



Published as: *Neural Comput.* 2012 April ; 24(4): 939–966.

Generalization and Multirate Models of Motor Adaptation

Hirokazu Tanaka,

Howard Hughes Medical Institute, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A

John W. Krakauer, and

Departments of Neurology and Neuroscience, Johns Hopkins University, Baltimore, MD 21287, U.S.A

Terrence J. Sejnowski

Howard Hughes Medical Institute, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A., and Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, U.S.A

Hirokazu Tanaka: hirokazu@salk.edu; John W. Krakauer: john.krakauer@gmail.com; Terrence J. Sejnowski: terry@salk.edu

Abstract

When subjects adapt their reaching movements in the setting of a systematic force or visual perturbation, generalization of adaptation can be assessed psychophysically in two ways: by testing untrained locations in the work space at the end of adaptation (slow postadaptation generalization) or by determining the influence of an error on the next trial during adaptation (fast trial-by-trial generalization). These two measures of generalization have been widely used in psychophysical studies, but the reason that they might differ has not been addressed explicitly. Our goal was to develop a computational framework for determining when a two-state model is justified by the data and to explore the implications of these two types of generalization for neural representations of movements. We first investigated, for single-target learning, how well standard statistical model selection procedures can discriminate two-process models from single-process models when learning and retention coefficients were systematically varied. We then built a two-state model for multitarget learning and showed that if an adaptation process is indeed two-rate, then the postadaptation generalization approach primarily probes the slow process, whereas the trial-by-trial generalization approach is most informative about the fast process. The fast process, due to its strong sensitivity to trial error, contributes predominantly to trial-by-trial generalization, whereas the strong retention of the slow system contributes predominantly to postadaptation generalization. Thus, when adaptation can be shown to be two-rate, the two measures of generalization may probe different brain representations of movement direction.

1 Introduction

Generalization of motor adaptation has been assessed in two different ways. One approach, postadaptation generalization, is to probe how a fully learned local remapping generalizes to other unlearned locations or directions (Imamizu, Uno, & Kawato, 1995; Gandolfo, Mussa-Ivaldi, & Bizzi, 1996; Krakauer, Pine Ghilardi, & Ghez, 2000; Mattar & Ostry, 2007). Another way is to assess adaptation by examining how errors experienced on trial k transfer to another part of the work space in trial $k + 1$ using a single-state state-space model (trial-by-trial generalization) (Thoroughman & Shadmehr, 2000; Baddeley, Ingram, Miall, 2003; Donchin, Francis, & Shadmehr, 2003; Cheng & Sabes, 2007; Francis, 2008). In both cases, the goal is to assess how a learned remapping generalizes to untrained areas of the work space and thereby determine the nature of the representation used by the brain to encode the learned remapping (Shadmehr, 2004).

This study focuses on directional generalization for multitarget adaptation. We have shown that in the case of eight-target visuomotor rotation adaptation, the two generalization approaches yield similar narrow generalization patterns that do not extend beyond adjacent targets separated by 45 degrees (Tanaka, Sejnowski, Krakauer, 2009). We demonstrated that a population-coding model composed of narrowly tuned computational units successfully explained both forms of generalization and several additional experimental observations, such as the wider pattern of generalization and the slower adaptation speed with an increasing number of training targets. In contrast with these results for rotation adaptation, for force-field adaptation, the two generalization approaches yielded different generalization functions, with trial-by-trial adaptation affecting directions as far as 180 degrees (Thoroughman & Shadmehr, 2000; Donchin et al., 2003), whereas postadaptation generalization was limited to directions within 90 degrees of the training targets (Mattar & Ostry, 2007).

How might these differences in generalization for rotations and viscous force fields be explained? Recently an innovative two-state model of short-term motor adaptation was proposed that posited a fast process with a fast learning rate but weak retention and a slow process with a slow learning rate but strong retention (Smith, Ghazizadeh, & Shadmehr, 2006). A subsequent study demonstrated that it is the slow processes that is retained as a motor memory (Joiner & Smith, 2008). This shows that when initial adaptation is made up of more than one process, these processes may go on to show divergent behavior on subsequent probing and may therefore be separable. In addition, the two-state model, or its multiple-time scale extension, was successfully applied to saccadic gain adaptation (Körding, Tenenbaum, & Shadmehr, 2007; Ethier, Zee, & Shadmehr, 2008). This suggests that an explanation for why rotation adaptation yielded a single generalization function for both approaches whereas force-field adaptation yielded two different generalization functions is that rotation adaptation is one-rate and force-field adaptation is two-rate.

This study had two goals. The first was to investigate the reliability of parameter estimation techniques to determine whether a process is indeed multirate instead of single rate. The second was to show analytically and by computational simulation that if adaptation is indeed two-rate, then the postadaptation approach probes generalization of the slow process, whereas the trial-by-trial approach evaluated with the single-state model probes generalization of the fast process.

2 Statistical Tests for Determination of a State-Space Model's Dimensionality and Parameters

In a recent study, a computational model with fast and slow timescales successfully explained savings, anterograde interference, spontaneous recovery, and rapid unlearning to adaptation in a viscous force field (Smith et al., 2006). One critical question when applying multirate models to experimental data is the degree to which the model's parameters need to differ in order for a standard statistical test to be able to detect multiple processes. We investigated this issue by fitting state-space models to artificially generated error data and performing statistical tests. We first simulated a two-state model and then determined how well standard statistical tests can recover the state-space model's parameters.

A two-state model with a state vector $\mathbf{x} = (x^{(f)} \ x^{(s)})^T$ includes a process equation and an observation equation defined as

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\Delta y_k + \mathbf{w}_k \\ y_k = \mathbf{C}\mathbf{x}_k + v_k \end{cases}, \quad (2.1)$$

where the matrices are

$$\mathbf{A} = \begin{pmatrix} a^{(f)} & 0 \\ 0 & a^{(s)} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b^{(f)} \\ b^{(s)} \end{pmatrix}, \quad \mathbf{C} = (1 \quad 1). \quad (2.2)$$

Here k is a trial number. We focus on trial-based (i.e., discrete time) state-space models with a fixed intertrial interval and briefly summarize the relation between discrete- and continuous-time state-space models in appendix A.

For simplicity, we assumed that motor adaptation obeyed the Markov property: the next state depends on only the current state and the error experienced at the current trial. The feedback error $\Delta y_k = u_k - y_k$ drives the adaptation process when an external perturbation (u_k) such as a visuomotor rotation or force-field perturbation is imposed. Here the two processes were assumed to develop independently (i.e., no off-diagonal components in \mathbf{A} ; for a discussion on the roles of off-diagonal components, see Criscimagna-Hemminger & Shadmehr, 2008) and to contribute to output equally. We will refer to a 's and b 's in equation 2.2 as retention and learning coefficients, respectively. The process and measurement noises are assumed to be gaussian as

$$\mathbf{w}_k \sim N(0, \mathbf{Q}) \quad \text{and} \quad v_k \sim N(0, R). \quad (2.3)$$

Given values in the process matrices (\mathbf{A} and \mathbf{B}) and the noise covariance matrices (\mathbf{Q} and R), it is possible to simulate artificial error data $\{\Delta y_k\}$.

When a constant perturbation, $u_k = \bar{u}$, is imposed, the average behavior of two-state model can be solved analytically. We used arbitrary units for the perturbations, so we set $\bar{u} = 1$ for the following discussion. We set the noise terms to zero in equation 2.1, and the fast and slow processes can develop as

$$\begin{cases} x_{k+1}^{(f)} = a^{(f)} x_k^{(f)} + b^{(f)} \Delta y_k \\ x_{k+1}^{(s)} = a^{(s)} x_k^{(s)} + b^{(s)} \Delta y_k \end{cases}, \quad (2.4)$$

$$\Delta y_k = \bar{u} - x_k^{(f)} - x_k^{(s)}, \quad (2.5)$$

where all variables are scalar. From these equations, the output at the k th trial is

$$y_k \approx [d_1 e^{-\lambda_+ k} + d_2 e^{-\lambda_- k} + d_3] \bar{u}, \quad (2.6)$$

where d_i are constants. Assuming the retention factors and learning gains are small, the decay constants λ_{\pm} are

$$\lambda_{\pm} = \frac{1}{2} [(1 - a^{(f)} + b^{(f)}) + (1 - a^{(s)} + b^{(s)})] \pm \sqrt{[(1 - a^{(f)} + b^{(f)}) - (1 - a^{(s)} + b^{(s)})]^2 + 4b^{(f)}b^{(s)}}. \quad (2.7)$$

Thus, a double-exponential time course of adaptation can occur with a two-state model but not with a single-state model. Therefore, if experimental data exhibit two timescales, a single-state model is incapable of modeling the data and a two-state model is justified. The same result also applies to multitarget training: a single-state model has a single timescale and hence cannot explain a learning curve with more than one time constant.

2.1 Maximum Likelihood Method

When there is more than one possible model for describing experimental data, they can be compared using a statistical test that provides a measure of the goodness of fit of a model. We tested how two standard methods of system identification estimated the model's dimensionality: maximum likelihood (ML) estimation and the prediction error (PE) estimation. The ML estimation finds a parameter value set θ that maximizes log likelihood,

$$\begin{aligned}\log L[\theta|y_{1:N}] &= \log p(y_{1:N}|\theta) \\ &= \log p(y_1|\theta) + \sum_{n=2}^N \log p(y_n|y_{1:n-1}, \theta) \\ &= -\frac{Nd_y}{2} \log(2\pi) - \frac{1}{2} \log \det(\mathbf{C} \sum_1 \mathbf{C}^T + \mathbf{R}) \\ &\quad - \frac{1}{2} \sum_{n=2}^N \{ \log \det \sum_{n|n-1}^y + (y - \hat{y}_{n|n-1})^T (\sum_{n|n-1}^y)^{-1} (y - \hat{y}_{n|n-1}) \},\end{aligned}\quad (2.8)$$

where the estimated value of y_n given $\{y_1, \dots, y_{n-1}\}$,

$$\hat{y}_{n|n-1} = E[y_n | \{y_1, \dots, y_{n-1}\}, \theta], \quad (2.9)$$

and its covariance,

$$\sum_{n|n-1}^y = \text{cov}[y_n | \{y_1, \dots, y_{n-1}\}, \theta] \quad (2.10)$$

are obtained using a standard Kalman filter forward computation (Kalman, 1960). Here N is the number of experimental trials in a single run. In the two-state model, we parameterized the covariance matrices as

$$\mathbf{Q} = \sigma_w^2 \mathbf{I}_2, \quad \mathbf{R} = \sigma_v^2 \quad \text{and} \quad \text{cov}[\mathbf{x}_1] = \sigma_1^2 \mathbf{I}_2, \quad (2.11)$$

and a nine-dimensional parameter vector,

$$\theta_{2\text{state}}^{\text{ML}} = (a^{(f)}, a^{(s)}, b^{(f)}, b^{(s)}, \sigma_w, \sigma_v, x_1^{(f)}, x_1^{(s)}, \sigma_1), \quad (2.12)$$

was optimized to maximize the log likelihood. In the single-state model defined as

$$\begin{cases} x_{k+1} = ax_k + b\Delta y_k + w_k \\ y_k = Cx_k + v_k \end{cases}, \quad (2.13)$$

a six-dimensional parameter vector,

$$\theta_{1\text{state}}^{\text{ML}} = (a, b, \sigma_w, \sigma_v, x_1, \sigma_1), \quad (2.14)$$

was optimized. These parameters can be optimized by either numerical maximization (Gupta & Mehra, 1974) or expectation-maximization (EM) algorithms (Shumway & Stoffer, 1982, 2000). We used a numerical maximization method (the Nelder-Mead simplex method) rather than EM algorithms because we found that the latter were slower and less robust in convergence. For a statistical comparison of multiple models, we used the Akaike information criterion (Akaike, 1974):

$$\text{AIC} = 2d - 2\log L[\hat{\theta}_{\text{ML}} | y_{1:N}] \quad (2.15)$$

Here d is the number of estimated parameters: $d=9$ for the two-state model and $d=6$ for the single-state model. We optimized both parameter vectors $\theta_{1\text{state}}^{\text{ML}}$ and $\theta_{2\text{state}}^{\text{ML}}$ that maximize the log likelihood, computed the AICs, and determined which state-space model was selected by the AIC criterion.

2.2 Prediction Error Estimation Method

Given an initial estimate of the system matrices, artificial error data were generated using the two-state model, equation 2.1, or the single-state model, equation 2.13. In the PE estimation method, the parameter vector was obtained by minimizing a squared difference between data and estimated error (i.e., prediction error; Goodwin & Sin, 1984):

$$V = \sum_{k=1}^N (\Delta y_k - \Delta \hat{y}_k)^2. \quad (2.16)$$

The state-space model parameter vector to be optimized is a three-dimensional vector

$\theta_{1\text{state}}^{\text{PEM}} = (a, b, x_1)$ for the single-state model or a six-dimensional vector and

$\theta_{2\text{state}}^{\text{PEM}} = (a^{(f)}, a^{(s)}, b^{(f)}, b^{(s)}, x_1^{(f)}, x_1^{(s)})$ for the two-state model. As a model selection criterion, we used a future prediction error (FPE) (Akaike, 1969; Ljung & Ljung, 1987), defined as

$$\text{FPE} = \frac{1+d/N}{1-d/N} V. \quad (2.17)$$

We used the same parameterization of the state-space models as defined above for ML estimation. As in the ML method, we optimized the parameter vectors, computed the FPEs, and selected the state-space model by the FPE criterion.

2.3 Numerical Results

2.3.1 Dimensionality Estimation—We created surrogate error data using the two-state state-space and estimated the dimensionality using the two statistical model selection methods described. Several factors affect statistical tests of a state-space model's dimensionality and parameter estimation: noise strength, length of artificial error data, and state-space model parameters. First, if noise strength is substantial, multiple time constants in error data may be misinterpreted as a single time constant process. Second, values of the retention and learning coefficients for the fast and slow processes in the two-state model need to differ enough to be distinguishable from the single-state model. Finally, the number of error trials, N , will determine the balance between the goodness of fit and the number of

parameters. To simplify our discussion, we focused on the influence of noise strength and state-space model parameter values by fixing the number of simulated trials to 100.

To generate the error data, the retention coefficient for the slow process and the learning coefficient for the fast process were fixed ($a^{(s)} = 0.99$ and $b^{(f)} = 0.20$), and the ratios,

$$r^{(a)} = \frac{a^{(f)}}{a^{(s)}}, \quad r^{(b)} = \frac{b^{(s)}}{b^{(f)}}, \quad (2.18)$$

were systematically varied from 0.10 to 0.97 in steps of 0.03, because we were interested in the difference between the coefficients of fast and slow processes. The same qualitative results were obtained (not shown) using four different sets of anchoring values for $a^{(s)}$ and $b^{(f)}$ [(0.96, 0.20), (0.93, 0.20), (0.99, 0.15) and (0.99, 0.20)]. We performed a Monte Carlo simulation for the estimation of dimensionality. First, 100 trials of artificial error data were generated by simulating the two-state state-space model, equation 2.1, and then the optimal dimensionality was determined (i.e., either the single- or two-state model was chosen) using the statistical tests (ML or PE). This procedure was repeated five times to obtain an average of estimated dimensionality. We considered three noise magnitudes: (1) $\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$, (2) $\sigma_w^2 = 5 \times 10^{-4}$ and $\sigma_v^2 = 15 \times 10^{-4}$, and (3) $\sigma_w^2 = 10 \times 10^{-4}$ and $\sigma_v^2 = 30 \times 10^{-4}$. Typical realizations of artificial error data are shown in Figure 1. Note that although the average error curves exhibited two time-constant behaviors (initial rapid decrease and subsequent gradual decrease in error), the noisier the learning curves become, the more difficult it is to identify multiple time constants from the data. The noise used in the first condition ($\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$) was roughly of the same order of magnitude as in our previously published result for visuomotor rotation adaptation (Zarahn, Weston, Liang, Mazzoni, & Krakauer, 2008) ($\sigma_w^2 = 2.530 \text{ (deg}^2\text{)}$ and $\sigma_v^2 = 12.791 \text{ (deg}^2\text{)}$ or $\sigma_w^2 = 1.95 \times 10^{-5}$ and $\sigma_v^2 = 9.87 \times 10^{-5}$ in dimensionless unit). Additional simulations confirmed that instead of fixing the ratio of the noise magnitudes, increasing the noise magnitude of either process or measurement noise independently yielded qualitatively similar dimensional estimates.

Figures 2A to 2C show the dimensionality estimated by the ML method and Figures 2D to 2F the PE method. Three general conclusions follow from these simulation results. First, as the noise strengths increased, multiple timescales were lost, and a single-state model with a single time constant was favored, as we expected. Second, retention coefficients (a 's in equation 2.2) needed to differ in order for multiple processes to be identified, whereas the difference between the learning coefficients (b 's in equation 2.2) had little influence on the dimensional estimation. Finally, the ML and PE methods yielded qualitatively similar estimation methods, although the results of the PE method appeared more robust in the setting of noise compared to the ML method.

2.3.2 Parameter Estimation—We then asked how well the two methods could reproduce the state-space parameters. This time we fixed the state-space parameters ($a^{(s)} = 0.99$, $b^{(f)} = 0.20$, $r^{(a)} = r^{(b)} = 0.20$) to generate artificial data and computed the estimated ratios $\hat{r}^{(a)} = \hat{a}^{(f)}/\hat{a}^{(s)}$, $\hat{r}^{(b)} = \hat{b}^{(s)}/\hat{b}^{(f)}$. This procedure was iterated 1000 times to obtain estimated distributions of ratios. We used two different noise levels; (1) $\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$ and (2) $\sigma_w^2 = 5 \times 10^{-4}$ and $\sigma_v^2 = 15 \times 10^{-4}$.

Figure 3 summarizes the parameter values obtained by the ML method (red histograms in Figures 3A and 3B) and by the PE method (blue histograms in panels A and B). The two estimation methods gave qualitatively similar results, although the ML provided tighter

distributions. The estimated distributions of the retention coefficient ratio were almost symmetrical and centered at the true value, whereas the estimated distributions of the learning coefficient ratio were skewed and peaked at the true value. The estimation of the learning coefficients was found to be more precise than that of the retention coefficients in both estimation methods. When the noise strengths were increased (see Figures 3C and 3D), the estimated distributions became broader, but essentially the same trend was observed. In summary, our simulations showed that the ML and PE estimation methods provided similar results for both the dimensionality and the parameter values of state-space models.

2.4 Robustness of the Estimation Methods

We here examine the robustness of the estimation methods in simulating motor adaptation processes when some simplifying assumptions are violated: (1) process and measurement noises are gaussian, (2) the learning function in response to movement error is linear, and (3) perturbations are constant. Although these assumptions are often made in system identification, they are not strictly valid, so there is a risk of applying an inappropriate state-space model and producing biased estimates in parameter estimation.

We therefore relaxed these assumptions in simulations of the two-state state-space model (see equation 2.1), and tested how well the parameters were estimated. Both the PE and ML methods yielded similar results in these tests, but only the results of ML method will be shown. The covariances of process and measurement noises were fixed at $\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$, respectively, and the state-space model parameters were $a^{(s)} = 0.99$, $b^{(f)} = 0.20$, and $r^{(a)} = r^{(b)} = 0.20$.

We first investigated the effects of having nongaussian processes and measurement noise. We used a Pearson distribution parameterized by the four first cumulants (mean, variance, skewness, and kurtosis). The skewness and kurtosis were systematically varied for both the processes and measurement noise. Highly nongaussian distributions such as bimodal distributions were not considered. The ML estimation method was applied to artificially generated noise data for two levels of skewness, -0.5 and 0.5 (shown in Figure 4A). The results in Figures 4B and 4C confirmed that the parameter estimation was little affected by these moderately skewed distributions. Next, simulated artificial error data were generated (see Figure 4D) at four levels of kurtosis (2.0 , 2.5 , 3.5 , and 4.0). The ML parameter estimation was minimally affected by varying the kurtosis over a wide range of subgaussian (kurtosis smaller than 3) and supergaussian (kurtosis larger than 3) distributions (see Figures 4E and 4F).

The influence of using a nonlinear learning function was investigated next. Fine and Thoroughman (2006) reported that perceived error during motor learning was linear for small errors but was saturated at an upper bound when the error was relatively large. To model this nonlinear effect, the linear error term (Δy_k) was replaced with a function that saturated for large inputs, $s \cdot \tanh(\Delta y_k/s)$, where s is a scale parameter controlling the degree of nonlinearity:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B} \cdot s \cdot \tanh\left(\frac{\Delta y_k}{s}\right) + \mathbf{w}_k \\ y_k = \mathbf{C}\mathbf{x}_k + v_k \end{cases} \quad (2.19)$$

When s is small, the learning function is less sensitive to large errors, resulting in slower learning curves, as illustrated in Figure 5A. Artificial error data were generated using this nonlinear state-space model and were then fitted by ML to a linear two-state state-space model with gaussian noise (see equation 2.1). Essentially the same results were obtained,

confirming the robustness of our estimation methods, although small estimation biases were observed, especially in estimating $\hat{r}^{(b)}$ (see Figures 5B and 5C). This occurred because the learning rate of the fast process was underestimated due to the saturation of nonlinear learning function.

Finally, a random perturbation (gaussian perturbations with a mean of 0 and standard deviation of 1) was used to generate artificial error data, to which the ML method was applied. As in the case of constant perturbations, the estimated values for the ratios of retention and learning coefficients were recovered (see Figures 5D and 5E). Estimation of the learning coefficients was more accurate (i.e., had smaller variance), but estimation of the retention coefficients was less accurate (i.e., larger variance) for random perturbation than for constant perturbation.

In summary, the estimation methods were robust even when some of the standard assumptions were violated, giving us more confidence that the fitting procedures can safely be applied to experimental data.

2.5 Reliability of Dimensional Estimation

To apply our estimation methods reliably to experimental data, we need to assess how often multiple processes are identified from data generated from a single-state model, and vice versa (false positives). We thus generated error data by simulating single- and two-state models with various numbers of trials and applied our dimensional estimation methods.

A range of trial numbers was tested from 60 to 240 in steps of 30 trials (the initial 30 trials were baseline, and the perturbation was turned on thereafter). Estimated dimensions were computed as functions of the number of trials. Dimensional and parameter estimation depends on factors other than the number of trials, such as values of retention and learning coefficients and noise variances. According to Figure 2, the retention coefficients (a) were more influential than the learning coefficients (b), so several values of the retention coefficients were examined while the learning coefficients were fixed. First, the ML estimation method was applied to error data generated with the single-state model (see equation 2.13). The value of the retention coefficient was set to 0.80, 0.90, or 0.95, and the learning coefficient was set to 0.2. Three levels of noise magnitudes were considered. For each number of trials, the ML estimation was applied to 50 independent realizations, and their mean value of estimated dimensions from each run was used as the estimated dimensionality. The results showed a gradually increasing trend of identifying the correct dimensionality (which was one in this case) with the increasing number of trials (see Figure 6). Also, our dimensional estimation method was quite robust against the noise. For noise magnitudes that fall in the range reported experimentally for visuomotor rotation (see Figure 6A), 100 trials should be sufficient to achieve the correct result with a probability of over 90%.

Next, the ML estimation method was applied to error data generated with the two-state model, equations 2.1 and 2.2. The retention coefficient of the slow process ($a^{(s)}$) was fixed to 0.99, and three values of the ratio $r = a^{(f)}/a^{(s)}$ were examined (0.80, 0.90, and 0.95). The learning coefficients $b^{(f)}$ and $b^{(s)}$ were set to 0.2 and 0.02, respectively. As in the previous simulation, three sets of noise variances were considered. The results showed that as in the case of the single-state model, the probability of correctly estimating the model dimensions was a gradually increasing function of the number of trials, and the difference between time constants of the two processes was relevant in recovering the correct dimensionality (see Figure 7). If the ratio of retention coefficients was 0.8, the estimation method could recover the correct dimensionality within 200 trials. When the number of trials is small or the noises

are significant, there is in general a bias toward a single-state model, so the AIC is biased in this case as we expected.

These results suggest a heuristic for estimating the dimensionality from experimental data: acquire error data for at least a few hundred trials and apply the dimensional estimation method to some of the data with an increasing number of trials. If the estimated dimensionality tends to increase with the number of trials, multiple timescales are indicated. If instead the estimated dimensionality decreases with an increasing number of trials, a single-state model is selected.

3 A Two-State State-Space Model for Multitarget Adaptation

The original formulation of two-state model (Smith et al., 2006) was designed to describe single-target adaptation and is not able to describe generalization to multiple targets. Here we extended their model to a general state-space model for directional errors to multiple targets with two states composed of fast and slow adaptation processes.

Consider a motor adaptation paradigm consisting of eight targets separated by 45 degrees and arrayed on a circle (0° , $\pm 45^\circ$, $\pm 90^\circ$, $\pm 135^\circ$, and 180°). The scalars of fast and slow processes ($x^{(f)}$ and $x^{(s)}$) in equation 2.1 are generalized to eight-dimensional vectors whose components represent the learning to eight targets. The two-state model consists of two eight-dimensional state vectors—a fast state ($\mathbf{X}^{(f)}$) representing the fast process and a slow state ($\mathbf{X}^{(s)}$) representing the slow process. Their trial-by-trial update is defined as

$$\begin{cases} \mathbf{X}_{k+1}^{(f)} = \mathbf{A}^{(f)} \mathbf{X}_k^{(f)} + \mathbf{B}^{(f)} \mathbf{H}_k^T \Delta y_k \\ \mathbf{X}_{k+1}^{(s)} = \mathbf{A}^{(s)} \mathbf{X}_k^{(s)} + \mathbf{B}^{(s)} \mathbf{H}_k^T \Delta y_k \end{cases} \quad (3.1)$$

Here the subscripts (k) denote trial number. The p th component of the state vectors $\mathbf{X}^{(f)}$ and $\mathbf{X}^{(s)}$ denotes the fast and slow processes' contribution to the hand movement direction when the target p ($p = 1, \dots, 8$) is shown. Each process in equation 3.1 is similar to single-state state-space models previously proposed but with minor differences: Thoroughman and Shadmehr (2000) used perturbation inputs rather than motor errors, whereas Donchin and colleagues (2003) defined their model without retention factors (i.e., \mathbf{A} was set to one). The 8×8 retention factor matrices, ($\mathbf{A}^{(f)}$, $\mathbf{A}^{(s)}$), are the part of the current states that will have decayed by the next state. For simplicity, these matrices are assumed to be identity matrices multiplied by scalars. The 8×8 generalization matrices, ($\mathbf{B}^{(f)}$, $\mathbf{B}^{(s)}$), are the part of the current error that generalizes to other target directions. The 1×8 observation matrix, \mathbf{H}_k , in equation 3.1 converts a scalar error into a vector. If a target is presented in the p th direction in the k th trial, the p th component of \mathbf{H}_k is set to one and the other components to zero. Those state vectors consist of eight components that correspond to hand movement directions for the eight targets, respectively. The diagonal components of the generalization matrices, ($\mathbf{B}^{(f)}$, $\mathbf{B}^{(s)}$), determine, separately for the two processes, how fast motor adaptation occurs for each target, and the off-diagonal components determine how much an error (Δy_k) experienced at one target at k th trial generalizes to other targets. In theory, generalization patterns from the two processes can take distinct shapes. The retention factors, ($\mathbf{A}^{(f)}$, $\mathbf{A}^{(s)}$), describe how the learned remapping decays. The fast process is assumed to learn and forget much more quickly than the slow process ($\mathbf{B}^{(f)} \gg \mathbf{B}^{(s)}$, $\mathbf{A}^{(s)} \gg \mathbf{A}^{(f)}$). Prior to adaptation, the initial values of the fast and slow processes were zero.

We assumed that the fast and slow processes linearly contribute to motor output y_k . The observation matrix \mathbf{H}_k in the observation equation,

$$y_k = \mathbf{H}_k(\mathbf{X}_k^{(f)} + \mathbf{X}_k^{(s)}), \quad (3.2)$$

extracts one performed movement direction depending on what target is presented; the movement direction is assumed to be the sum of fast and slow processes. The change $\Delta y_k = y_k - y_{k-1}$, where y_k is a constant imposed perturbation, is the directional error measured in work-space coordinates that drives motor adaptation. Equations 3.1 and 3.2 can be used to generate artificial directional error data for any sequence of target presentations.

We show below that the multitarget two-state model (see equations 3.1 and 3.2) can be reduced to the single-target two-state model (see equation 2.1) if averages over independent runs with randomized target orders are considered. Because the state-space model is not time invariant (the observation matrix \mathbf{H}_k depends on the target direction at the k th trial), an analytical solution cannot be obtained. If, however, target directions are presented in a uniformly random order, the mean learning curve averaged over repeated runs can be obtained by replacing \mathbf{H}_k with its expected value, $\bar{\mathbf{H}} = (1/8, \dots, 1/8)^T$. With this approximation, the fast and slow state vectors can be reduced to their mean scalars as

$$x_k^{(f)} = \frac{1}{8} \sum_{i=1}^8 (\mathbf{X}_k^{(f)})_i, \quad x_k^{(s)} = \frac{1}{8} \sum_{i=1}^8 (\mathbf{X}_k^{(s)})_i. \quad (3.3)$$

The multitarget state-space model, equations 3.1 and 3.2, is then reduced to the single-target state-space model, equations 2.1 and 2.2 with the following replacements:

$$a^{(f)} = \frac{1}{8} \sum_{i=1}^8 (\mathbf{A}^{(f)})_{ii}, \quad a^{(s)} = \frac{1}{8} \sum_{i=1}^8 (\mathbf{A}^{(s)})_{ii}, \quad (3.4)$$

$$b^{(f)} = \frac{1}{8^2} \sum_{i,j=1}^8 (\mathbf{B}^{(f)})_{ij}, \quad b^{(s)} = \frac{1}{8^2} \sum_{i,j=1}^8 (\mathbf{B}^{(s)})_{ij}. \quad (3.5)$$

Therefore, we can apply the same statistical tests (the AIC criterion of ML or the FPE criterion of the PE method) to multitarget state-space models to determine whether there is a single process or there are multiple concurrent processes. In psychophysical experiments, in order to determine whether there is a single process or there are multiple processes in multitarget motor adaptation, the statistical methods described in section 2 can be applied to learning curves obtained by averaging independent runs with randomized target presentation orders.

Here, to reduce the multitarget model to a single-target model, the average over independent runs with randomized target order was assumed. Although this is a standard procedure in machine learning, it is not possible in psychophysical experiments, where fixed pseudo-randomized target order is frequently used. We therefore simulated the two-state model equations 3.1 and 3.2, with a fixed target sequence by adding process noise ($\sigma_w^2 = 1 \times 10^{-4}$) and measurement noise ($\sigma_v^2 = 3 \times 10^{-4}$). Each learning curve was noisy, but their averages clearly demonstrated multiple-time behaviors. Therefore, even when a fixed target sequence is used, it is possible to detect multiple timescales if there are enough independent runs with independent noise.

Generalization of the fast and slow processes in principle could be determined by directly fitting the two-state model to the experimental trial-by-trial errors. It was, however, difficult to fit the two-state model reliably due to the large number of parameters. Instead, we used the single-state model approach, described below, to evaluate trial-by-trial generalization.

4 Analysis of Trial-by-Trial Generalization Using a Single-State State-Space Model

We briefly review an approach using a single-state state-space model to evaluate trial-by-trial generalization from trial error data. (Details can be found in Tanaka et al., 2009.) To assess how an error experienced in one direction influences learning in another direction, we used a single-state state-space model originally introduced for force-field adaptation. Here, we recast the single-state analysis into a form that is applicable to any form of motor adaptation, including visuomotor rotation learning. The analysis begins by defining a state-space model, which consists of a state vector, which is usually not directly observable, and an observable variable. The equations that define the single-state state-space model are

$$\begin{cases} \mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{B}\mathbf{H}_k^T \Delta y_k \\ y_k = \mathbf{H}_k \mathbf{X}_k \end{cases} \quad (4.1)$$

The first equation describes how an error Δy_k in the k th trial updates the current state vector \mathbf{X}_k to the next state \mathbf{X}_{k+1} , and the second equation describes which component of the state vector is actually observed. The matrix \mathbf{B} in equation 4.1 represents how the angular error experienced in the current trial updates the state vector and thus defines trial-by-trial generalization. To avoid overfitting, we assumed that the degree of trial-by-trial generalization depends on only the angular difference between the current and subsequent direction (-135° , -90° , -45° , 0° , 45° , 90° , 135° , and 180°). We also optimized an initial value of state vector \mathbf{X}_1 in order to account for initial biases.

In section 2, we showed that the ML and PE methods give similar results for estimating state-space model parameters. Therefore, for estimating trial-by-trial generalization, we used the PE method for its speed and simplicity. We looked for the parameter value that best explained the data, and the 16 parameters $\theta \equiv (\mathbf{B}, \mathbf{X}_1)$ were optimized so as to minimize the quadratic error:

$$\hat{\theta} = \underset{\lambda}{\operatorname{argmin}} \sum_k (\Delta y_k^{\text{data}} - \Delta y_k^{\text{trial-analysis}})^2 \quad (4.2)$$

Here the error time course $\{\Delta y_k^{\text{data}}\}$ was generated artificially with the two-state model introduced above (see equations 3.1 and 3.2). Optimal values of the parameters were searched for using the downhill simplex method (Press, Teukolsky, Vetterling, & Flannery, 2007). A confidence interval for the estimated parameters was computed using a standard bootstrap method (Efron, 1982), and we found that 200 independent samples were sufficient to obtain reasonably robust results.

5 Trial-by-Trial and Postadaptation Generalization Target the Fast and Slow Adaptation Processes, Respectively

Starting with the premise that learning has fast and slow processes, it is of interest to ask how each of these two processes generalizes. One way to compare generalization functions

for the fast and slow processes would be to directly fit a two-state state-space model, equations 3.1 and 3.2, to experimental data and compute the generalization matrices ($\mathbf{B}^{(f)}$ and $\mathbf{B}^{(s)}$). This is difficult to do, however, because of the high dimensionality of the state-space model with two states. The full model has 34 parameters (16 for initial conditions; 16 for generalization functions, assuming rotational symmetry; and 2 for the retention factors, assuming $\mathbf{A}^{(f)}$ and $\mathbf{A}^{(s)}$ to be diagonal with equal diagonal components).

Instead, we show formally that trial-by-trial generalization and postadaptation generalization are characterized by fast and slow processes, respectively. Specifically, using analysis and computational simulations, we show that trial-by-trial generalization of the state-space model with a single state arises mainly from fast learning, whereas the postadaptation generalization function derives from the slow process, provided that certain conditions regarding learning and retention speed are satisfied. This claim can be understood intuitively. Because the fast process has a larger gain than the slow process, the fast process will accordingly make larger corrections to trial errors. Trial-by-trial generalization reflects mainly the change in the fast process, namely, the $\mathbf{B}^{(f)}$ matrix in equation 3.1. After completion of adaptation, the fast process will have largely decayed due to weak retention, leaving the residue of the slow process. Thus, postadaptation generalization reflects mainly the net change in the slow process, namely, the $\mathbf{B}^{(s)}$ matrix in equation 3.1.

In the update equation, 3.1, during the initial phase of adaptation, the fast and slow states are close to their initial values, so the retention terms are negligible. Thus, trial-by-trial generalization is well approximated by the average of \mathbf{B} matrices ($b^{(f)}$ for $\mathbf{B}^{(f)}$ and $b^{(s)}$ for $\mathbf{B}^{(s)}$), and the fast process dominates trial-by-trial adaptation. Postadaptation generalization refers to the degree to which the fast and slow systems have retained the imposed perturbation after the learning has approached its asymptote. The asymptotic values of the fast and slow processes can be computed using the average learning curve, respectively:

$$\begin{cases} x^{(f)} = \frac{(1-a^{(s)})b^{(f)}}{\|\mathbf{I}-\mathbf{A}+\mathbf{B}\mathbf{C}\|} \bar{u} \\ x^{(s)} = \frac{(1-a^{(f)})b^{(s)}}{\|\mathbf{I}-\mathbf{A}+\mathbf{B}\mathbf{C}\|} \bar{u} \end{cases} \quad (5.1)$$

Therefore, the ratio of the average fast and slow processes is given by

$$\frac{x^{(s)}}{x^{(f)}} = \frac{(1-a^{(f)})b^{(s)}}{(1-a^{(s)})b^{(f)}}. \quad (5.2)$$

If $(1-a^{(s)})^{-1}b^{(s)} > (1-a^{(f)})^{-1}b^{(f)}$ is satisfied, then the learned remapping (probed by testing for post-adaptation generalization) is stored mainly in the slow process. Note that for single-target viscous force field adaptation, the values for the learning gains and retention factors found experimentally satisfy the above conditions: $(1-a^{(s)})^{-1}b^{(s)} = 2.5 > (1-a^{(f)})^{-1}b^{(f)} = 0.51$, calculated from the parameter values best fit to the experimental data ($a^{(f)} = 0.59$, $a^{(s)} = 0.992$, $b^{(f)} = 0.21$ and $b^{(s)} = 0.02$) (Smith et al., 2006).

To confirm this analysis, we performed computational simulations with three combinations of narrow, intermediate, and broadly tuned \mathbf{B} matrices (see Figure 8). In the following simulations, the retention-factor matrices were assumed to be diagonal, with all diagonal values $\mathbf{A}^{(f)} = 0.95 \cdot \mathbf{I}_8$ and $\mathbf{A}^{(s)} = 0.998 \cdot \mathbf{I}_8$, respectively. When parameterizing \mathbf{B} , we assumed rotational symmetry, as for the single-state model, so each matrix was described by eight row components. We assumed that the diagonal components of the generalization

matrices, which determine trial-by-trial generalization to the same direction, took a maximal value, and that the value of the off-diagonal components decreased as the angular difference between the learned and tested directions increased. The maximal values of the generalization matrices were fixed at 0.12 and 0.03, respectively, for the fast and slow processes. With these parameter values, the statistical tests described in section 2 identify two concurrent processes rather than one. Similar results were also observed when the values for the fast process were larger than those of the slow process and $(1 - a^{(s)})^{-1} b^{(s)} > (1 - a^{(f)})^{-1} b^{(f)}$. Artificial error data were generated using the two-state model with various gaussian widths for $\mathbf{B}^{(f)}$ and $\mathbf{B}^{(s)}$: (A) $(\sigma^{(f)}, \sigma^{(s)}) = (1^\circ, 60^\circ)$, (B) $(\sigma^{(f)}, \sigma^{(s)}) = (30^\circ, 30^\circ)$, and (C) $(\sigma^{(f)}, \sigma^{(s)}) = (60^\circ, 1^\circ)$. Our analysis indicated that the sums over the components of \mathbf{B} matrices, equation 4.1, determine how much the fast and slow processes contribute to trial-by-trial and postadaptation generalization, respectively. The values of $b^{(f)}$, $b^{(s)}$, $(1 - a^{(f)})^{-1} b^{(f)}$, and $(1 - a^{(s)})^{-1} b^{(s)}$ computed from the various gaussian widths are summarized in Table 1. The conditions $b^{(f)} > b^{(s)}$ and $(1 - a^{(s)})^{-1} b^{(s)} > (1 - a^{(f)})^{-1} b^{(f)}$ are satisfied for these three sets of gaussian widths.

We first investigated which process contributes more to trial-by-trial generalization. The two-state model was trained with eight targets to produce artificial error data, and trial-by-trial generalization was then assessed using the single-state state-space model. The shapes of the trial-by-trial generalization functions were found to reflect mainly the width of the \mathbf{B} matrix for the fast process (the middle row of Figure 8). We next investigated which process contributes most to postadaptation generalization. The two-state model was trained with a single target with the motor perturbation, and the degree of generalization was computed after the training was completed. As predicted, the width of the postadaptation generalization function mainly reflects the slow process (the bottom row of Figure 8). Trial-by-trial and postadaptation generalization have similar shapes only when the fast and slow processes have similar widths for their \mathbf{B} matrices.

In summary, we conclude that if motor adaptation has a fast and a slow adaptation process, then trial-by-trial generalization will reflect the fast process and postadaptation generalization the slow process.

A recent study demonstrated that the decrease in relearning rate in A_1BA_2 paradigms, or anterograde interference, can be explained with a two-state model (Sing & Smith, 2010). Anterograde interference is defined as how the memory of task A_1 interferes with the subsequent learning of task B. A two-state model attributes anterograde interference to the memory stored in the slow process, and the degree of anterograde interference increases with the duration of learning of task A_1 . These predictions from the two-state model were confirmed experimentally.

Our multitarget two-state model posits that the slow process contributes mainly to postadaptation generalization once adaptation is almost complete. The model also makes an interesting prediction: the pattern of postadaptation should reflect that of the fast process if adaptation is prematurely terminated and should approach that of slow process asymptotically. Therefore, in addition to the degree of anterograde interference, a buildup of the slow process should be manifest as a gradual transition in the shape of postadaptation generalization if the fast and slow systems have distinct generalization matrices.

We simulated anterograde interference and postadaptation generalization with various numbers of single-target adaptation trials. We considered a narrow (1°) and a broad (90°) shape of fast and slow process, respectively, and vice versa. In initial 10, 20, 40, and 160 trials, positive perturbations (+1) were imposed, and in subsequent trials, negative perturbations (−1) were delivered. The adaptation rate became slower as the number of

adaptation trials increased, suggesting anterograde interference of prior learning of the positive perturbations to subsequent learning of the negative perturbations (see Figure 9A). Accordingly, the shape of postadaptation generalization changed gradually from one mostly reflecting the fast process to another mostly reflecting the slow process (see Figures 9B and 9C).

6 Discussion

Starting with the premise that motor adaptation is mediated by slow and fast processes (Smith et al., 2006), we showed analytically and with computational simulations that these two processes lead to double exponential learning curves and are reflected, respectively, in the postadaptation and the trial-by-trial measures of generalization. The question of whether the fast and slow processes had the same or different generalization functions was not addressed in the original two-rate adaptation model. Our formal demonstration suggests that psychophysical investigation using the two generalization measures can probe their corresponding neural representations.

6.1 Visuomotor Rotation Learning

The generalization function for rotation learning had the same shape whether it was obtained by testing in untrained directions at the end of adaptation or derived from a state-space model with a single state (Tanaka et al., 2009). There are two potential explanations. The first is that rotation adaptation is single-rate. This is certainly the case for one-target rotation adaptation (Zarahn et al., 2008). The second is that multitarget rotation adaptation is two-rate. We found that double exponentials fits were significantly better than single exponentials to eight-target rotation adaptation data in our previous paper (Tanaka et al., 2009), consistent with an initial fast adaptation process followed by a slower one (Smith et al., 2006). This difference between one-target and multitarget learning could be biologically interesting in and of itself, or it could be an artifact of the parameter estimation procedures. Multitarget learning curves were assessed using 264 trials (Tanaka et al., 2009), but in single-target learning, only 80 trials were analyzed (Zarahn et al., 2008). As we noted, the AIC criterion prefers single-state models to two-state models when the number of error trials is insufficient (see equation 2.15 and section 2.3.1). Further experiments with different-sized rotations and different target numbers may help resolve this issue.

6.2 Force-Field Adaptation

For force-field adaptation, different generalization functions have been described depending on the approach used to derive them. The postadaptation approach to generalization for viscous force-field learning found no generalization beyond 90 degrees from the training direction (Mattar & Ostry, 2007). In contrast, state-space models yielded a broader bimodal pattern of trial-by-trial generalization (Thoroughman & Shadmehr, 2000; Donchin et al., 2003). Our theoretical analysis suggests that this apparent contradiction arises because Mattar and Ostry (2007) probed the slow adaptation process, whereas previous force-field generalization studies studied the fast process. This distinction indicates that in the case of force-field adaptation, fast and slow processes have distinct neural correlates with different directional tuning properties. Thus, force-field adaptation differs from rotation adaptation with respect to both the dimensionality of the state-space model and the widths of the generalization functions. This is in agreement with our previous contention that kinematic and dynamic adaptation are distinct learning processes (Krakauer, Ghilardi, & Ghez, 1999). It is also of interest that recent work has shown that the fast process in force-field adaptation may be mechanistically distinct from the slow process (Smith et al., 2006; Keisler & Shadmehr, 2010). Thus, differences in generalization for the two processes might be less surprising.

Further comment is required, however, with respect to the state-space modeling results for viscous curl force-field learning that have suggested substantial bimodality, with generalization to opposite movement directions reaching nearly 90% if invariant desired trajectories were assumed (Donchin et al., 2003). However, results from the same paper suggest less than 10% generalization to opposite-movement directions if desired trajectories were assumed to be variable. State-space modeling results from Thoroughman and Shadmehr (2000) suggest about 27% generalization to opposite-movement directions. A direct comparison between those results is difficult because these two papers adopted different definitions for their generalization functions, with motor error given in terms of either a scalar or vector. Whether the disparity between these results resulted from this methodological subtlety will require future confirmatory experiments.

6.3 Nonlinear Interactions

We should emphasize that the state-space framework assumes that adaptation processes are linear time-invariant (LTI). It is interesting that under certain experimental conditions, adaptation paradigms do indeed generate several phenomena that are well predicted by LTI state-space models (Smith et al., 2006; Tanaka et al., 2009). However, adaptation experiments also yield phenomena that are not LTI. For example, we recently showed that savings can occur for rotation adaptation after complete washout (Zarahn et al., 2008), a finding that is incompatible with any LTI multirate model. Similarly, in previous work, we showed that interference can persist in the absence of aftereffects (Krakauer, Ghez, & Ghilardi, 2005). Thus, dissociation of savings and interference effects cannot be explained by current LTI state-space models. One proposed solution is that memories can be protected by contextual cues (Lee & Schweighofer, 2009); modulation by context is by definition nonlinear. Alternatively, qualitatively distinct memory processes may be operating over trials compared to those operating over longer time periods; that is, time and trial may have distinct effects, with only trials corresponding to LTI-based state-space models. We suggest that error-based LTI models capture short-term adaptation processes that are likely cerebellar based (Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007; Criscimagna-Hemminger, Bastian, & Shadmehr, 2010; Taylor & Ivry, 2011) and that deviations from LTI models represent the presence of additional learning processes, and not necessarily an error in the models themselves (Huang, Haith, Mazzoni, & Krakauer 2011). It is to be hoped that the structure of behavioral deviation from the predictions of LTI state-space models will be informative and can lead to new experiments and the formulation of new models.

Acknowledgments

We thank Eric Zarahn for his help in simulating how the two approaches to the study of generalization map onto the fast and slow adaptation processes, Maurice Smith for critical comments, and two anonymous reviewers for helpful comments. This research was supported by the Howard Hughes Medical Institute (T.J.S.) and NINDS R01 052804 (J.W.K.).

References

- Akaike H. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*. 1969; 21:243–247.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control*. 1974; 19:716–723.
- Baddeley RJ, Ingram HA, Miall RC. System identification applied to a visuomotor task: Near-optimal human performance in a noisy changing task. *J Neurosci*. 2003; 23:3066–3075. [PubMed: 12684493]
- Cheng S, Sabes PN. Calibration of visually guided reaching is driven by error-corrective learning and internal dynamics. *J Neurophysiol*. 2007; 97:3057–3069. [PubMed: 17202230]

- Criscimagna-Hemminger SE, Bastian AJ, Shadmehr R. Size of error affects cerebellar contributions to motor learning. *J Neurophysiol.* 2010; 103:2275–2284. [PubMed: 20164398]
- Criscimagna-Hemminger SE, Shadmehr R. Consolidation patterns of human motor memory. *J Neurosci.* 2008; 28:9610–9618. [PubMed: 18815247]
- 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.* 2003; 23:9032–9045. [PubMed: 14534237]
- Efron, B. The jackknife, the bootstrap and other resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics; 1982.
- Ethier V, Zee DS, Shadmehr R. Spontaneous recovery of motor memory during saccade adaptation. *J Neurophysiol.* 2008; 99:2577–2583. [PubMed: 18353917]
- Fine MS, Thoroughman KA. Motor adaptation to single force pulses: Sensitive to direction but insensitive to within movement pulse placement and magnitude. *J Neurophysiol.* 2006; 96:710–720. [PubMed: 16707722]
- Francis JT. Error generalization as a function of velocity and duration: Human reaching movements. *Exp Brain Res.* 2008; 186:23–37. [PubMed: 18030456]
- Gandolfo F, Mussa-Ivaldi FA, Bizzi E. Motor learning by field approximation. *Proc Natl Acad Sci USA.* 1996; 93:3843–3846. [PubMed: 8632977]
- Goodwin, GC.; Sin, KS. Adaptive filtering prediction and control. Englewood Cliffs, NJ: Prentice-Hall; 1984.
- Gupta N, Mehra R. Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. *IEEE Trans Automatic Control.* 1974; 19:774–783.
- Huang VS, Haith A, Mazzoni P, Krakauer JW. Rethinking motor learning and savings in adaptation paradigms: Model-free memory for successful actions combines with internal models. *Neuron.* 2011; 70:787–801. [PubMed: 21609832]
- Imamizu H, Uno Y, Kawato M. Internal representations of the motor apparatus: Implications from generalization in visuomotor learning. *J Exp Psychol Hum Percept Perform.* 1995; 21:1174–1198. [PubMed: 7595246]
- Joiner WM, Smith MA. Long-term retention explained by a model of short-term learning in the adaptive control of reaching. *J Neurophysiol.* 2008; 100:2948–2955. [PubMed: 18784273]
- Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng.* 1960; 82:35–45.
- Keisler A, Shadmehr R. A shared resource between declarative memory and motor memory. *J Neurosci.* 2010; 30:14817–14823. [PubMed: 21048140]
- Körding KP, Tenenbaum JB, Shadmehr R. The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat Neurosci.* 2007; 10:779–786. [PubMed: 17496891]
- Krakauer JW, Ghez C, Ghilardi MF. Adaptation to visuomotor transformations: Consolidation, interference, and forgetting. *J Neurosci.* 2005; 25:473–478. [PubMed: 15647491]
- Krakauer JW, Ghilardi MF, Ghez C. Independent learning of internal models for kinematic and dynamic control of reaching. *Nat Neurosci.* 1999; 2:1026–1031. [PubMed: 10526344]
- Krakauer JW, Pine ZM, Ghilardi MF, Ghez C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci.* 2000; 20:8916–8924. [PubMed: 11102502]
- Lee JY, Schweighofer N. Dual adaptation supports a parallel architecture of motor memory. *J Neurosci.* 2009; 29:10396–10404. [PubMed: 19692614]
- Ljung, L.; Ljung, EJ. System identification: Theory for the user. Englewood Cliffs, NJ: Prentice-Hall; 1987.
- Mattar AA, Ostry DJ. Modifiability of generalization in dynamics learning. *J Neurophysiol.* 2007; 98:3321–3329. [PubMed: 17928561]
- Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical recipes: The art of scientific computing. Cambridge: Cambridge University Press; 2007.
- Shadmehr R. Generalization as a behavioral window to the neural mechanisms of learning internal models. *Hum Mov Sci.* 2004; 23:543–568. [PubMed: 15589621]
- Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. *J Time Series Analysis.* 1982; 3:253–264.

- Shumway, RH.; Stoffer, DS. Time series analysis and its applications. Berlin: Springer-Verlag; 2000.
- Sing GC, Smith MA. Reduction in learning rates associated with anterograde interference results from interactions between different timescales in motor adaptation. *PLoS Comput Biol.* 2010; 6
- Smith MA, Ghazizadeh A, Shadmehr R. Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* 2006; 4:e179. [PubMed: 16700627]
- Tanaka H, Sejnowski TJ, Krakauer JW. Adaptation to visuomotor rotation through interaction between posterior parietal and motor cortical areas. *J Neurophysiol.* 2009; 102:2921–2932. [PubMed: 19741098]
- Taylor JA, Ivry RB. Flexible cognitive strategies during motor learning. *PLoS Comput Biol.* 2011; 7:e1001096. [PubMed: 21390266]
- Thoroughman KA, Shadmehr R. Learning of action through adaptive combination of motor primitives. *Nature.* 2000; 407:742–747. [PubMed: 11048720]
- Tseng YW, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J Neurophysiol.* 2007; 98:54–62. [PubMed: 17507504]
- Zarahn E, Weston GD, Liang J, Mazzoni P, Krakauer JW. Explaining savings for visuomotor adaptation: Linear time-invariant state-space models are not sufficient. *J Neurophysiol.* 2008; 100:2537–2548. [PubMed: 18596178]

Appendix A: Trial-Based and Time-Based State-Space Models

Here we discuss the relationship between continuous- and discrete-time state-space models and show that the learning coefficients in discrete-time models depend on the forgetting factors for continuous-time models and the intertrial interval (see also Ethier et al., 2008). The decay of motor memory is a continuous-time process, whereas error-based learning is a trial-based, discrete-time process:

$$\frac{d\tilde{\mathbf{x}}}{dt}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{x}}(t) + \tilde{\mathbf{B}}\Delta\tilde{\mathbf{y}}(t), \quad (\text{A.1})$$

where the error correction term is $\Delta\tilde{\mathbf{y}}(t) = \sum_k \Delta y_k \delta(t - k\tau_{\text{TI}})$, which takes nonzero values only at movement trials. τ_{TI} is an intertrial interval, and the k th trials occur at $t = k\tau_{\text{TI}}$. Here we use a tilde notation to indicate continuous-time variables. The matrices are defined as

$$\tilde{\mathbf{A}} = \begin{pmatrix} \tilde{a}^{(f)} & 0 \\ 0 & \tilde{a}^{(s)} \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \tilde{b}^{(f)} \\ \tilde{b}^{(s)} \end{pmatrix}. \quad (\text{A.2})$$

$-\tilde{a}^{(f)}$ and $-\tilde{a}^{(s)}$ are the reciprocals of decay time constants of fast and slow systems, respectively. Here, by definition, $0 > \tilde{a}^{(s)} > \tilde{a}^{(f)}$ and $\tilde{b}^{(f)} > \tilde{b}^{(s)} > 0$. By integrating equation A.1 over the time period $[k\tau_{\text{TI}}, (k+1)\tau_{\text{TI}}]$, we obtain the discrete-time state-space model, equation 2.1. The matrices are equated as

$$\mathbf{A} = \begin{pmatrix} a^{(f)} & 0 \\ 0 & a^{(s)} \end{pmatrix} = \begin{pmatrix} e^{\tilde{a}^{(f)}\tau_{\text{ISI}}} & 0 \\ 0 & e^{\tilde{a}^{(s)}\tau_{\text{ISI}}} \end{pmatrix}, \quad (\text{A.3})$$

$$\mathbf{B} = \begin{pmatrix} b^{(f)} \\ b^{(s)} \end{pmatrix} = \begin{pmatrix} e^{\tilde{a}^{(f)}\tau_{\text{ISI}}} \tilde{b}^{(f)} \\ e^{\tilde{a}^{(s)}\tau_{\text{ISI}}} \tilde{b}^{(s)} \end{pmatrix}.$$

Note that the learning coefficients in the discrete-time state-space model ($b^{(f)}$ and $b^{(s)}$) depend not only those in the continuous-time model ($\tilde{b}^{(f)}$ and $\tilde{b}^{(s)}$) but also the retention coefficients ($\tilde{a}^{(f)}$ and $\tilde{a}^{(s)}$) and the intertrial interval (τ_{TI}). We obtain the following expression of the ratio of the discrete-time learning coefficients:

$$\frac{b^{(f)}}{b^{(s)}} = \frac{\tilde{b}^{(f)}}{\tilde{b}^{(s)}} \exp[(\tilde{a}^{(f)} - \tilde{a}^{(s)})t_{\text{ITI}}]. \quad (\text{A.4})$$

The exponential factor stems from the decay of the fast and slow processes during the intertrial intervals; the weak retention factor of the fast system reduces its learned gain, whereas the learned gain of the slow system is maintained due to its strong retention. Since $\tilde{a}^{(f)} - \tilde{a}^{(s)} < 0$, the ratio of the learning coefficients decays exponentially as the intertrial interval increases. Thus, in order to best observe the contribution of the fast process in trial-by-trial adaptation, the intertrial interval should be minimized (Joiner & Smith, 2008).

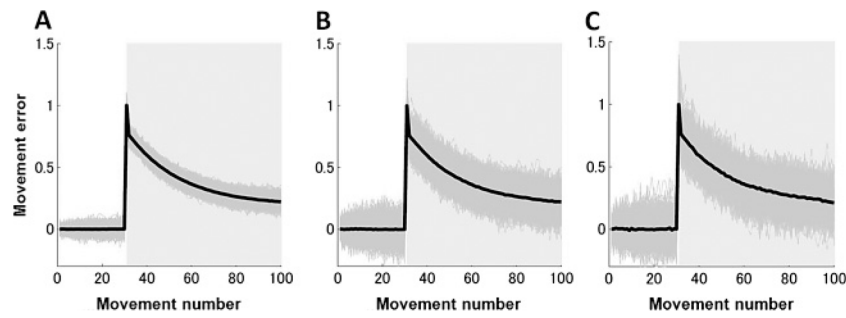


Figure 1.

Typical error curves generated by simulating the two-state model, equation 2.1, with (A) $\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$, (B) $\sigma_w^2 = 5 \times 10^{-4}$ and $\sigma_v^2 = 15 \times 10^{-4}$, and (C) $\sigma_w^2 = 10 \times 10^{-4}$ and $\sigma_v^2 = 30 \times 10^{-4}$. Thin gray lines depict individual realizations, and thick black lines depict average error curves. External perturbations are 0 in trials 1–30 and 1 in trials 31–100 (shaded areas).

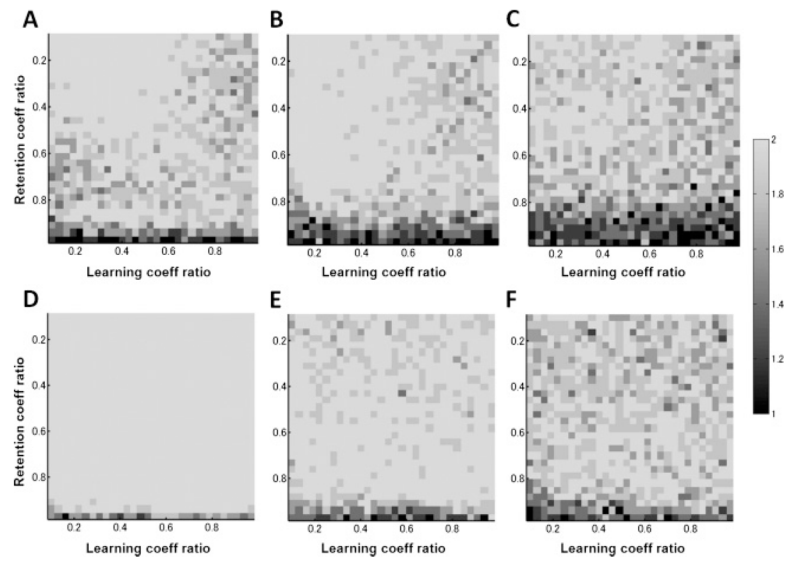


Figure 2.

Estimation of state-space dimensionality using ML estimation (A, B, C) and PE estimation (D, E, and F). Three levels of noises were considered: (A, D) $\sigma_w^2=1 \times 10^{-4}$ and $\sigma_v^2=3 \times 10^{-4}$; (B, E) $\sigma_w^2=5 \times 10^{-4}$ and $\sigma_v^2=15 \times 10^{-4}$; and (C, F) $\sigma_w^2=10 \times 10^{-4}$ and $\sigma_v^2=30 \times 10^{-4}$.

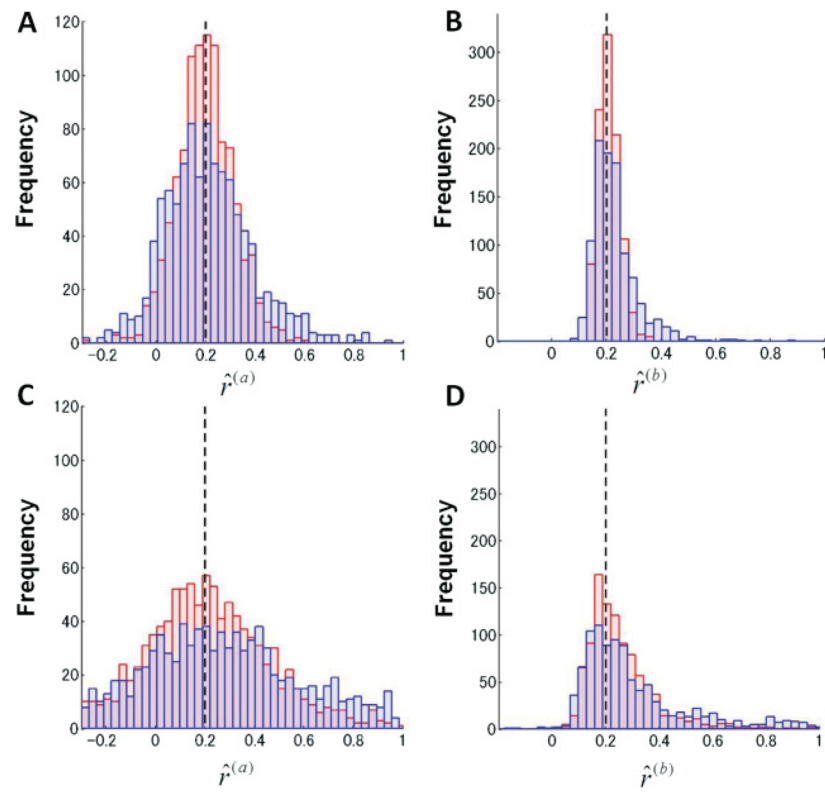


Figure 3.

Distributions of the estimated coefficient ratios by Monte Carlo simulations using ML estimation (red histograms) and PE estimation (blue histograms) for noise values (A, B) $\sigma_w^2=1 \times 10^{-4}$ and $\sigma_v^2=3 \times 10^{-4}$ and (C, D) $\sigma_w^2=5 \times 10^{-4}$ and $\sigma_v^2=15 \times 10^{-4}$. The dashed vertical lines indicate the value of the ratios used to generate artificial error data.

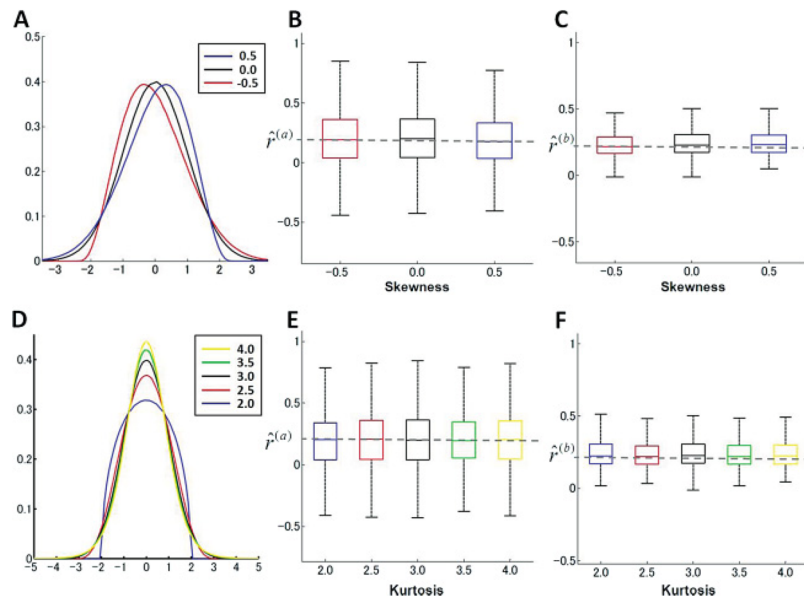


Figure 4. Robustness of parameter estimation for nongaussian noise. (A) Skewness systematically varied. (B, C) Box plots of estimated ratios of parameters for the skewed distributions. (D) Kurtosis was systematically varied. (E, F) Box plots of estimated ratios of parameters for sub- and supergaussian distributions.

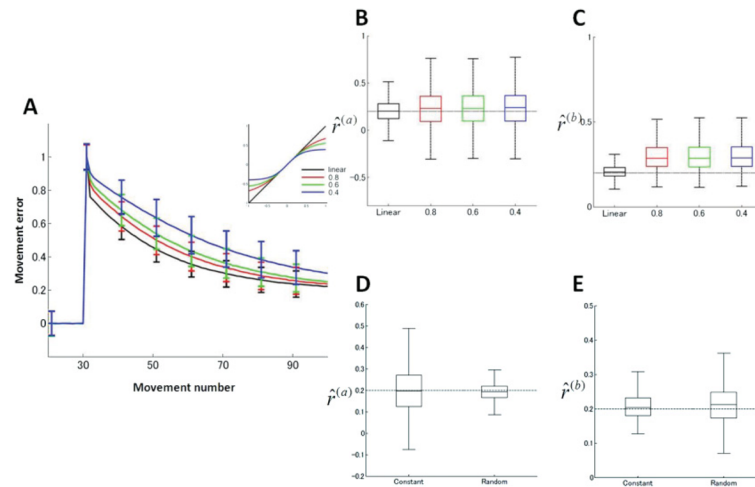


Figure 5.

Robustness of the learning for nonlinear learning functions. (A) Learning curves for a nonlinear, saturating function (see equation 2.19). (B, C) Box plots of estimated ratios of model parameters. (D, E) Box plots of estimated ratios of model parameters for random perturbations.

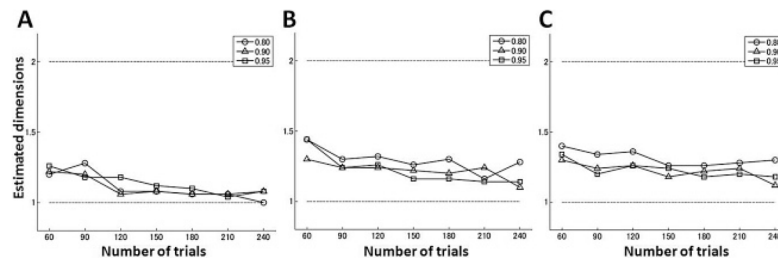


Figure 6.

Averaged estimated dimension as a function of the number of trials when error data were generated from a single-state model. The noise variances were (A) $\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$, (B) $\sigma_w^2 = 5 \times 10^{-4}$ and $\sigma_v^2 = 15 \times 10^{-4}$, and (C) $\sigma_w^2 = 10 \times 10^{-4}$ and $\sigma_v^2 = 30 \times 10^{-4}$.

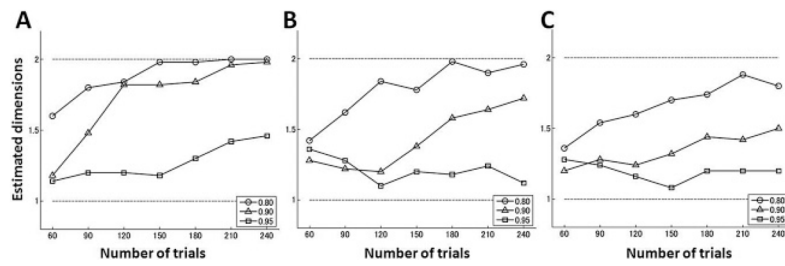


Figure 7.

Averaged estimated dimension as a function of the number of trials when error data were generated from a two-state model. The noise variances were (A) $\sigma_w^2 = 1 \times 10^{-4}$ and $\sigma_v^2 = 3 \times 10^{-4}$, (B) $\sigma_w^2 = 5 \times 10^{-4}$ and $\sigma_v^2 = 15 \times 10^{-4}$, and (C) $\sigma_w^2 = 10 \times 10^{-4}$ and $\sigma_v^2 = 30 \times 10^{-4}$.

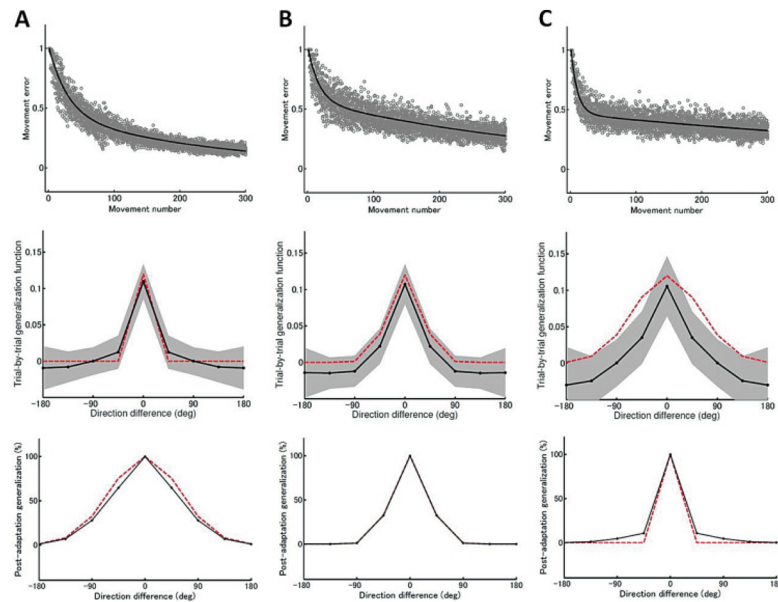


Figure 8.

Double exponential learning curves (top row), trial-by-trial generalization functions (middle row), and postadaptation generalization functions (bottom row) computed from a two-state state-space model with various widths for the fast process's $\mathbf{B}^{(f)}$ and the slow process's $\mathbf{B}^{(s)}$ matrices defined in equations 3.1 and 3.2. (A) narrow $\mathbf{B}^{(f)}$ and broad $\mathbf{B}^{(s)}$, (B) moderate $\mathbf{B}^{(f)}$ and $\mathbf{B}^{(s)}$, and (C) broad $\mathbf{B}^{(f)}$ and narrow $\mathbf{B}^{(s)}$. Red lines denote generalization functions used to generate error data. The learning curves were computed from 10 independent runs and fitted by a double-exponential function (thick black line). Trial-by-trial generalization was assessed using the initial 50 trials, and postadaptation generalization was assessed after 300 trials.

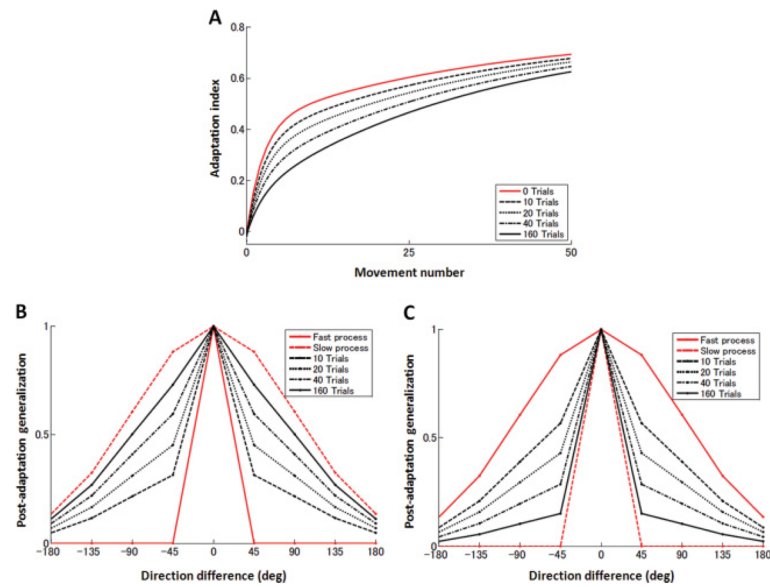


Figure 9.

(A) Learning curves of opposite perturbations (-1) (aligned at the zero crossing point) after learning positive perturbations for 10, 20, 40, and 160 trials. The red solid line denotes the learning curve without prior learning as a comparison. (B, C) Gradual transition in the shape of postadaptation generalization after learning positive perturbations for 10, 20, 40, and 160 trials. The red solid and dashed lines show the normalized generalization functions for the fast and slow systems, respectively.

Table 1

Summary of Parameter Values Used for the Simulation of Figure 8.

	$b^{(f)}$	$b^{(s)}$	$(1 - a^{(f)})^{-1}b^{(f)}$	$(1 - a^{(s)})^{-1}b^{(s)}$
(A) $(\sigma^{(f)}, \sigma^{(s)}) = (1^\circ, 60^\circ)$	0.12	0.099	2.40	49.95
(B) $(\sigma^{(f)}, \sigma^{(s)}) = (30^\circ, 30^\circ)$	0.20	0.050	4.012	25.05
(C) $(\sigma^{(f)}, \sigma^{(s)}) = (60^\circ, 1^\circ)$	0.40	0.030	8.10	15.00