7 Learning faster

The title of this chapter sounds a little like a product that you might find advertised on TV in the wee hours of the night. "We will not only help you learn (your favorite task here) faster, but if you order now, we will also send you this amazing new ..." It sounds like the promises of a snake oil salesman. But as we will see here, if we view learning in biological systems as a Bayesian state estimation process, then the mathematical framework that we developed provides us with some useful ideas on how to encourage the process of learning.

7.1 Increased sensitivity to prediction errors

One of the properties of the state estimation framework is that the sensitivity to error (the Kalman gain) depends on the relative size of the state uncertainty with respect to the measurement uncertainty. Said in another way, the sensitivity to error depends on the ratio of the confidence in the state that we are trying to estimate with respect to the confidence of our measurements. For example, suppose that we are attempting to estimate state x, which is governed by the following generative model:

$$\begin{aligned} x^{(n+1)} &= ax^{(n)} + \varepsilon_x^{(n)} \qquad \varepsilon_x \sim N\left(0,\sigma_x\right) \\ y^{(n)} &= x^{(n)} + \varepsilon_y^{(n)} \qquad \varepsilon_y \sim N\left(0,\sigma_y\right) \end{aligned} \tag{7.1}$$

The sensitivity to prediction error, written as our Kalman gain $k^{(n)}$, will depend on the ratio of the prior state uncertainty $p^{(n|n-1)}$ with respect to measurement uncertainty $p^{(n|n-1)} + \sigma_y$:

$$k^{(n)} = \frac{p^{(n|n-1)}}{p^{(n|n-1)} + \sigma_y} \tag{7.2}$$

When the learner observes a prediction error but is uncertain about the sensory measurement (noise σ_y is large), he will have a relatively small learning rate. When the learner observes the same prediction error but is uncertain about his own predictions (the state uncertainty p is large), he will have a relatively large learning rate. Therefore, the state estimation framework predicts that through simple techniques, you should be able to up-regulate or down-regulate rates of adaptation.

Let us do some simulations to explore how sensory and state noise of the underlying system affects uncertainty and sensitivity to error (i.e., learning rate). Fig. 7.1 shows that if the variance of the state noise σ_x were to increase (reflecting our uncertainty about how the underlying state changes with time), then state uncertainty p increases, resulting in an increase in the sensitivity to error k. Fig. 7.1 also shows that if the variance of the measurement noise σ_y were to increase (reflecting our uncertainty about our observations), our state uncertainty p increases, but this increase is generally dwarfed by the increase in the measurement noise. As a result, the sensitivity to error k decreases. Therefore, if one could alter the learner's uncertainty about the state that is being estimated, or her confidence about the observation, one might be able to change the learner's sensitivity to error. The learning rate should be adaptable to the conditions of the experiment.

Johannes Burge, Marc Ernst, and Martin Banks (2008) tested this idea in a simple reaching task. They had people hold a pen in hand and reach to a visual target. Visual feedback was withheld until the end of the reach (Fig. 7.2). The form of the visual feedback was a blurry collection of pixels in which intensities of the pixels were described by a Gaussian function. To manipulate the subject's measurement uncertainty, they changed the variance of this Gaussian. They considered two conditions, a low measurement uncertainty condition in which the Gaussian's variance was $\sigma_y = 4$ deg, and a high uncertainty condition in which $\sigma_y = 24$ deg. Normally, the center of the Gaussian corresponded to the hand position at end of the reach. However, to produce adaptation, a bias was added to this relationship. In summary, the feedback $y^{(n)}$ on trial *n* was a blurry Gaussian centered at $h^{(n)} + r$, where h is hand position at end of trial *n*, and r is the visual perturbation, with variance σ_y. The prediction was that with the more uncertain feedback (larger σ_y), rates of learning would be reduced. Indeed, Burge et al. (2008) found that the adaptation rate was slower when σ_y was larger (Fig. 7.2A).

You might be a little skeptical about this result, because with a really blurry feedback, there may be little motivation to adapt. For example, in Fig. 7.2A the plot for σ_y may never reach zero, so it is not clear that we are really looking at a reduced sensitivity to error. We will return to this problem shortly, but for now consider a more interesting question: how do we increase the learning rate? The answer would have important practical implications because rehabilitation of people who have suffered brain injury relies on learning to recover lost function, something akin to adaptation. Finding ways to improve learning rates might result in better ways to train patients during rehabilitation.

The estimation framework predicts that if one could make the learner uncertain about the state of the perturbation, then the sensitivity to error will likely increase. Burge et al. (2008) approached this problem by adding trial to trial variability to the perturbation, i.e., by making the perturbation $r^{(n)}$ act like a random walk about a mean value r_0:

$$\begin{aligned} x^{(n)} &= x^{(n-1)} + \varepsilon_x \qquad & \varepsilon_x \sim N\left(0, \sigma_x\right) \\ r^{(n)} &= r_0 + x^{(n)} \end{aligned} \tag{7.3}$$

They then assumed that the subjects would be estimating this perturbation using a generative model of the usual form:

$$\begin{aligned} r^{(n)} &= r^{(n-1)} + \varepsilon_r \qquad & \varepsilon_r \sim N\left(0, \sigma_r\right) \\ y^{(n)} &= h^{(n)} + r^{(n)} + \varepsilon_y \qquad & \varepsilon_y \sim N\left(0, \sigma_y\right) \end{aligned} \tag{7.4}$$

Their experimental results are plotted in Fig. 7.2B. When the perturbation was more variable from trial to trial, the learning was somewhat faster. For a given feedback blurriness (for example, $\sigma_y = 24$ deg), a larger variance in the random walk of the perturbation (σ_r) produced a more rapid adaptation.

Kunlin Wei and Konrad Kording (2008) also tested this idea of altering learning rates, but they approached it in a way that did not rely on fitting exponential functions to the data. Rather, their approach relied on trial-to-trial measures of sensitivity to error. They asked volunteers to move their hand through a visual target (Fig. 7.3) and provided the volunteers with visual feedback as the hand crossed the target area. This visual feedback represented the hand position, plus a random perturbation. To setup our generative model for this problem, suppose that on trial *n*, at the end of the reach the hand is at $h^{(n)}$ and the imposed perturbation is $x^{(n)}$. The computer displays its feedback at $y^{(n)}$. However, for some blocks the visual feedback is a single dot at $y^{(n)}$, while for other blocks it is a distribution of dots centered at $y^{(n)}$. For example, the feedback can be five dots drawn from a small variance distribution, or five dots drawn for a large variance distribution. The larger this variance, the larger the variance of the measurement noise σ_y in Eq. (7.5). Our generative model takes the form:

$$\begin{aligned} x^{(n+1)} &= x^{(n)} + \varepsilon_x & \varepsilon_x &\sim N\left(0, \sigma_x\right) \\ y^{(n)} &= h^{(n)} + x^{(n)} + \varepsilon_y & \varepsilon_y &\sim N\left(0, \sigma_y\right) \end{aligned} \tag{7.5}$$

Notice that here, the perturbation is zero mean and there is no consistent bias in the movements. Therefore, it is not possible to plot a learning curve because in fact, there is nothing to learn. However, it is still possible to measure sensitivity to error. On trial *n*, the computer instructs the subject to move the hand to target location $y_t^{(n)}$. To do so, the subject predicts the perturbation on this trial $\hat{x}^{(n|n-1)}$ and moves the hand to cancel that perturbation:

$$h^{(n)} = y_t^{(n)} - \hat{x}^{\left(n|n-1\right)} \tag{7.6}$$

The computer provides feedback to the subject by displaying a cursor at position $y^{(n)}$. The subject observes an error between the feedback $y^{(n)}$ and the predicted feedback $h^{(n)} + \hat{x}^{\left(n|n-1\right)}$ and updates his estimate of the perturbation:

$$\hat{x}^{\left(n|n\right)} = \hat{x}^{\left(n|n-1\right)} + k^{(n)}\left(y^{(n)} - h^{(n)} - \hat{x}^{\left(n|n-1\right)}\right) \tag{7.7}$$

Wei and Kording tested the idea that increasing observation noise σ_y (uncertainty about the measurement) would reduce the learning rate $k^{(n)}$. For example, in Fig. 7.3A we see that when the vertical perturbation on trial n was zero, vertical position of the hand on the trial $n+1$ had a zero mean. This means that on average, subjects did not alter their movements on trial $n+1$ if on trial n their movement was not perturbed. Once a -2cm visual perturbation was imposed on trial n, on trial $n+1$ the distribution of hand positions had a slightly positive mean. That is, the brain responded to the errors in trial n by changing behavior on the next trial (to compensate for the -2cm perturbation, on trial $n+1$ the subject moved their hand to +0.4). However, we are particularly interested in the sensitivity to error, i.e., the learning rate $k^{(n)}$. The sensitivity to error is the slope of the line in Fig. (7.3A). To see this, consider that hand position on trial $n+1$ is simply

$$\begin{aligned} h^{(n+1)} &= y_t^{(n+1)} - \hat{x}^{\left(n+1|n\right)} \\ &= y_t^{(n+1)} - \hat{x}^{\left(n|n-1\right)} - k^{(n)}\left(y^{(n)} - h^{(n)} - \hat{x}^{\left(n|n-1\right)}\right) \\ &= y_t^{(n+1)} + h^{(n)} - y_t^{(n)} - k^{(n)}\left(y^{(n)} - y_t^{(n)}\right) \\ h^{(n+1)} - h^{(n)} &= -k^{(n)}\left(y^{(n)} - y_t^{(n)}\right) \end{aligned} \tag{7.8}$$

The steeper this slope becomes (more negative), the larger the sensitivity to error $k^{(n)}$. Wei and Kording noted that as the subject's measurement uncertainty increased (single dot, five dots with a small distribution, five dots with a large distribution), sensitivity to error became smaller (the slope became more shallow, Fig. 7.3C). Therefore, when the feedback that subjects received about their actions became more uncertain and noisy, sensitivity to prediction errors declined and people learned less from their errors.

The estimation framework's key prediction is that the learners will exhibit increased sensitivity to error when they are made uncertain about their own predictions. Wei and Kording attempted to increase this prior uncertainty in two ways. First, they had their subjects make a minute of reaching movements without any visual feedback. They argued that in effect, this increased the magnitude of the noise variance σ_x in Eq. (7.5). In the subsequent test period, they provided visual feedback regarding end of the reach but on random trials they added a disturbance. They measured the change in performance on trial $n+1$ as a function of the error on trial n. Indeed, they found a slightly larger sensitivity to error than in a control condition in which the prior movements had feedback (Fig. 7.3B). They next tried a different way to increase the subject's uncertainty about their state: they had the subjects sit quietly for a minute in darkness. This time passage can be argued to increase one's uncertainty about the state of one's hand as it removes both visual feedback as well as velocity dependent proprioceptive feedback that occurs when one makes movements. Indeed, the sensitivity to error in the subsequent test trials was even larger than in the control condition (Fig. 7.3B). [However, the experiment may also need another control: having subjects sitting still with full vision of the hand. The effect observed by Wei and Kording could reasonably be attributed to the period of inactivity reducing the reliability of motor commands for some time immediately following the pause.]

In summary, if we view learning in a Bayesian framework, we can make certain predictions about how to speed up or slow down the process of learning. Speeding up learning can occur if the learner becomes more sensitive to prediction errors, which can be achieved by making the learner more uncertain about her own predictions. Slowing down learning can occur if the learner's sensory measurements become less reliable.

7.2 Modulation of forgetting rates

In theory, another way to manipulate sensitivity to error is via our belief regarding how the state that we are trying to estimate changes from trial to trial. Suppose our generative model takes the form:

$$\begin{aligned} x^{(n+1)} &= a x^{(n)} + \varepsilon_x \qquad & \varepsilon_x \sim N\left(0, \sigma_x\right) \\ y^{(n)} &= x^{(n)} + \varepsilon_y \qquad & \varepsilon_y \sim N\left(0, \sigma_y\right) \end{aligned} \tag{7.9}$$

If the state x that we are trying to estimate is basically constant from one trial to the next, i.e., $a \approx 1$, then what we learn in one trial should be retained for the next trial. That is, the posterior estimate $\hat{x}^{(n|n)}$ in trial *n* should be 'retained' and become the prior estimate $\hat{x}^{(n+1|n)}$ in trial *n*+1:

$$\hat{x}^{(n+1|n)} = a\hat{x}^{(n|n)} \tag{7.10}$$

However, if we believe that the state x is uncorrelated from one trial to the next, i.e., $a \approx 0$, then what we learn in one trial should be forgotten by the next trial. In a sense, if we believe that $a \approx 0$, then the state that we are trying to estimate is pure noise and whatever prediction errors we observe on a given trial should be forgotten as we form our prior for the next trial.

Do people reduce their sensitivity to prediction error when the perturbations are pure noise? Do they increase their sensitivity when the perturbations are correlated? To answer this question, we need a way to estimate the subject's sensitivity to error. There are, in principle, two things that we can measure on each trial: the subject's prediction $\hat{y}^{(n)}$, and the subject's prediction error $y^{(n)} - \hat{y}^{(n)}$. If we assume that the $\hat{y}^{(n)}$ reflects the subject's prior, then the change in the subject's predictions from one trial to the next is a function of the prediction error, with a sensitivity that is proportional to the learning rate k:

$$\begin{aligned} \hat{y}^{(n)} &= \hat{x}^{(n|n-1)} \\ \hat{x}^{(n+1|n)} &= a\hat{x}^{(n|n-1)} + ak^{(n)}\left(y^{(n)} - \hat{y}^{(n)}\right) \\ \hat{y}^{(n+1)} &= \hat{x}^{(n+1|n)} \\ \hat{y}^{(n+1)} - \hat{y}^{(n)} &= (1-a)\,\hat{x}^{(n|n-1)} + ak^{(n)}\left(y^{(n)} - \hat{y}^{(n)}\right) \end{aligned} \tag{7.11}$$

If the subject has learned that $a \approx 0$, then the sensitivity to error $ak^{(n)}$ should also go to zero. In effect, if the learner is optimal, then she will stop learning when the perturbations are simply noise with no correlations from trial to trial.

Maurice Smith and Shadmehr (2004) tested this idea by having people reach while holding a robotic arm that produced force perturbations. The force perturbation $f^{(n)}$ on trial *n* was described by a discrete function in which the correlation between one trial to the next was controlled within each block of training by parameter *a*:

$$f^{(n+1)} = af^{(n)} + \varepsilon \qquad \varepsilon \sim N\left(0, p^2\right) \tag{7.12}$$

In Fig. 7.4A we have examples of the perturbations as a function of trial. When $a = +0.9$, the perturbation on trial *n* were positively correlated with trial *n*+1. This correlation disappeared when $a = 0$, and became negative when $a = -0.9$. The authors estimated the change in the motor output $\hat{y}^{(n+1)} - \hat{y}^{(n)}$ as a function of the error in the previous trial (similar to Fig. 7.3A). This sensitivity to error is plotted in Fig. 7.4B. When $a = 0$, the sensitivity to error was around 0.19, meaning that subjects compensated for about 19% of their prediction errors. When $a = +0.9$, sensitivity to error was about 0.22. When $a = -0.9$, the sensitivity to error was around 0.06. Therefore, the apparent learning rate was modulated by the statistics of the perturbation.

However, this last experiment unmasks a troubling feature of our attempt to couch biological learning in the framework of state estimation. Note that when $a = -0.9$, the theoretical prediction is that the Kalman gain will have the opposite sign as to when $a = +0.9$, but equal magnitude. That is, negatively correlated perturbations are just as predictable (and therefore learnable) as positively corrected ones. However, what we see in Fig. 7.4B is that people tend to shut down their learning system when they are presented with negatively correlated perturbations. What is happening here?

The theoretical problem that we are facing is that in optimal estimation (as in the Kalman framework), we need to have knowledge of the generative model, i.e., the model that describes how the hidden states that we are trying to estimate are related to the data that we are observing. In other words, we need to have an internal model of the data and its relationship to our actions. But in reality, no one gives us this generative model. We do not have the very thing that we need in order to do optimal estimation. For example, no one provides the learner in Fig. 7.4, or in any of our examples thus far, the generative model that is producing the perturbations.

One way to view the results of the experiment in Fig. 7.4 is to assume that the brain is estimating both the parameter *a* of the generative model, and the state of the perturbation $f^{(n)}$. If indeed the brain was attempting to estimate the parameter *a*, then what we should see is a modulation of the forgetting rates from trial to trial. For example, suppose that the perturbations are governed by Eq. (7.12) in which $a \approx 1$, and suppose that you as a learner have an accurate estimate of this parameter (because of prior learning or experience with the perturbations). Now suppose that through some artificial means during ten trials in a row whatever predictions you make, you observe zero prediction error:

$$y^{(n)} - \hat{y}^{(n)} = 0 \tag{7.13}$$

Let us call these *error clamp* trials because the prediction error is clamped to zero. If in your generative model $a \approx 1$, then you should exhibit little or no change in your motor output during these ten error-clamp trials. That is, you should show very little forgetting. On the other hand, if the perturbations are governed by $a \approx 0$, and you have an accurate internal model of this generative process, then your motor output should show complete forgetting during these same error-clamp trials.

This way of thinking allows us to arrive at an interesting idea: decay of our memories (forgetting rate) should be – at least in part - a reflection of our internal model of the environment in which we acquired that memory. We should retain what we learned in environments that changed slowly, but forget what we learned in environments that changed rapidly. In the motor learning literature there are a number of reports demonstrating that the brain retains a recently acquired memory better if that memory was acquired as a consequence of a gradual rather than a sudden change in the environment (Kagerer et al. 1997; Klassen et al. 2005; Michel et al. 2007). For example, adaptation to prism glasses can produce long-lasting motor memories if the visual distortion is introduced gradually (Hatada et al. 2006), but tends to decay quickly if the distortion is introduced suddenly (Martin et al. 1996). Adaptation to gradual perturbations of gait produces longer lasting after-effects than adaptation to rapidly changing perturbations (Reisman et al. 2007). Why should the rate of change in the environment influence decay rates of motor memory?

In principle, if an environment changes slowly it might lead to the implicit belief that the changes are likely to be lasting, which in turn should influence the retention properties of the memory. On the other hand, if the environment changes suddenly, one should still adapt to the changes, but the speed of the change might imply that whatever is learned should be rapidly forgotten because the changes in the environment are unlikely to be sustained. This view implies that the long-term statistics of the experience, which presumably leads to formation of a more accurate generative model, i.e., an estimate of parameter a in Eq. (7.12), should have a strong effect on the 'forgetting' rate associate with the state estimate.

Vincent Huang and Shadmehr (2009) found that when people reached in an environment in which perturbations changed rapidly, motor output adapted rapidly but then decayed rapidly in a post-adaptation period in which performance errors were eliminated (error-clamp trials). On the other hand, in an environment that changed gradually motor output adapted gradually but then decayed gradually. They used this behavioral assay to test whether prior experience with a different rate of environmental change would affect forgetting rates of the motor output. The idea was that through experience with a sequence of perturbations, the brain builds a generative model of those perturbations. Part of this generative model is the parameter a in Eq. (7.12). If this parameter indicates that the perturbations are changing slowly, then one should remember what one has learned in the subsequent error clamp trials.

To test this idea, volunteers were provided with a prior experience with a perturbation sequence, and then tested on a second sequence. For example, Group 1 initially trained with a perturbation sequence that was step-like, while Group 3 trained with a gradual perturbation sequence (Fig. 5A, 'learn' period). After this brief period of training, a very long period of washout trials were presented for which there were no perturbations. Following the washout, the subjects once again were exposed to a sequence of perturbations. In this re-learning block, the perturbations for both Groups 1 and 3 were gradual. Retention was then assayed in the following error-clamp block.

Huang and Shadmehr found that the initial training in the learning block biased the forgetting rates in the re-learning block. People who had initially observed a rapidly changing sequence of perturbations subsequently had rapid forgetting after the gradual re-learning (Fig. 5B, Group 1). People who had initially observed a gradually changing sequence of perturbations subsequently had slow forgetting after the same gradual re-learning (Fig. 5B, Group 3). Similarly, subjects that had prior experience with a rapidly changing perturbation (learn block, Group 4) exhibited more rapid forgetting in the error-clamp period that followed re-learning (Group 4, Fig. 5C) as compared to subjects whose prior experience was with a gradually changing perturbation (Groups 2 and 5, Fig. 5C). Therefore, prior experience with a rapid change in the environment biased retention of memories acquired later in response to a gradual change in the same environment, increasing forgetting rates. Similarly, prior experience in an environment that changed gradually enhanced retention in response to a rapid change in that same environment.

These results suggest that as our brain is presented with a prediction error, it tries to learn a generative model. This generative model is a description of how the state of the perturbation is related to the observations or measurements. It then uses this generative model to optimally estimate the state of the perturbation.

Summary

If we view learning in a Bayesian state estimation framework, we can make certain predictions about how to speed up or slow down the process of learning. Learning will speed up if the learner becomes more sensitive to prediction errors, which can be achieved by making the learner more uncertain about her own predictions. Learning will slow down if the learner's sensory measurements become less reliable.

To make a prediction, the learner relies on her prior estimate. The prediction error allows the learner to form a posterior estimate. However, the prior belief on the next trial depends on the generative model and how it dictates trial-to-trial change in the hidden states. In effect, the generative model's trial-to-trial change is a forgetting rate. Presumably, biological learning is not only a process of state estimation, but also a process in which the brain learns the structure of the generative model, i.e., how the hidden states change from one trial to the next. This can account for the fact that retention, as measured by decay of motor output in error-clamp trials, is slower after perturbations that change gradually vs. those that change rapidly.

Figure 7.1. Learning rates depend on uncertainties of the learner. The left figure shows the Kalman gain and the right figure shows the uncertainty in the state estimate for the generative model of Eq. 7.1. As the noise in the state update equation σ_x increase, the uncertainty in the state estimate p increases, and the sensitivity to error k increases. As the noise in the measurement σ_y increases, the uncertainty in the state estimate p increases, but the sensitivity to error k decreases.

Figure 7.2. Sensitivity to prediction error can be increased when the subject is made more uncertain about the state of the perturbation. Volunteers reached to a target while a visuomotor perturbation altered their sensory feedback. **A**) Performance when the measurement uncertainty was small ($\sigma_y = 4$) or large ($\sigma_y = 24$). **B**) For a given measurement uncertainty, increased state uncertainty (increasing σ_x) produced a faster learning rate. (From Burge et al. (2008) with permission.)

Figure 7.3. Altering the sensitivity to error through manipulation of feedback. Subjects reached to a target and feedback was provided regarding their endpoint. **A**) Hand deviations on trial $k+1$ as a function of perturbation on trial k. A positive perturbation on trial k produces a negative compensation on trial $k+1$. The slope of the line is sensitivity to error. The more negative this slope, the greater the sensitivity to error in trial k. **B**) Sensitivity to error increased after a period in which no feedback was provided after a movement, and after a period of sensory deprivation in which no movements or feedback were provided (sitting in darkness). **C**) When feedback was blurred, adaptation slope became less negative, indicating a reduced sensitivity to error. (From Wei and Kording (2008), with permission.)

Figure 7.4. Altering the sensitivity to error through manipulation of the state update equation. Subjects held the handle of a robotic arm and made reaching movements to a target. The robot perturbed the hand with a force field. **A**) Force perturbations are plotted as a function of trial number. These forces were generated via Eq. (7.12). **B**) As the auto-correlation parameter a increased, sensitivity to error increased. (From Smith and Shadmehr (2004) with permission.)

Figure 7.5. Retention of motor memory as a reflection of the statistics of the perturbations. **A**) On each trial, subjects reached to a target and were perturbed by a force that either increased

abruptly (as a function of trial) or gradually. The gray region indicates error-clamp trials, i.e., trials in which errors in movement were artificially clamped to zero, enabling measurement of motor output while minimizing error dependent learning. In block 1, no perturbations were applied. In block 2, learning was followed by a measure of retention in error-clamp trials. In block 3, no perturbations were applied (termed washout). In block 4, re-learning took place followed by a measure of retention in error-clamp trials. **B**) In block 3 (re-learning), groups 1 and 3 both experienced a gradually increasing force. However, the prior experience of group 3 was with the same gradually increasing force, while the prior experience of group 1 was with a rapidly increasing force. In error-clamp trials, force output of group 3 gradually declined, while output of group 1 rapidly declined. **C**) In block 3, groups 2, 4 and 5 experienced a rapidly increasing force. However, the prior experience of both groups 2 and 5 was with a gradually increasing force. In error-clamp trials, force output of groups 2 and 5 gradually declined. (From Huang and Shadmehr (2009) with permission.)

Reference List

Burge J, Ernst MO, Banks MS (2008) The statistical determinants of adaptation rate in human reaching. J Vis 8:20-19.

Huang VS, Shadmehr R (2009) Persistence of motor memories reflects statistics of the learning event. J Neurophysiol 102:931-940.

Smith M, Shadmehr R (2004) Modulation of the rate of error-dependent learning by the statistical properties of the task. Advances in Computational Motor Control 3.

Wei K, Kording K (2008) Uncertainty in state estimate and feedback determines the rate of motor adaptation. Advances in Computational Motor Control 7.