters $w^{(1)}, \dots, w^{(P)}$. For example, in the force adaptation task, $-w^{(1)}, \dots, -w^{(P)}$ represent the estimated magnitude of the force perturbation and $w^{(1)}, \dots, w^{(P)}$ represent the estimated magnitude that the subject should produce to counteract the force perturbation (i.e., the internal model of the task). On each trial, the produced output \hat{y}_n for a particular component x_n depends on the corresponding parameter plus some motor noise ρ_n , a random variable with zero-mean and variance σ^2

$$\hat{y}_n = w_n^{(x_n)} + \varepsilon_n \quad (A1)$$

After an action on trial n , the system learns from the performance error, the difference between the actual \hat{y}_n and optimal output y_n . Thus on each trial

$$w_{n+1}^{(x_n)} = w_n^{(x_n)} + K_n(y_n - \hat{y}_n) \quad (A2)$$

After the criterion proposed by Cohn et al. (1996), the best skill component to train on trial n is x_n^* , the component that, after learning, will reduce the expected squared error on trial $n + 1$ (the expected value is taken over all possible components x_{n+1} and produced movements)

$$x_n^* = \arg \min_{x_n} \left(E_{\hat{y}_{n+1}} \left[(y_{n+1} - \hat{y}_{n+1})^2 \middle| x_n \right] \right) \quad (A3)$$

Under the assumption of an unbiased learner (i.e., a learner that does on average not show a systematic error), the expected squared error of the output is the uncertainty about the relevant parameter $\Sigma_{n+1}^{(x)}$ plus the variance of producing the output (σ^2), again averaged across all possible skill components on trial $n + 1$

$$E_{\hat{y}_{n+1}} \left[(y_{n+1} - \hat{y}_{n+1})^2 \right] = E_{x_{n+1}} \left(\Sigma_{n+1}^{(x_{n+1})} + \sigma^2 \right) \quad (A4)$$

The uncertainty is defined as the expected squared distance from the unknown ideal model parameters $w^{(x)*}$

$$\Sigma_{n+1}^{(x)} = E \left[\left(w^{(x)*} - w_{n+1}^{(x)} \right)^2 \right] \quad (A5)$$

Now we have to calculate how observing the error of trial n for skill component x_n influences the uncertainty of the model parameter on the next trial after using the learning rule in Eq. A2. To do so, we can use Eq. A2 to expand the term $(w^{(x)*} - w_{n+1}^{(x)})$

$$\begin{aligned} w_{n+1}^{(x)} &= w_n^{(x)} + K_n(y_n - \hat{y}_n) \\ w_{n+1}^{(x)} &= w_n^{(x)} + K_n \left(w^{(x)*} - w_n^{(x)} - \varepsilon_n \right) \\ w^{(x)*} - w_{n+1}^{(x)} &= w^{(x)*} - w_n^{(x)} - K_n w^{(x)*} + K_n \varepsilon_n + K_n w_n^{(x)} \\ &= (1 - K_n) \left(w^{(x)*} - w_n^{(n+1)} \right) - K_n \varepsilon_n \end{aligned} \quad (A6)$$

We assume that the motor errors ε_n have variance σ^2 and are independent of the parameter uncertainty. Thus we can express the uncertainty around $w_{n+1}^{(x)}$ after perceiving behavior $x^{(n)}$ as

$$\begin{aligned} \Sigma_{n+1}^{(x)} &= E \left[\left(w^{(x)*} - w_{n+1}^{(x)} \middle| x_n \right)^2 \right] \\ &= E \left[\left((1 - K_n) \left(w^{(x)*} - w_n^{(x)} \right) - K_n \varepsilon_n \right)^2 \right] \\ &= (1 - K_n)^2 \Sigma_n^{(x)} + K_n^2 \sigma^2 \end{aligned} \quad (A7)$$

From this we can see that the change in uncertainty after learning for a particular skill component x is

$$\Sigma_{n+1}^{(x)} - \Sigma_n^{(x)} = (K_n - K_n^2) \Sigma_n^{(x)} + K_n^2 \sigma^2 \quad (A8)$$

Thus for a constant learning rate $0 < K < 1$, it follows directly from Eq. A8 that the average uncertainty will be reduced most, if we pick a behavior x_n^* , for which the corresponding parameter uncertainty $\Sigma_x^{(n)}$ is highest.

Thus the decision rule in Eq. A3 can be simplified as

$$\begin{aligned} x_n^* &= \arg \min_{x_n} E \left[(y_{n+1} - \hat{y}_{n+1})^2 \middle| x_n \right] \\ &= \arg \min_{x_n} E \left[\Sigma_{n+1}^{(x_n)} + \sigma^2 \middle| x_n \right] \\ &= \arg \max_{x_n} \Sigma_n^{(x_n)} \end{aligned} \quad (A9)$$

We also considered the result for adaptive learning rates. The optimal K_n , known as the Kalman gain, is the learning rate that results in the lowest possible uncertainty in the parameter after learning. To obtain this learning rate, we differentiate Eq. A7 with respect to K_n . The resulting optimal adaptive learning rate depends on the parameter uncertainty $\Sigma^{(x)}$ and the motor noise σ^2

$$K_n = \frac{\Sigma_n^{(x)}}{\Sigma_n^{(x)} + \sigma^2} \quad (A10)$$

With such flexible learning rate the updated optimal parameter uncertainty becomes

$$\Sigma_{n+1}^{(x)} = \left(\Sigma_n^{(x)} - \frac{\Sigma_n^{(x)} \Sigma_n^{(x)}}{\Sigma_n^{(x)} + \sigma^2} \right) \quad (A11)$$

The decrease in uncertainty between trial n and $n + 1$ therefore is maximal when $\Sigma_n^{(x)}$ is maximal. Therefore the optimal selection rule (Eq. A9) remains valid. Indeed this can be shown for a number of different choices of K_n . It should be noted that we assume that there is negligible generalization between behaviors, that the motor noise σ^2 is constant across all components, and that the amount of motor noise cannot be changed by learning.

APPENDIX 2

Does the optimal active learning rule (Eq. A9) remain valid for a system or learner that forgets with time? One might suspect the opposite: if one forgets, it would be good to repeat immediately what was learned before. Here we show that the derivation in APPENDIX 1 remains valid as long as we have an unbiased learner: a learner that has a rate of forgetting that is matched to the rate of change in the environment (Kording et al. 2007).

To set out, let us assume that the environment (v) changes following a simple auto-regressive process of order 1 with $0 < A_T < 1$

$$v_{n+1}^{(x)} = A_T v_n^{(x)} + \varepsilon_n^{(x)}; \quad \varepsilon \sim N(0; Q_T) \quad (A12)$$

An optimal Bayesian learner should then mirror the rate of change in the environment with a forgetting factor of the same size. Thus the learning rule (Eq. A2) becomes

$$w_{n+1}^{(x)} = \begin{cases} A_L w_n^{(x)} + K_n (y_n - \hat{y}_n); & \text{if observation on } x \text{ was made} \\ A_L w_n^{(x)}; & \text{otherwise} \end{cases} \quad (A13)$$

Thus in the absence of observations and $-1 < A < 1$, all weights drift back toward zero.

The uncertainty would also be updated to match the uncertainty in the environment

$$\Sigma_{n+1}^{(x)} = \begin{cases} A_L^2 \Sigma_n^{(x)} - K_n \Sigma_n^{(x)} + Q_L; & \text{if } x \text{ observed} \\ A_L^2 \Sigma_n^{(x)} + Q_L; & \text{otherwise} \end{cases} \quad (A14)$$

As long as the learner is unbiased, i.e., $A_T = A_L$ and $Q_T = Q_L$, the expected squared error will still be as in Eq. A4. Thus the only thing that changed from APPENDIX 1 is that we have added a constant Q

to the uncertainty and scaled the uncertainty by A^2 on every trial. Neither of these manipulation changes where the minimum for the choice in Eq. A9 lies. Thus as long as we have an unbiased learner, the selection rule in Eq. A9 remains optimal.

What if the forgetting rate of the learner (A_L) and the “true” forgetting rate of the environment (A_T) are not the same? Then we will see a discrepancy between $A_L w_n$ and $A_T w_n$, and this difference will be bigger the further w_n is away from zero (the prior). So if learner has a weight that has a large absolute value, then the forgetting will make the internal estimate systematically closer to zero than in the environment. The optimal strategy would then to repeat these movements or observations more to offset the faster forgetting rate with repeated training.

The argument here does not rest on the assumption that the forgetting rate of the learner and the forgetting rate of a specific experimental environment are matched. Rather we propose that the forgetting rate is matched to the average forgetting rate in the environment and that under these conditions the active learning rule (Eq. A9) is optimal.

GRANTS

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

REFERENCES

- Brydges R, Carnahan H, Backstein D, Dubrowski A.** Application of motor learning principles to complex surgical tasks: searching for the optimal practice schedule. *J Mot Behav* 39: 40–48, 2007.
- Burdet E, Osu R, Franklin DW, Milner TE, Kawato M.** The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature* 414: 446–449, 2001.
- Chiviacowsky S, Wulf G.** Feedback after good trials enhances learning. *Res Q Exerc Sport* 78: 40–47, 2007.
- Cohn DA, Ghahramani Z, Jordan MI.** Active learning with statistical models. *J Art Intell Res* 4: 129–145, 1996.
- Cohen JD, McClure SM, Yu AJ.** Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond B Biol Sci* 362: 933–942, 2007.
- Daw ND, O’Doherty JP, Dayan P, Seymour B, Dolan RJ.** Cortical substrates for exploratory decisions in humans. *Nature* 441: 876–879, 2006.
- 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.
- Einhäuser W, Mundhenk TN, Baldi P, Koch C, Itti L.** A bottom-up model of spatial attention predicts human error patterns in rapid scene recognition. *J Vis* 7: 6 1–13, 2007.
- Huang VS, Shadmehr R.** Evolution of motor memory during the seconds after observation of motor error. *J Neurophysiol* 97: 3976–3985, 2007.
- 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.
- Keller GJ, Li Y, Weiss LW, Relyea GE.** Contextual interference effect on acquisition and retention of pistol-shooting skills. *Percept Mot Skills* 103: 241–252, 2006.
- Kording KP, Tenenbaum JB, Shadmehr R.** The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat Neurosci* 10: 779–786, 2007.
- Kording KP, Wolpert DM.** Bayesian integration in sensorimotor learning. *Nature* 427: 244–247, 2004.
- Memmert D.** Long-term effects of type of practice on the learning and transfer of a complex motor skill. *Percept Mot Skills* 103: 912–916, 2006.
- Mussa-Ivaldi FA, Hogan N, Bizzi E.** Neural, mechanical, and geometric factors subserving arm posture in humans. *J Neurosci* 5: 2732–2743, 1985.
- Shea JB, Morgan RL.** Contextual interference effects on the acquisition, retention and transfer of a motor skill. *J Exp Psychol (Hum Learn)* 3: 179–187, 1979.
- Shebilske WL, Tubre T, Tubre AH, Oman CM, Richards JT.** Three-dimensional spatial skill training in a simulated space station: random vs. blocked designs. *Aviat Space Environ Med* 77: 404–409, 2006.
- Thoroughman KA, Shadmehr R.** Learning of action through adaptive combination of motor primitives. *Nature* 407: 742–747, 2000.
- Wulf G, Shea CH.** Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon Bull Rev* 9: 185–211, 2002.