The Binding of Learning to Action in Motor Adaptation

Text S1

Example of credit assignment for error-based learning: the gradient-descent learning rule

The familiar gradient-descent learning rule [1-2] provides an example of credit assignment during error-based learning. This rule updates the state of a learning system so that for a given (arbitrarily small) step size, maximal reduction of error will be achieved for a subsequent trial with the same input. If the state of the learning system can be characterized by a vector of parameters ($\boldsymbol{w}$), then the output of the system (y) is related to its input ($\boldsymbol{x}$) by:

$$y = f(\boldsymbol{w}, \boldsymbol{x})$$

For such a system, the error (e) is defined as the difference between actual output (y) and desired output (y^*).

$$e = y^* - y$$

The gradient-descent learning rule maximizes the reduction of squared error for a given step size (α) as follows:

$$\Delta \boldsymbol{w} = -\alpha \frac{\partial e^2}{\partial \boldsymbol{w}} = -2\alpha \frac{\partial e}{\partial \boldsymbol{w}} e = -2\alpha \frac{\partial y}{\partial \boldsymbol{w}} e = -2\alpha \frac{\partial f(\boldsymbol{w},\boldsymbol{x})}{\partial \boldsymbol{w}} e$$

Note that since y^* is independent of $\boldsymbol{w}$, $\partial e/\partial \boldsymbol{w} = \partial y/\partial \boldsymbol{w}$ as implied above. Rewriting this equation in expanded notation we have:

$$\begin{bmatrix} \Delta w_1 \\ \vdots \\ \Delta w_N \end{bmatrix} = -2\alpha \cdot \begin{bmatrix} \partial f(\boldsymbol{w},\boldsymbol{x})/\partial w_1 \\ \vdots \\ \partial f(\boldsymbol{w},\boldsymbol{x})/\partial w_N \end{bmatrix} \cdot e$$

Note that each element of the gradient vector ($f(\boldsymbol{w}, \boldsymbol{x})/\partial w_k$) represents the sensitivity of the output with respect to the corresponding component (w_k) of the parameter vector ($\boldsymbol{w}$). Thus each parameter will be updated in proportion to both the size of the error and this specific sensitivity. Then, the parameters which have the greatest ability to drive error on a given trial, receive the greatest credit for that error in the sense that their values are modified to the greatest extent: The learned change to each parameter is proportional to both the size of the error and the credit assigned to it.

Modeling and simulation of credit assignment for force-field interference experiments

We simulated the adaptation process in the FF interference experiments for the PRL and MRL credit assignment schemes. We used a linear multi-rate learning model (Smith et al., 2006) with Gaussian basis functions across movement directions to model adaptation and generalization, Equations S1-S4, tailored to the plan-referenced and motion-referenced learning credit assignment schemes.

(S1) $$x(n+1,\theta) = x_s(n+1,\theta) + x_f(n+1,\theta)$$

(S2) $$x_s(n+1,\theta) = A_s x_s(n,\theta) + B_s(\theta)e(n)$$

(S3) $$x_f(n+1,\theta) = A_f x_f(n,\theta) + B_f(\theta)e(n)$$

(S4) $$e(n) = \theta_{desired} - \theta_{actual}(n)$$

In this model $x(n+1, \theta)$ (Equation S1) represents the learning level after trial $n+1$ in direction θ. This learning, is the sum of the learning of two distinct states, termed slow and fast (Equations S2 and S3), with the same mathematical form. Specifically, the sum of the learning level at the previous trial discounted by a retention factor, A, and the error in the previous trial, $e(n)$ – the angular difference between the desired movement direction and the actual movement direction (Equation S4) – weighted by a learning function, $B(\theta)$.

We modeled motor adaptation using Gaussian functions for the generalization of learning centered at either the planned movement direction (for PRL) or the actual movement direction (for MRL) and constant retention factors (Thoroughman and Shadmehr, 2000). The parameters for the peaks of the Gaussian functions (B_s = 0.020 and B_f = 0.210) as well as the retention coefficients (A_s = 0.992 and A_f = 0.590) for the two states were obtained from a previously published unidirectional 2-state model of motor adaptation (Smith et al., 2006). The standard deviations (σ_{Bs} and σ_{Bf}) of the Gaussian learning functions were set to 37 based on the results of an independent generalization experiment [3]. In this experiment, the full width at half maximum (FWHM) for the generalization functions was measured to be 87° on average (and for a Gaussian function, $= FWHM / \left[2\sqrt{2\ln(2)}\right]$).

The learning rule for the plan-referenced learning hypothesis is given by Equation S5. Here, the amount of adaptation arising from a given trial depends on the size of the error, and this adaptation is centered on the planned movement direction.

(S5) $$x(n+1,\theta) = Ax(n,\theta) + B\big(\theta - \theta_{planned}(n)\big)e(n)$$

On the other hand, for the motion-referenced credit assignment hypothesis, the update for each of the internal processes of the system is given by Equation S6. Here, the amount of adaptation arising from given trial depends on the size of the error, and this adaptation is centered on the actual movement direction.

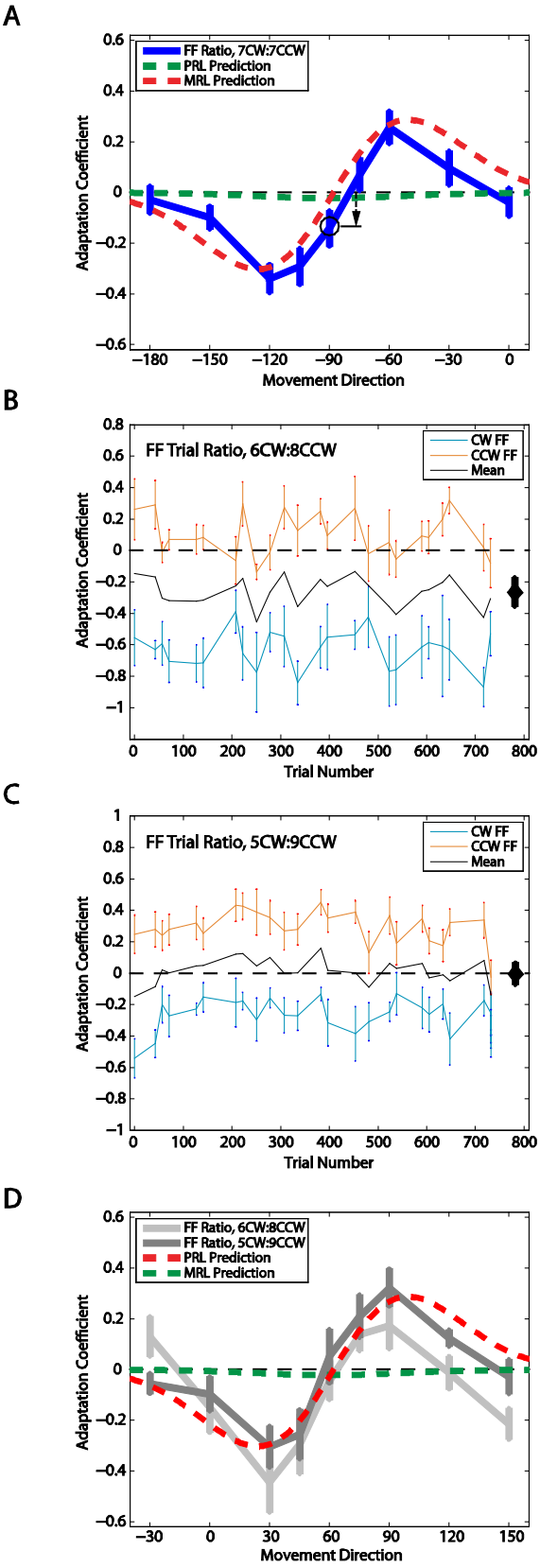
(S6) $$x(n+1,\theta) = Ax(n,\theta) + B(\theta - \theta_{actual}(n))e(n)$$

Note that these general equations (Equations S5 and S6) for updating the state of adaptation (x), each hold independently for the fast and slow learning processes (x_f and x_s), but with different parameter values for the learning rates (B_s and B_f) and retention factors (A_s and A_f) as in Equations S2 and S3 above.

Adjustment of force-field trial ratios in the force-field interference experiments

The first group of subjects (n=12) in the force-field interference experiments were trained with a target location of -90° (270°). In this group the ratio of CW to CCW FF trials was 7:7, and we noticed that the FF learning was biased towards the CW FF. Figure 1A shows that the adaptation level at the training direction (-90°; see black circle) is negative (about -0.15) consistent with greater learning of the CW FF (these data are also shown in Figure 2C in the main manuscript). Although we did not notice a difference in learning rates for the CW and CCW force-fields in an additional generalization experiment [3], this small bias is not surprising because we have previously observed it during single-trial motor adaptation (Sing et al., 2009; see *CW and CCW Learning* section in the Supplemental Information as well as Figure S7 of that study).

In order to replicate the results of the force-field interference experiment training in the -90° direction we conducted a second experiment, this time training subjects along the 60° direction. Having noticed that the learning had previously been biased towards the CW FF, we decided to modify the ratio of CW to CCW FF trials to 6:8 and track the evolution of the learning with randomly interspersed EC trials during the training blocks. After collecting data from six subjects we noticed that during training the FF learning was still biased towards the CW FF (see Figure 1B), a fact that was also reflected in the generalization pattern obtained for this group which appear to be shifted down (Figure 1D, light grey trace). We then decided to modify the ratio of CW to CCW FF trials to 5:9. After this adjustment, the FF learning observed during training became balanced and the bias was eliminated (see Figure 1C). This gave rise to a generalization pattern with peaks of almost symmetric heights (see Figure 1D, dark grey trace). Notice that irrespective of the balancing of FF learning, the shape of the generalization patterns for the three data sets (-90° CW:CCW = 7:7, 60° CW:CCW = 6:8 and 60° CW:CCW = 5:9) is in line with the predictions of the motion-referenced learning hypothesis –



the correlation coefficients for these three data sets are 0.92 ($F(1,7) = 36.87$, $p < 0.001$), 0.75 ($F(1,7) = 8.81$, $p < 0.05$), and 0.93 ($F(1,7) = 42.61$, $p < 0.001$), respectively.

Figure 1. Results of FF interference generalization experiments. **A**. Generalization pattern for force-field interference experiment training at -90°. Notice that the adaptation is biased towards the CW FF (negative adaptation). This is apparent at the training direction (black circle). **B**. Evolution of adaptation during replication experiment, training at 60° and using a CW to CCW FF trial ratio of 6:8. Notice that the mean adaptation remains consistently biased towards the CW FF (black trace and black diamond for overall mean adaptation). **C**. Evolution of adaptation during replication experiment, training at 60° and using a CW to CCW FF trial ratio of 5:9. Notice that here the mean adaptation is not biased toward the CW or CCW FF (black trace and black diamond for overall mean adaptation). **D**. Generalization pattern for the interference experiments with training at 60°. Notice that although the adaptation is biased towards the CW FF (negative adaptation) for a 6:8 CW:CCW FF trial ratio (light grey trace), this bias is removed when the ratio is modified to 5:9 (dark grey trace). All generalization patterns are consistent with the prediction of the MRL hypothesis ($r > 0.7$ in all cases). The error bars represent standard errors.

Comparison of force-field adaptation in different training paradigms for extended training

The MRL hypothesis predicts that the LST training paradigm should produce faster learning than the standard STT paradigm, and the control RST paradigm. We found that as predicted by the MRL simulations (Figure 2A) the LST paradigm does produce faster learning early in training than the STT (t_{58} = -2.17, $p < 0.02$) and RST (t_{58} = -5.05, $p < 3 \times 10^{-6}$) paradigms (Figure 2B). However as training proceeds and angular errors decrease, the misalignment between prescribed and actual motion directions decreases leading to the convergence of the adaptation levels for all three training paradigms at the end of the training period (Figure 2B; last five EC trials; one-way ANOVA, $F(2,87) = 0.23$, $p > 0.05$).

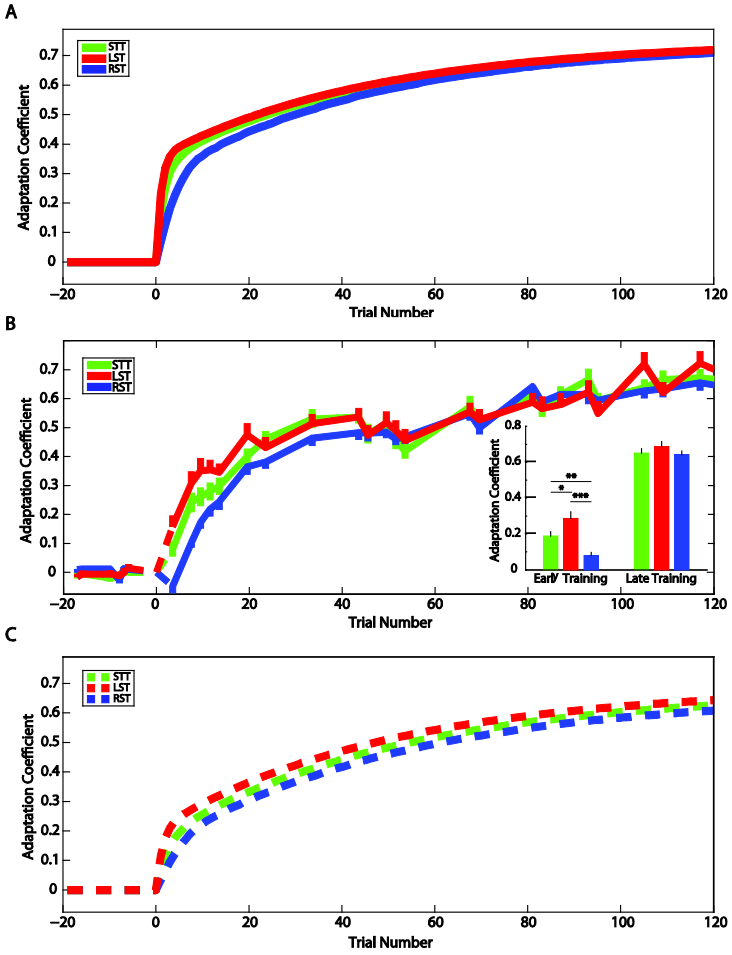


Figure 2. Comparison of Different Training Paradigms. A. Simulated learning curves for STT, LST and RST paradigms according to the MRL hypothesis during extended training – similar to 4D. **B**. Experimental results showing the extended learning curves for the STT, LST and RST training paradigms. **C**. Simulated learning curves for chronic stroke patients with reduced learning rates and higher residual errors (based on parameters from Patton et al., 2006, and Scheidt and Stoeckmann, 2007) trained with the STT, LST and RST paradigms according to the MRL model. The error bars represent standard errors.

Learning simulations for chronic stroke patients

The improvement afforded by the LST paradigm might even be more substantial if MRL-based training is used in chronic stroke patients undergoing neurorehabilitation. Chronic stroke patients are able to adapt to dynamic environments, but display slower learning rates and higher residual errors than healthy controls (Patton et al., 2006; Scheidt and Stoeckmann, 2007). The expected learning behavior for chronic stroke patients was simulated as explained above (see *Modeling and Simulation of Credit Assignment for Force-Field Interference Experiments*) for all three training paradigms according to the motion-referenced learning hypothesis. In the simulation, the peaks of the learning Gaussian functions were decreased (B_s = 0.015; B_f = 0.100) to account for decreased learning rates reported in the literature [4]. The steady state angular error was increased ($e_{\theta,\infty}$ = 12˚) also to account for findings in chronic stroke patients [5]. Our modeling results suggest that the improvement of LST relative to STT is predicted to be greater in magnitude and longer lasting as

shown in Figure 2C. Further studies would be required to determine whether this training paradigm could lead to clinically significant improvements in neurologically impaired subjects.

Previous implementations of motion-referenced learning

Learning rules that incorporate motion-referenced learning have been previously proposed in the literature. An example can be found in the inverse dynamics adaptation law derived in Chapter 5 (Adaptive Control of Manipulators) of Craig (1998) for a manipulator with an "adaptive element". The equation of motion (i.e., the physical dynamics) for the manipulator is given by (Equation 5.1),

$$T = M(\theta)\ddot{\theta} + Q(\theta, \dot{\theta})$$

where M is the mass matrix and Q is a matrix of Coriolis terms. Together these matrices are responsible for the physical dynamics of the manipulator given above. In order to control the manipulator, its physical dynamics must be taken into account. These dynamics aren't known exactly, however, estimates of these dynamics, $\widehat{M}$ and $\widehat{Q}$, could be used to effectively control the manipulator if these estimates are sufficiently accurate. The general approach is to iteratively update the estimates of these dynamics, $\widehat{M}$ and $\widehat{Q}$, so that they gradually approach the true dynamics, M and Q. Thus there are two key parts of this control scheme (i) the control law which generates torque commands which depend on the parameter estimates, $\widehat{M}$ and $\widehat{Q}$, and (ii) the learning rule for iteratively updating (improving) the parameter estimates, $\widehat{M}$ and $\widehat{Q}$. The proposed control law for the manipulator echoes the manipulator's true dynamics (i.e., equations of motion) and is given by (Equation 5.3),

$$T = \widehat{M}(\theta)\ddot{\theta}^* + \widehat{Q}(\theta, \dot{\theta})$$

where $\widehat{M}$ and $\widehat{Q}$ are estimates of M and Q, and $\ddot{\theta}^*$ is the acceleration which characterizes the motion that is currently desired. Craig uses a robust formulation of this desired acceleration given by (Equation 5.4),

$$\ddot{\theta}^* = \ddot{\theta}_D + K_v\dot{E} + K_pE$$

where the currently desired acceleration, $\ddot{\theta}^*$, is equal to the sum of the original desired acceleration, $\ddot{\theta}_D$, and a feedback correction term, $K_v\dot{E} + K_pE$, that prevents the manipulator from drifting too far off course. Specifically, E, is the error which is given by the difference between the desired and actual motion ($E = \theta_D - \theta$). The multidimensional control signal T can be broken down into its individual components (individual torques τ_j) used to control the movement of the manipulator. A key point is that in the

expressions for the individual components (τ_j) two terms can generally be separated from one another: a parameter term which denotes a physical property of the manipulator and a time-varying structural term which involves the current motion state of the manipulator. For the inertia part of the dynamics the static parameter is denoted by $\hat{m}$ and the structural term is denoted by $f(\theta, \ddot{\theta}^*)$. For the Coriolis part of the dynamics the static parameter is denoted by $\hat{q}$ and the structural term is denoted by $g(\theta, \dot{\theta})$. This is shown in Equation 5.9 of Craig (1988),

$$\tau_j = \sum_{i=1}^{a_j} \hat{m}_{ji} f_{ji}(\theta, \ddot{\theta}^*) + \sum_{i=1}^{b_j} \hat{q}_{ji} g_{ji}(\theta, \dot{\theta})$$

If $\hat{m}$ and $\hat{q}$ are merged into an overall vector of parameters, *P*, and *f* and *g* are merged into an overall matrix of structural terms, *W*, a learning rule for the parameter estimates can be identified which is provable stable, meaning that the parameter estimates will definitely converge on the true values of the parameters given sufficient training. This learning rule is given in Equation 5.27:

$$\dot{\hat{P}} = \Gamma W(\theta, \dot{\theta}, \ddot{\theta})^T \hat{M}^{-1}(\theta) E_1$$

This learning rule was derived (and proven to be stable) using Lyapunov theory. Here, Γ is a positive definite diagonal matrix of constants denoting the learning rates for the different system parameters, $W(\theta, \dot{\theta}, \ddot{\theta})$ is a matrix of structural terms which is a concatenation of $f(\theta, \ddot{\theta})$ and $g(\theta, \dot{\theta})$, which in turn depend on the actual motion θ (and not on the desired motion θ_D). Note that $\hat{M}^{-1}(\theta)$ is also a function of the actual motion and not the desired motion. So overall, the update for the parameters which define the inverse model used for control ($\hat{m}$, $\hat{q}$) collectively represented as $\dot{\hat{P}}$ can be partitioned into three terms which are multiplied together: the learning rate (Γ), a credit assignment ($W(\theta, \dot{\theta}, \ddot{\theta})^T \hat{M}^{-1}(\theta)$) and the error ($E_1$). Note that although the control law contains the desired acceleration, the credit assignment, $W(\theta, \dot{\theta}, \ddot{\theta})^T \hat{M}^{-1}(\theta)$, does not. Thus it is purely motion-referenced in this learning rule.

Another example of motion-referenced learning was implemented by Wolpert and Kawato (1998) in a system that combined multiple paired forward-inverse models for motor adaptation [6]. The output of the *i*-th forward model of the system is given by (Equation 5),

$$\hat{x}_{t+1}^i = \phi(w_t^i, x_t, u_t)$$

where w_t^i are the weights of the neural primitives underlying the model, x_t is the true state of the system and u_t is the motor command. After comparing the predicted next state with the actual state, each forward model

can be weighted using a responsibility signal that determines the extent to which each forward model predicted the behavior of the system. For the *i*-th forward model this signal would be (Equation 6),

$$\lambda_t^i = \frac{e^{-|x_t - \hat{x}_t^i|^2/\sigma^2}}{\sum_{j=1}^{n} e^{-|x_t - \hat{x}_t^j|^2/\sigma^2}}$$

where σ is a scaling constant. The combined output of all the forward models is then given by (Equation 7),

$$\hat{x}_{t+1} = \sum_{i=1}^{n} \lambda_t^i \, \phi(w_t^i, x_t, u_t)$$

and after this prediction a gradient descent learning rule is used to update the weights of the neural primitives of each individual model (Equation 8),

$$\Delta w_t^i = \epsilon \frac{d\hat{x}_t^i}{dw_t^i} \lambda_t^i (x_t - \hat{x}_t^i)$$

Note that the gain on the error signal, $x_t - \hat{x}_t^i$, is:

$$\epsilon \frac{d\hat{x}_t^i}{dw_t^i} \lambda_t^i$$

This gain reflects motion-referenced learning. Notice that the motion-referenced adaptation takes place at two levels in the system proposed by Wolpert and Kawato. First, the "responsibility signal", λ_t^i, for each individual forward model depends on actual rather than desired motion (Equation 6), and second, the term $d\hat{x}_t^i/dw_t^i$ depends on actual rather than desired motion [6].

Optimality of motor learning according to PRL and MRL

It is interesting to note that in Donchin et al. (2003), where motor adaptation is modeled with a PRL learning rule, the authors provide a proof in their supplementary materials that PRL-based learning is optimal. Inspection of this proof reveals that what they actually set about to find is how credit assignment should be aligned in order for motor adaptation to produce maximal benefit for subsequent trials to the same target. In other words, this proof investigated what the optimal credit assignment procedure should be for STT and found that PRL maximizes the benefit of motor adaptation for STT. Since PRL is optimal for STT, MRL must be suboptimal for STT (as our simulations predict; see Figure 4B-D in the main manuscript). This suggests that some training procedure other than STT would be optimal for MRL, and our data show that, for a clockwise FF, left-shifted training (LST) is indeed more effective than STT. Effectively, Donchin et al.

(2003) assumed that the credit assignment procedure optimized for performance on STT rather than for stability would be used by the nervous system. Here we show that this is not the case, and as a result, STT produces slower learning than another training procedure (LST).

Previous work on credit assignment during motor adaptation

Although recent studies [7-11] have looked at the contribution of internal and external (body versus world) sources of error during motor adaptation, they have not shed light on how the state-dependent associations [12-14] that underlie motor adaptation are formed. In these studies the authors have categorized the sources of error into two groups, internal (due to sensory discrepancies) and external (due to environmental uncertainty), and have argued that internal errors are of greater relevance for motor adaptation than external errors. Although this body of work helps to address the question of to what extent should adaptation occur in different circumstances, i.e. when sensory information is miscalibrated and/or the learning environment is uncertain, it sidesteps the question of how the error-dependent associations that underlie the adaptation arise, irrespective of the error source, which is the issue that we have addressed.

References

1. Hertz J, Krogh A, Palmer RG (1991) Introduction to the theory of neural computation. Redwood City, Calif.: Addison-Wesley Pub. Co. xxii, 327 p. p.
2. Battiti R (1992) First-order and second-order methods for learning - between steepest descent and Newton's method. Neural Computation 4: 141-166.
3. Smith MA, Gonzalez Castro LN, Monsen CB, Brayanov J. Adaptive changes in arm dynamics are experience-dependent rather than goal-dependent; 2008; Washington, D.C. Society for Neuroscience.
4. Scheidt RA, Stoeckmann T (2007) Reach adaptation and final position control amid environmental uncertainty after stroke. Journal of Neurophysiology 97: 2824-2836.
5. Patton JL, Stoykov ME, Kovic M, Mussa-Ivaldi FA (2006) Evaluation of robotic training forces that either enhance or reduce error in chronic hemiparetic stroke survivors. Experimental Brain Research 168: 368-383.
6. Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Networks 11: 1317-1329.
7. Berniker M, Körding K (2008) Estimating the sources of motor errors for adaptation and generalization. Nature Neuroscience 11: 1454-1461.
8. Ethier V, Zee DS, Shadmehr R (2008) Changes in Control of Saccades during Gain Adaptation. Journal of Neuroscience 28: 13929-13937.
9. Wei KL, Koerding K (2009) Relevance of Error: What Drives Motor Adaptation? Journal of Neurophysiology 101: 655-664.
10. Körding KP, Tenenbaum JB, Shadmehr R (2007) The dynamics of memory as a consequence of optimal adaptation to a changing body. Nature Neuroscience 10: 779-786.
11. Kluzik J, Diedrichsen J, Shadmehr R, Bastian AJ (2008) Reach adaptation: What determines whether we learn an internal model of the tool or adapt the model of our arm? Journal of Neurophysiology 100: 1455-1464.
12. Conditt MA, Gandolfo F, MussaIvaldi FA (1997) The motor system does not learn the dynamics of the arm by rote memorization of past experience. Journal of Neurophysiology 78: 554-560.
13. Conditt MA, Mussa-Ivaldi FA (1999) Central representation of time during motor learning. Proceedings of the National Academy of Sciences of the United States of America 96: 11625-11630.
14. Sing GC, Joiner WM, Nanayakkara T, Brayanov JB, Smith MA (2009) Primitives for Motor Adaptation Reflect Correlated Neural Tuning to Position and Velocity. Neuron 64: 575-589.