Three timescales in prism adaptation

Masato Inoue¹, Motoaki Uchimura¹,², Ayaka Karibe¹, Jacinta O’Shea³,⁴, Yves Rossetti⁴,⁵, and Shigeru Kitazawa¹,²,⁶,⁷

¹Department of Neurophysiology, Graduate School of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo, Tokyo 113-8421, Japan
²Dynamic Brain Network Laboratory, Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka, 565-0871, Japan
⁴ImpAct, Centre des Neurosciences de Lyon, Inserm UMR-S 1028 ; CNRS UMR 5292; Université Lyon 1, 16 Avenue Lépine, 69676 Bron, France
⁵Plate-forme Mouvement et Handicap, Hospices Civils de Lyon; Centre de Recherche en Neurosciences de Lyon, Service de Rééducation Neurologique ; Hôpital Henry Gabrielle, 20 Route de Vourles, 69230 Saint Genis Laval, France
⁶Department of Brain Physiology, Graduate School of Medicine, Osaka University, Suita, Osaka, 565-0871 Japan
⁷Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, and Osaka University, Suita, Osaka, 565-0871, Japan

Running head Three timescales in prism adaptation

Corresponding author: Shigeru Kitazawa
kitazawa@fbs.osaka-u.ac.jp
Tel +81-6-6879-4431
Fax +81-6-6879-4437

Copyright © 2014 by the American Physiological Society.
Abstract

It has been proposed that motor adaptation depends on at least two learning systems, one that learns fast but with poor retention, and another that learns slowly but with better retention (Smith et al. 2006). This two-state model has been shown to account for a range of behaviour in the force field adaptation task. In the present study, we examined whether such a two-state model could also account for behaviour arising from adaptation to a prismatic displacement of the visual field. We first confirmed that an "adaptation rebound", a critical prediction of the two-state model, occurred when visual feedback was deprived after an adaptation-extinction episode. We then examined the speed of decay of the prism after-effect (without any visual feedback) after repetitions of 30, 150, and 500 trials of prism exposure. The speed of decay decreased with the number of exposure trials, a phenomenon that was best explained by assuming an "ultra-slow" system, in addition to the fast and slow systems. Finally we compared retention of after-effects 24 hours after 150 or 500 trials of exposure: retention was significantly greater after 500 than 150 trials. This difference in retention could not be explained by the two-state model but was well explained by the three-state model as arising from the difference in the amount of adaptation of the "ultra-slow process". These results suggest that there are not only fast and slow systems but also an "ultra-slow" learning system in prism adaptation that is activated by prolonged prism exposure of 150-500 trials.

Keywords Prism adaptation, Reaching, Motor learning, two-state model, three-state model
Introduction

It has been proposed that motor adaptation depends on at least two learning systems with different speeds, one that learns fast but has poor retention, and another that learns slowly but has better retention (Smith et al. 2006). The idea was proposed by Smith et al. (2006) to account for data obtained during adaptation of saccadic eye movements (Kojima et al. 2004) and was further tested in adaptation of reaching movements in response to a force field perturbation (Smith et al. 2006) and visuomotor rotations (Lee and Schweighofer 2009; Shmuelof et al. 2012; Zarahn et al. 2008). The model was successful in explaining many phenomena, such as savings, adaptation rebound (Smith et al. 2006), spontaneous recovery (Ethier et al. 2008), memory decay (Shmuelof et al. 2012; Vaswani and Shadmehr 2013), and generalization (Tanaka et al. 2012).

Joiner and Smith (2008), using a force field adaptation paradigm, further suggested that the magnitude of retention at 24 hours could be predicted by the amount by which ‘the slow system’ had adapted: retention at 24 hours reflected ~65% of the magnitude of adaptation of the slow process after adaptation with 11, 30, 103, and 160 trials. It was noteworthy that the amount of retention at 24 hours approached a plateau as the number of adaptation trials increased to 103 and 160 trials (Fig. 3 in Joiner and Smith, 2008). By extrapolating these results to a still larger number of trials, it is reasonable to predict that the magnitude of adaptation of the slow process (and the amount of retention at 24 hours) in the force field task should converge to an asymptote after an even larger number of adaptation trials.
However, Yin and Kitazawa (2001) demonstrated in monkeys that 250 trials of prism adaptation (i.e. adaptation of reaching movements in response to an optical displacement of the visual field) did not result in significant retention of prism after-effects at 24 hours. By contrast, 500 trials of adaptation produced significant retention, which amounted to ~70% of the visual displacement (Fig. 2 in Yin and Kitazawa, 2001). This difference in 24 hour retention rates with prism adaptation contrasts with those reported above for force field adaptation. This observed difference appears to challenge the standard two-state model, which posits a single slow state to account for retention. We, therefore, set out to test the hypothesis that there is a third system for prism adaptation that learns much more slowly than ‘the slow system’, and that this would be engaged and grow significantly over the period between 250 and 500 exposure trials.

In the present study, we tested the three-state hypothesis for prism adaptation, by testing three inter-related predictions from that model, namely: 1) the process of prism adaptation can be approximated by the two-state model up to 250 trials of adaptation, but 2) the third component, an ultra-slow system, becomes essential to explain data after 500 adaptation trials, and 3) long-term retention at 24 hours reflects the amount of adaptation in the “ultra-slow” system, not in the slow system.

To address the first issue, we first examined whether a so-called “adaptation rebound”, a critical prediction of the two-state model (Smith et al. 2006), could be observed following 250 trials of prism adaptation (Experiment 1). In the work by Smith and colleagues, participants adapted to a force field perturbation (e.g., a counter-clockwise force field), and errors induced by the perturbation decreased with learning on a trial-by-trial basis. After 200 trials of adapting to the first force field, a second field
with the opposite perturbation direction was introduced (e.g., a clockwise force field) for 15 trials. When the second perturbation was removed, after-effects from the second force field disappeared quickly, whereas after-effects from the first force field reappeared as an “adaptation rebound”. We designed a similar protocol with prism adaptation, and first examined whether the phenomenon of “adaptation rebound” would occur in this task, and whether the process can be approximated by the two-state model. Three- and four-state models were also fitted to the data for comparison.

To address the second issue, we designed Experiment 2. Naïve participants adapted their reaching movements while wearing prisms in a block of 30, 150 or 500 trials. As in Experiment 1, the rate of memory decay was assayed by trials in which visual feedback was removed. We chose three adaptation doses (30, 150 or 500 trials), with the expectation that data involving 500 adaptation trials would be better explained by assuming the third ultra-slow component.

To test the third prediction, we examined long-term retention by assessing the magnitude of prism after-effects after a delay of 24 hours for each of the 150- and 500-trials adaptation conditions. We critically examined whether the magnitude of the after-effect correlated better with the amount of adaptation in the slow system or in the “ultra-slow” system.
Materials and methods

Participants

Seven male volunteers (aged 27-47) participated in Experiment 1. Twenty-four volunteers (10 female, aged 20-47) participated in Experiment 2. All participants were right-handed, and had normal or corrected-to-normal visual acuity and had no significant neurological history. The study received approval from the institutional ethics committee, and all participants gave informed written consent before the experiments.

Apparatus and task procedure

The participant was seated facing a tangent 14-inch CRT screen, 400 mm from the eyes, with the head restrained by a chin rest. The participant wore a pair of spectacles with liquid-crystal shutters (PLATO, Translucent Technologies) and viewed the CRT screen through opaque tubes that restricted the field of view. The shutters could change from opaque white to transparent by means of a control signal from a personal computer (Dell, Precision 370). A trial started when the participant pressed a button with the index finger. The button was positioned 300 mm below and 100 mm ahead of the eyes in the mid-sagittal plane. When the participant pressed the button for 1.5 sec, a target (5 mm diameter circle), superimposed on a cross, 15 mm × 15 mm) appeared in a random location within a 40mm square in the centre of the visual field. The participant was required to release the button within 300 ms after the appearance of the target, to touch the screen within 300 ms after the release, and to hold the final position of the hand for
1 second until a 1 second beep instructed the participant to return the hand to the starting position.

Vision of the hand and the arm was blocked at the release of the button by the liquid-crystal shutters to eliminate visual feedback during the movement. In trials with visual feedback (visual-feedback trials), the shutters were opened again at the touch for 300 ms to allow participants to see the target and the final static position of the hand. In trials without visual feedback (no-visual-feedback trials), the shutters were kept closed until the next trial, so that participants could not see the final static position of the hand and the target.

Whenever the visual field was displaced by wedge prisms placed within the opaque tubes, the CRT screen was also displaced (rotated) around the vertical axis that passed the middle point of the participant’s eye to compensate for the visual displacement of the screen caused by the prisms: it was displaced by the amount of refraction (8.5 degree). Because the distance between the screen and the eyes was 400 mm, the screen was displaced 50 mm along an arc of 400 mm radius. This procedure kept the visual image of the screen in the exposure period almost identical to that in the pre-test period: the participant viewed the tangent CRT screen in the same straight ahead direction as in the first period (Kitazawa et al. 1995).

**Experimental Design**

**Experiment 1**

The prism adaptation task consisted of four phases: 1) pre-test (60 trials); 2) the first exposure (Exposure 1, 200 trials), 3) the second exposure (Exposure 2, 15 trials),
and 4) post-test (50 trials). The pre- and post-test blocks were performed without visual
displacement. During the first exposure block the visual field was displaced by 60 mm
(8.5 degrees, 15 diopters) to the left, and during the second exposure block it was
displaced by 60 mm to the right.

In the pre-test block, visual feedback was given in the first and the last 15 trials
(trials 1-15 and 46-60), but was deprived for the intervening 30 trials (trials 16-45). In
the first exposure block, visual feedback was given during the first 30 trials and
deprived during the next 10 trials; and this alternating cycle (30 visual feedback
followed by 10 no-visual feedback trials, termed as an epoch thereafter) was repeated 5
times, yielding 200 trials in sum. This manipulation, interleaving blocks of trials with
versus without visual feedback throughout the different phases of the experiment, was
designed to track changes in the magnitude and stability of learning and prism after-
effects over time. By alternating between blocks of prism exposure versus after-effect
assessment, we expected to observe a ‘zig-zag’ change in the sign and magnitude of
reaching error, which would be beneficial for estimating adaptation and decay
parameters for the computational model, formulated later in Eq. (1). The length of the
first exposure block (a total of 150 trials with the initial visual displacement excluding
50 trials without visual feedback) was chosen to ensure a sufficient number of trials to
trigger adaptation of the hypothesized ‘slow’ system. During the second exposure block,
the visual field was displaced by 60 mm in the direction opposite to that during the first
exposure block, and visual feedback was given in all 15 trials. This rather short block
length was chosen after Smith et al. (2006) as sufficient to ensure the fast system would
adapt to the second displacement, but the slow system would still retain adaptation to
the initial displacement. In the post-test, participants performed 50 trials without visual feedback. Critically, we expected to observe “adaptation rebound” during the post-test, namely a re-emergence of adaptation to the initial visual displacement.

Experiment 2

On Day 1, participants performed blocks of: pre-test (30 trials), adaptation (30, 150, or 500 trials), and post-test (50, 100, or 100 trials). In the pre-test visual feedback was present in the first 10 trials, deprived in the next 10 trials, and present again in the last 10 trials.

During the adaptation block, the visual field was displaced by 50 mm to the left (n = 12) or to the right (n = 12) by the prisms (7.1 degrees, 12.5 diopters). The CRT screen was also displaced to compensate for the visual displacement of the screen, as described in Experiment 1. Adaptation was performed with visual feedback at the end of each movement, in a block of either 30, 150, or 500 trials.

In the post-test block visual feedback was deprived in order to test the speed of prism after-effect decay following a short delay. By contrast with Experiment 1, in Experiment 2, both the prismatic and CRT displacement were maintained throughout both the adaptation and the post-test block. Logically, during the post-test block, the visual displacements imposed during the exposure period (i.e.: by the prisms and the CRT), should be removed. However, as this removal did not change any visual experience, these displacements were maintained throughout the post-test block in Experiment 2 to save time. That is, regardless of whether the prism and the CRT were in place or not, the initial visual input (before movement onset) was no different. Further,
the pointing measures taken during the post-test block were obtained under conditions of no visual feedback, so the presence or absence of the prism and CRT was irrelevant during the post-test period in which behavior was sampled. By skipping the process of removing the displacements in Experiment 2, this afforded us greater temporal resolution, enabling us to test the speed of memory decay without a delay that would otherwise be unavoidable. The post-test consisted of 50 trials (in the 30-trials adaptation condition) and 100 trials (in the 150- and 500-trials adaptation conditions).

On Day 2, retention of the prism after-effect was tested, 24 hours after the adaptation and post-test blocks of Day 1. During the post-test on Day 2, participants performed 30 trials of reaching without prisms and without visual feedback.

Each volunteer (n = 24) participated in three sessions, once for each of the 30-, 150-, and 500-trial adaptation conditions. The order of the three conditions was counterbalanced across participants. Inter-session intervals were 7 days on average (s.e.m. = 1 day). The direction of the visual displacement was fixed to the left or to the right for each participant.

Data analysis

Two-, three-, and four-state models (Experiments 1 and 2)

We analyzed horizontal errors because the visual field was displaced in the horizontal direction. The error was defined as the discrepancy (distance) between the actual position of the target and the reach endpoint, irrespective of whether the participants were allowed or not allowed to see the errors. The mean horizontal error in trials with no visual feedback during the pre-exposure period was adjusted to zero. We
then fitted the two-, three- and four-state models to the data from each participant, and to the group mean of the horizontal errors averaged across the 7 (Experiment 1) and 24 participants (Experiment 2). To further estimate confidence intervals for each parameter, we prepared 100 bootstrap data samples for each experiment by resampling 7 (Experiment 1) and 24 (Experiment 2) participants with replacement, and applied the two-, three-, and four-state models to the mean of each bootstrap data sample (Burnham and Anderson 2002).

The two-state model was formulated after Smith et al. (2006) with a modification as follows:

\[
\begin{align*}
  e(n) &= d(n) + x(n) - \text{bias} \\
  x(n) &= x_f(n) + x_s(n) \\
  x_f(n) &= A_f x_f(n-1) - B_f e(n-1) \\
  x_s(n) &= A_s x_s(n-1) - B_s e(n-1) \\
  0 &< A_f < A_s < 1, \\
  0 &< B_s < B_f < 1
\end{align*}
\]  

(1)

where the horizontal error on the n-th trial, \( e(n) \), was assumed to be the sum of the visual displacement, \( d(n) \), the total amount of adaptation, \( x(n) \), and a bias term. The bias term was added to capture the small difference at baseline between the error on trials with visual feedback versus on trials without visual feedback. The total amount of adaptation consisted of the sum of states of the fast system, \( x_f(n) \), and the slow system, \( x_s(n) \), both of which were assumed to grow in proportion to the horizontal error on the previous trial, \( e(n-1) \), with constant “learning rates” of \( B_f \) and \( B_s \). The larger the learning rate the faster the adaptation. \( A_f \) and \( A_s \) denote “retention factors”, which determine the speed of memory decay. The smaller the retention factor the faster the decay. Thus, the
two inequalities reflect our assumption that the fast system learns faster and decays faster than the slow system.

The three- and four-state models were formulated as follows:

\begin{align}
  e(n) &= d(n) + x(n) - \text{bias} \\
  x(n) &= x_f(n) + x_s(n) + x_{us}(n) \\
  x_f(n) &= A_f x_f(n-1) - B_f e(n-1) \\
  x_s(n) &= A_s x_s(n-1) - B_s e(n-1) \\
  x_{us}(n) &= A_{us} x_{us}(n-1) - B_{us} e(n-1) \\
  x_{hs}(n) &= A_{hs} x_{hs}(n-1) - B_{hs} e(n-1)
\end{align}

\begin{align}
  0 &< A_f < A_s < A_{us} < 1, \\
  0 &< B_{us} < B_s < B_f < 1
\end{align}

and

\begin{align}
  e(n) &= d(n) + x(n) - \text{bias} \\
  x(n) &= x_f(n) + x_s(n) + x_{us}(n) + x_{hs}(n) \\
  x_f(n) &= A_f x_f(n-1) - B_f e(n-1) \\
  x_s(n) &= A_s x_s(n-1) - B_s e(n-1) \\
  x_{us}(n) &= A_{us} x_{us}(n-1) - B_{us} e(n-1) \\
  x_{hs}(n) &= A_{hs} x_{hs}(n-1) - B_{hs} e(n-1)
\end{align}

\begin{align}
  0 &< A_f < A_s < A_{us} < A_{hs} < 1, \\
  0 &< B_{hs} < B_{us} < B_s < B_f < 1
\end{align}

where subscripts \( f, s, \text{us} \) and \( hs \) represent parameters for the fast, slow, ultra-slow and the hyper-slow systems, respectively.

The models were fitted to the data by applying the least squares method using Matlab® (version R2013a; optimization and global optimization toolboxes). The optimization under the condition of inequalities was achieved by using the \textit{fmincon} function implemented in the optimization toolbox, and repeated 100 times with different initial conditions by using the \textit{multistart} function in the global optimization toolbox. We further restricted the range of search for the retention factor of the fast system between...
0.7 and 0.95 (0.7 < $A_f$ < 0.95), and the range of search for the learning rate between 0 and 0.3 (0 < $B_f$ < 0.3). These ranges cover the reported values for the fast system ($A_f$ = ~0.9, $B_f$ < 0.11) in previous studies (Joiner and Smith 2008; Smith et al. 2006). We set upper bounds for the learning rates ($B_s$ < 0.10, $B_{us}$ < 0.033, $B_{hs}$ < 0.011), so that the upper bounds decreased by a factor of 1/3.

To compare the two-, three-, and four-state models for their relative goodness of fit, the determination coefficient (d.c.) and Akaike’s Information Criterion (AIC) were calculated for each model. The d.c. and AIC were calculated using the formula as follows:

\[
d.c. = 1 - \frac{\text{var(residual error)}}{\text{var(horizontal error)}} \quad (4),
\]

\[
AIC = n \ln(\text{var(residual error)}) + 2k \quad (5),
\]

where $n$ denotes the number of data points ($n = 50 + 100 + 100$) and $k$ denotes the degrees of freedom in each model ($k = 5$ for the two-state, 7 for the three-state, and 9 for the four-state models). The d.c. represents the proportion of the variance of the data explained by the model: it takes the maximum value of one when there are no residual errors. However, the d.c. is not suitable for choosing the best of the three models, because the d.c. increases with the number of parameters. AIC takes not only the residual error but also the degrees of freedom ($k$) into account: the degrees of freedom are added as a penalty to the natural logarithm of the variance of the residual error. Thus, the model that yields the smallest AIC can be judged as the best model. When we applied the two-, three-, and four-state models to each dataset, the AIC difference was calculated by subtracting the smallest value of the three: the model that yielded zero was judged to be the best model (Burnham and Anderson 2002). To test whether the AIC
difference was significant or not, we applied the sign test (non-parametric paired sample sign test) for the results obtained for the 7 (Experiment 1) and 24 (Experiment 2) participants. We further applied the sign test to the results obtained for the 100 bootstrap samples for each experiment. The non-parametric sign test was used because the AIC difference was bounded by zero and did not follow the normal distribution.

An exponential model for approximating the immediate decay (Experiment 2)

To assess the speed of after-effect decay after a short delay, we analyzed horizontal errors during the Day 1 post-test in Experiment 2. As the error tended to increase across trials (reflecting decay) to reach a plateau, the mean horizontal error was fitted to the exponential function with a plateau, formulated as follows:

\[ e(n) = (e(1)-\text{plateau}) \exp(-k \cdot n) + \text{plateau} \]  

where \( e(n) \) is the horizontal error in the \( n \)-th trial, and \( k \) is the constant of decay. We further tested whether the mean horizontal error at the end of the post-test period (i.e.: on trial 50) was the same across the 30-, 150-, and 500-adaptation conditions. Student t-tests were used after correcting the level of significance to 0.05/3 (Bonferroni correction).

Assessments of long-term retention (Experiment 2)

We examined long-term retention by assessing the magnitude of prism after-effects after a delay of 24 hours. The after-effect, defined as the mean error on the initial 5 trials on Day 2 (with no visual feedback), was calculated for each participant for each
of the 150- and 500- adaptation conditions, and conditions were compared using a
paired-samples t-test. We further examined whether the magnitude of the after-effect
correlated better with the amount of adaptation in the slow system or in the ultra-slow
system.
**Results**

**Experiment 1**

Figure 1A shows the average data from 7 participants. During the pre-test, the horizontal error was distributed around zero. During the first exposure phase (Exposure 1), participants initially made leftward errors in the direction of the prismatic displacement, with maximal errors of ~60 mm (trial 61), i.e.: the magnitude of the displacement, but they learned to gradually reduce their errors, with the initial error decreasing to less than 15 mm within 30 trials. When visual feedback was deprived, during the subsequent 10 trials (trials 91-100), the error increased from 14 to 26 mm on average, reflecting decay of what was learned. Further learning occurred in the next set of visual feedback trials, reflected in a reduction of horizontal errors, but these increased (i.e. decayed) again in the next phase when visual feedback was deprived, although the magnitude of error increase (decay) was smaller in this block than in the previous one. In the fifth epoch of no-visual feedback trials, the error was small but still increased from 4.8 mm on the first trial (trial 251) to 8.3 mm on the 10th trial (trial 260) (Fig. 1B).

During the second exposure phase (Exposure 2), in which the visual field was displaced 60 mm to the right, participants initially made a large leftward error (about 100 mm), but then the error decreased rapidly. During the first few trials (trial 276 - 279) of the post-test, participants made errors as small as 5 mm to the left (Fig. 1A and C). This may indicate that the adaptation to the initial visual displacement was completely washed out, and an after-effect to the second visual displacement was observed. However, the errors did not simply converge to zero, but rather crossed the zero-line and had increased up to
15 mm rightward within ~10 trials (trial 286 – 295, Fig. 1A and C), suggesting that the
adaptation to the initial visual displacement had re-emerged (adaptation rebound).

We assessed goodness of fit of the models to individual participant data as well as
to the group mean data. Data from a representative individual are shown in Fig. 2A,
showing changes in horizontal error and a clear adaptation rebound as for the group
mean data, although the noise level is of course much larger. The data agreed
quantitatively with the two-state model. Eighty percent of the variance in the
participant’s data was explained by the two-state model (red line in Fig. 2A,
determination coefficient = 0.800). For the individual data, a small improvement was
observed, in terms of the determination coefficient (d.c. = 0.802), when a third, slower
(ultra-slow) component was added to the model (Fig. 2B). However, this was
accompanied by an increase in AIC of ~2 (ΔAIC = 1.7). No further improvement was
observed with the four-state model in terms of d.c. As a result, the AIC increased by 4
in the 4-state model as compared to the three-state model (ΔAIC = 5.8). Therefore, the
two-state model was judged to be the best in terms of AIC for this particular participant.

Likewise, the two-state model performed best in 6 of 7 participants, and the four-
state model was the worst in all (Fig. 2E). The paired sign tests showed that the AIC
was significantly larger with the four-state model as compared to the two- and three-
state models (Fig. 2E, p = 0.016 < 0.05/3, Bonferroni correction), but the difference
between the two- and three-state models did not reach significance.

As for the group mean data, more than 90% of the variance was explained by the
two-state model (Fig. 2C, d.c. = 0.934). However, the d.c. was not improved by adding
the ultraslow component because the learning constant converged to zero (Fig. 2D,
magenta line). Accordingly, and the two-state model yielded the smallest AIC. By applying the three models to the mean of 100 bootstrap data samples, we found that the two-state model was selected 91 times (model selection frequency = 0.91), the three-state model was selected 9 times (0.09), but never was the four-state model (0). Accordingly, the sign tests showed that the two-state model was significantly better than the three- and the four-state models in terms of the AIC difference (Fig. 2H, $p < 0.001$, 0.05/3, in all combinations).

The results show that the process of adaptation rebound after ~200 trials of the first exposure and 15 trials of the second exposure was best explained by the two-state model, as we initially expected. The median of the estimated parameters were $0.834 (A_f)$, $0.994 (A_s)$, $0.182 (B_f)$, $0.0379 (B_s)$, and $9.4 \text{ mm} (bias)$, with 95% confidence intervals of $[0.741, 0.890] (A_f)$, $[0.988, 1] (A_s)$, $[0.128, 0.234] (B_f)$, $[0.0198, 0.0640] (B_s)$, and $[7.1, 11.6] (bias)$ (Figs. 2I, J).

Experiment 2

Figure 3A shows the average data from 24 participants. During the pre-test period, the horizontal error was distributed around zero. During the exposure period, the initial error was approximately the magnitude of the visual displacement (~50 mm), but decreased across trials to less than 10 mm within 30 trials, and reached a plateau of ~5 mm thereafter.

During the post-test period, the horizontal error increased: that is, the after-effect decayed in an exponential manner across trials in all conditions (Fig. 3A and B). The rate of decay was the largest in the 30-trial adaptation condition ($k = 0.11$), was nearly

18
halved in the 150-adaptation condition (0.058), and further near-halved in the 500-adaptation condition (0.026). The mean error at trial 50 was significantly different in any combination of the three conditions \((p < 0.0077 < 0.05/3\), paired t-tests). This result indicates that the rate of decay was significantly different across the three conditions.

Next, we examined which of the two-, three-, and four-state, models could best explain these data. In a typical example (participant #19), the d.c. improved from 0.471 (Fig. 4A, two-state) to 0.475 (Fig. 4B, three-state), and the AIC was also improved by 4.4, when the ultra-slow component was added to the two-state model. However, there was no further improvement when the hyper-slow component was added to the three-state model. As for the averaged data, the d.c. was improved from 0.888 (two-state) to 0.8928 (three-state), and the AIC was improved by 37 (Figs. 4C, D) by adding the ultra-slow component to the two-state model. The d.c. was further improved by adding the hyper-slow component \((d.c. = 0.8931\) but the AIC increased by 1.7, showing that the three-state model was the best. This was further confirmed by applying the models to the 100 bootstrap data samples (Fig. 4H): the model selection frequency was the largest for the three-state model (0.41), followed by the four-state model (0.32), and the least for the two-state model (0.27). The three AIC comparisons were all significant (paired sign tests, \(p < 0.0005 < 0.05/3\)). The results clearly show that the three-state model best explained the different decays after 30-, 150-, and 500-adaptation trials.

The median of the estimated parameters were 0.914 \((A_f)\), 0.995 \((A_s)\), 0.9996 \((A_{us})\), 0.202 \((B_f)\), 0.0612 \((B_s)\), 0.0330 \((B_{us})\), and -3.0 mm \((bias)\), with 95% confidence intervals of [0.795, 0.95] \((A_f)\), [0.987, 0.999] \((A_s)\), [0.997, 1] \((A_s)\), [0.119, 0.296] \((B_f)\), [0.0307, 0.10] \((B_s)\), [0.0082, 0.0330] \((B_{us})\), and [-6.9, 0.47] \((bias)\) (Figs. 4I, J).
The long-term after-effect at 24 hours was tested in the 150- and the 500-
adaptation conditions. We defined the magnitude of long-term retention as the mean of
the horizontal errors in the first 5 trials in the post-test period on Day 2. The mean prism
after-effect magnitude after 500 trials of adaptation (18 ± 12 mm, mean ± s.d.) was
significantly larger than that after 150 trials of adaptation (8.7 ± 11 mm) (Fig. 5B, \( p =
0.0002 \), paired t-tests).

Previous work (Joiner and Smith 2008) suggested that retention at 24 hours
depended on the amount of adaptation of the “slow process” of the two-state model.
Hence, we calculated the magnitude of adaptation of the “slow process” at the end of
Day 1 for each participant, based on the two-state model, and compared that value with
the magnitude of prism after-effect retained at 24 hours (Fig. 5C). There was no
significant correlation between the magnitude of adaptation of the “slow process” and
the size of the after-effect at 24 hours (\( r = 0.10, \ p = 0.50 \)). By contrast, the amount of
adaptation of the “ultra-slow process” in the three-state model did correlate with the
amount of prism after-effect retained at 24 hours (Fig. 5D; \( r = 0.52, \ p = 0.00016 \)). The
slope of an orthogonal regression line was close to 1 (1.0; 95% confidence interval (c.i.)
\( = [0.55 \ 1.84] \); bootstrap method, \( n = 10000 \)) with an y-intercept of -6.1 mm (c.i. = [-18
3.1]), suggesting that the adaptation of the ultra-slow process decayed little after an
interval as long as 24 hours. (Alternatively, the adaptation of the ultra-slow process may
have de-adapted during the 24 hours, and copied to another unknown system for long-
term retention.) The amount of adaptation of the “hyper-slow process” in the four-state
model significantly correlated with the amount of prism after-effect retained at 24 hours,
but the correlation was weaker than that yielded by the three-state model (Fig. 5E; \( r =

0.37, \( p = 0.0103 \)). It is also worth noting that there was no significant correlation between the magnitude of adaptation, quantified as the last 5 trials of the adaptation block on Day 1, and the size of the after-effect at 24 hours (\( r = 0.08, \ p = 0.60 \)). The results clearly indicate that what was carried over to the next day was proportional to the adaptation in the ultra-slow system of the three-state model, with a retention ratio close to one.
Discussion

In Experiment 1, we confirmed that an “adaptation rebound”, a critical prediction of the two-state model, occurred in the post-test block (in which visual feedback was deprived) following a block of adaptation and de-adaptation. To our knowledge, this is the first demonstration of adaptation rebound during visuomotor adaptation that occurs immediately after an adaptation-extinction episode (however, see Hatada et al. 2006 for adaptation rebound that developed over days). The fit of the two-state model to the data support the hypothesis that prism adaptation also depends on two systems, one that learns fast but forgets fast, and another that learns slowly but forgets less. We further fitted three-state and four-state models to the data in Experiment 1, which contained 150 adaptation trials. We predicted that adding the slower systems would not improve the fit of the model with this number of adaptation trials. As expected, the two-state model was shown to be the best in terms of the AIC.

In Experiment 2, which included 500 adaptation trials in one condition, we predicted that the three-state model would better represent the data as a whole than the two-state model. Our prediction was confirmed again: the three state model was shown to be better than the two- or four-state models in terms of the AIC. In addition, the amount of retention on Day 2 significantly correlated with the amount of adaptation in the ultra-slow system of the three-state model at the end on Day 1, but not with the amount of adaptation in the slow system (two-state model). These results consistently support our hypothesis that there is a third slower component in prism adaptation in addition to the conventional fast and slow systems.
Is the ultraslow system unique to prism adaptation?

Why were three systems required to account for these data in prism adaptation, whereas only two systems were sufficient to explain short-term as well as long-term retention at 24 hours in force field adaptation (Joiner and Smith 2008; Smith et al. 2006)? Recently, Vaswani and Shadmehr (2013) examined decay of after-effects after adaptation to a force field perturbation for 300 trials. In the post-test period, errors were clamped to zero (error-clamp trials). They found that the after-effect decayed but reached a plateau, similar to our present results. They examined whether the fast-slow model could explain their data. With the best-fit parameters for the period of adaptation

$A_f = 0.876, B_f = 0.315, A_s = 0.995, B_s = 0.056$, ~70% of the variance during a long post-exposure period was explained by the two-state model, but the model failed to reproduce the plateau: the two-state model predicted that the after-effects would approach zero. This was true even when we fitted the two-state model to the overall data during and after adaptation (Fig. 6A, $A_f = 0.886, B_f = 0.266, A_s = 0.995, B_s = 0.055$; d.c. $= 0.864$). This discrepancy between the model predictions and the observed data can be resolved by positing an additional ultra-slow system (Fig. 6B, $A_f = 0.690, B_f = 0.189, A_s = 0.977, B_s = 0.147, A_{us} = 0.9991, B_{us} = 0.013$). With the three-state model, the endpoint of the decay (~0.2) was also reproduced (Fig. 6B). This result suggests that the ultra-slow system may not be unique to prism adaptation but might be at work even during motor learning in response to force field perturbation.

Neural mechanisms
The present study has demonstrated that prism adaptation behaviour can be explained by a simple state-space model that posits the existence of three adaptive processes operating in parallel with different learning rates and retention capacities. One question is where in the brain these proposed functional mechanisms are implemented.

There are several human neuroimaging studies that have used target-reaching with visual displacement induced by wedge prisms (Clower et al. 1996; Danckert et al. 2008; Luauté et al. 2009). However, Luauté et al. (2009) was the only one that examined blood oxygenation level-dependent (BOLD) signal changes during exposure to a visual displacement over an extended period, sufficient for the fast and the slow systems to develop (i.e., more than 20 trials; see traces in Fig. 4). Extrapolating from Fig. 4D (see green trace), and assuming that fMRI signals in the fast system grow over the first few trials, peak at around trial 10, and gradually subside thereafter, BOLD signal changes in the left parieto-occipital sulcus (POS) in the Luaute et al. study would appear to match the profile of the fast system. On the other hand, slower signal development was observed in lobules IV and V of the cerebellar cortex, in the hemisphere ipsilateral to the hand that was used for reaching: signals in the cerebellar cortex reached a peak at trial 16 and subsided thereafter (Luauté et al. 2009). These results raise the possibility that the fast system is implemented in the cerebral cortex (POS), whereas the slow system is implemented in lobules IV and V in the cerebellar cortex. In monkeys, complex spikes recorded from Purkinje cells in lobules IV to VI in the cerebellar hemisphere ipsilateral to the hand appear to encode end-point errors in reaching (Kitazawa et al. 1998). We may speculate that long-term depression, triggered by complex spikes in the cerebellar cortex (Ito 2001; 1989), is a part of the slow learning
process. A causal role for the cerebellum in prism adaptation is indicated by impaired performance following cerebellar lesions in humans (review in Jacquin-Courtois et al. 2013; Martin et al. 1996; Pisella et al. 2005; Thach et al. 1992; Weiner et al. 1983) as well as in monkeys (Baizer et al. 1999).

Studies have recently suggested that the memory trace of adaptation in ocular reflexes is initially acquired in the cerebellar cortex (flocculus) and undergoes short term storage and is then “transferred” to the vestibular nuclei, whereby the memory lasts over days (Anzai et al. 2010; Kassardjian et al. 2005; Shutoh et al. 2006). Extending these speculations regarding the ocular reflex to prism adaptation, we may suggest that plastic changes in the deep cerebellar nuclei are a likely candidate for the ultra-slow process that would follow fast-process changes in the cerebellar cortex (Nagao et al. 2013).

The above-mentioned possibility that the fast, slow, and the ultra-slow processes implicated in prism adaptation are implemented in the left POS (fast), the cerebellar cortex (slow), and the deep cerebellar nuclei (ultra-slow), may seem to contradict a report by Galea et al. (2011), in which the cerebellar cortex was implicated in a fast-learning/forgetting process and the primary motor cortex (M1) was implicated in a slow-learning/forgetting process during adaptation of reaching movements to visuo-motor rotation. However, this seeming contradiction may be resolved by taking the following points into account. Firstly, adaptation to visuo-motor rotation generally occurs much more slowly than prism adaptation. From Galea et al.’s results, it took at least 20 trials of learning for the amount of initial error to be halved (from 30 degrees to 15 degrees), whereas, by contrast, it took only a few trials (less than 5 trials) for the
initial error to be halved in the present and other studies on prism adaptation (Kitazawa et al. 1997; Luauté et al. 2009; Weiner et al. 1983). Hence, the fast process for prism adaptation may be distinct from the fast process implicated in adaptation to visuo-motor rotation, with the latter possibly corresponding to the slow process in prism adaptation. Assuming this, both studies agree in that each posits that the cerebellar cortex is involved in learning with a similar time constant, irrespective of whether that process is termed “fast”, as in Galea et al. (2011), or “slow” as in the present study. Secondly, involvement of the primary motor cortex in retention does not, of course, preclude additional involvement of the deep cerebellar nuclei, which receive inputs from cerebellar lobules IV-VI, particularly since this part of the dentate nucleus is closely connected with M1 via the thalamus and the pontine nucleus (Dum et al. 2002). It would not be surprising if two brain structures, both located within the same circuit inter-connecting M1 and the dentate nucleus, would be involved in long-term retention of a motor memory. We therefore speculate that the ultra-slow process may be implemented in a more or less distributed fashion across a cerebro-cerebellar loop inter-connecting the primary motor cortex, pontine nucleus, dentate nucleus, thalamus, and back again to M1.

Future directions

In the present study, we assumed that the same error drives all three components simultaneously, as has been hypothesized by the original study of Smith and colleagues (Joiner and Smith 2008; Smith et al. 2006). However, it is possible that the brain is solving a credit assignment problem so that the error is efficiently assigned to one of the
many components by using a Bayesian solution (Kording et al. 2007). It is also possible
that reinforcement learning partially contributes to the acquisition and stabilization of
the ultra-slow component (Shmuelof et al. 2012).

Future work combining computational modeling (Berniker and Kording 2011;
Kording et al. 2007; Lee and Schweighofer 2009) with detailed characterization of
adaptation behavior (O’Shea et al. 2014), together with brain imaging techniques,
should help to further elucidate the temporal dynamics of the learning systems
mediating prism adaptation and help to determine the underlying neural substrates.
Grants

This research was supported by a research grant from the Human Frontiers Science Program Organization, and by a Grant-in-Aid for Scientific Research on Innovative Areas (#25119002). JO’S is supported by the Royal Society, U.K.
References


Jacquin-Courtois S, O'Shea J, Luaute J, Pisella L, Revol P, Mizuno K, Rode G,


Figure legends

Figure 1 Adaptation rebound in prism adaptation (Experiment 1). A: Group mean horizontal errors plotted against trial sequence (n = 7). The thick solid line indicates the size of the prismatic displacement of the visual field. Open circles indicate horizontal errors in the visual-feedback trials; filled circles indicate horizontal errors in the no-visual-feedback trials. Note the adaptation rebound during the post-test period (trials 276-325). B: The mean horizontal errors in trials 1 and 10 during the first and the 5th epoch of no-visual-feedback trials during the first exposure period. C: The mean horizontal error during the post-test period. Note that the mean error during trials 11-20 is inverted and larger than the mean error during trials 1-3 indicating that clear “adaptation rebound” had occurred. Error bars show the standard error of the mean.

Figure 2 Application of the multi-state models to the data showing adaptation rebound. A, B: The two- (A) and three-state models (B) fitted to horizontal errors from a typical participant (subject 1). C, D: The two- (C) and three-state models (D) fitted to horizontal errors from the group mean across 7 participants (average). Data in (C) and (D) are the same as those in Figure 1. Green, blue, and magenta traces show development of the fast (green), slow (blue), and ultra-slow (magenta) systems, and red traces show the motor output predicted from the two- (A, C) and three-state (B, D) models. E, H: Comparison of AIC among the three models applied to data from individual participants (E, n = 7) and the mean of bootstrap data sample (n = 100). Note that the smallest AIC (ΔAIC = 0) was yielded by the two-state model in 6 of 7 participants (E), and 91 of 100 bootstrap data samples. Brackets show pairs of
comparison with significant differences (p < 0.05/3, Bonferroni correction). F, G, I, J:
Distribution of the retention factors (F, I) and learning rates (G, J) estimated by the two-
state model for each participant (F, G) and for each of 100 bootstrap data samples.
Notches and horizontal lines in the box plots show 2.5, 25, 50, 75 and 97.5 percentiles.

**Figure 3** Comparison of the prism after-effect decay after 30, 150, and 500 trials of
adaptation (Experiment 2, Day 1). A: The mean horizontal errors plotted against trial
sequences (n = 24). Note that the prisms were not removed in the post-test period, and
hence the horizontal errors are continuous across the adaptation and the post-test period.
B: The exponential decay model fitted to the mean horizontal error during the post-test
period in the 30- (cyan circle), 150- (brown circle), and 500- (purple circle) adaptation
conditions.

**Figure 4** Application of the multi-state models to the data in Experiment 2. A, B: The
two- (A) and three-state models (B) fitted to horizontal errors from a typical participant
(subject 19). C, D: The two- (C) and three-state models (D) fitted to the group mean
across 24 participants (C, D; average). Data in (C) and (D) are the same as those in
Figure 3A. Green, blue, and magenta traces show development of the fast (green), slow
(blue), and ultra-slow (magenta) systems, and red traces show the motor output
predicted from the two- (A, C) and three-state (B, D) models. E, H: Comparison of AIC
among the three models applied to data from individual participants (E, n = 24) and the
mean of bootstrap data sample (n = 100). Brackets show pairs of comparison with
significant differences (p < 0.05/3, Bonferroni correction). Note in (H) that the AIC
yielded by the three-state model was significantly smaller than those yielded by the two-
state model ($p = 0.0000068$) and four-state model ($p = 0.00047$). F, G, I, J: Distribution of the
retention factors (F, I) and learning rates (G, J) estimated by the three-state model for
each participant (F, G) and for each of 100 bootstrap data samples. Other conventions
are the same as in Figure 2.

**Figure 5** After-effects at 24 hours (Day 2). A: The mean horizontal errors at 24 hours
plotted against trial sequences in the 150- and 500-adaptation conditions (n = 24).
Broken lines show the standard error of the mean. B: Comparison of the size of the
after-effect averaged over the initial 5 trials. Error bars show the 95% confidence
interval across participants. C-E: The size of the prism after-effect at 24 hours (Day 2,
retention) plotted against adaptation at the end of Day 1: adaptation of the slow process
predicted by the two-state model (C), adaptation of the ultra-slow process predicted by
the three-state model (D), adaptation of the hyper-slow process predicted by the four-
state model (E), and adaptation calculated from the raw data (E). Note that the
correlation was significant with the three-state model (D) ($r = 0.52$, $p=0.00016$), and
with the four-state model (E) ($r = 0.37$, $p = 0.010$), but not with the two-state model (C).
The broken line in (D) shows the orthogonal regression line with a slope close to 1 ($y =
1.0 x – 6.1$ mm).

**Figure 6** Application of the two-state (A) and three-state model (B) to the experimental
data replotted from Figure 8A in Vaswani and Shadmehr (2013). Note that the
asymptote in the original data (black dots) is reproduced by the three-state model (red trace in B), but not by the two-state model (red trace in A).
Fig. 1

A

Pre-test Exposure-1 Exposure-2 Post-test

Horizontal error (mm)

Visual-feedback trial
No-visual-feedback trial
Visual displacement

Prism displacement (mm)

B

Epoch

Horizontal error (mm)

Trial 1 Trial 10

C

Horizontal error (mm)

Trial 1-3 Trial 11-20
Fig. 2

2-state model Subject 1

- $A_f = 0.887$, $B_f = 0.0703$
- $A_s = 0.993$, $B_s = 0.0184$
- d.c. = 0.800
- $\Delta AIC = 0$

3-state model Subject 1

- $A_f = 0.854$, $B_f = 0.061$
- $A_s = 0.979$, $B_s = 0.0258$
- $A_{us} = 1$, $B_{us} = 0.00461$
- d.c. = 0.802
- $\Delta AIC = 1.7$

2-state model Average

- $A_f = 0.845$, $B_f = 0.176$
- $A_s = 0.994$, $B_s = 0.0356$
- d.c. = 0.934
- $\Delta AIC = 0$

3-state model Average

- $A_f = 0.845$, $B_f = 0.176$
- $A_s = 0.994$, $B_s = 0.0356$
- $A_{us} = 0.999$, $B_{us} = 0$
- d.c. = 0.934
- $\Delta AIC = 4.0$

Retention factor ($A$)

- $A_f$ vs. $A_s$

Learning rate ($B$)

- $B_f$ vs. $B_s$
Fig. 5

A  24h after 150-adaptation condition  24h after 500-adaptation condition

Horizontal error (mm)  Trial

B  Day 1 condition

Day 2 retention (mm)

0 5 10 15 20

0 5 10 15 20

2-state model  3-state model

C  Day 1 adaptation (slow process, mm)

Day 2 retention (mm)

0 20 40 60 80

r = 0.101  p = 0.496

D  Day 1 adaptation (ultra-slow process, mm)

Day 2 retention (mm)

0 20 40 60 80

r = 0.518  p = 0.000161

E  Day 1 adaptation (hyper-slow process, mm)

Day 2 retention (mm)

0 20 40 60 80

r = 0.367  p = 0.0103

F  Raw data

Day 2 retention (mm)  Day 1 adaptation (net, mm)

0 50 100

-50 0 50 100
Fig. 6

A 2-state model

B 3-state model