Intrinsic rewards modulate sensorimotor adaptation

Hyosub E. Kim[1,2]*, Darius E. Parvin[1,2], Richard B. Ivry[1,2]

[1]Department of Psychology and [2]Helen Wills Neuroscience Institute, University of California, Berkeley

ABSTRACT

Recent studies have demonstrated that task success signals can modulate behavioral changes during sensorimotor adaptation tasks, primarily through the engagement of explicit processes. In a series of reaching experiments with human participants, we explore a potential interaction between reward-based learning and implicit adaptation, using a method in which feedback is not contingent on task performance. We varied the size of the target to compare conditions in which visual feedback indicated an invariant angular error that either hit or missed the target. Hitting the target attenuated the behavioral changes from adaptation, an effect we attribute to the generation of an intrinsic reward signal. We evaluated two models, one in which reward and adaptation systems operate in parallel, and a second in which reward acts directly on the adaptation system. The results favor the latter, consistent with evidence showing communication, and possible overlap, between neural substrates underlying reward-based learning and sensorimotor adaptation.

INTRODUCTION

Multiple learning systems contribute to successful goal-directed actions in the face of changing physiological states, body structures, and environments (Huberdeau, Krakauer, & Haith, 2015; McDougle, Ivry, & Taylor, 2016; Jordan A. Taylor & Ivry, 2014). Among these different learning processes, implicit sensorimotor adaptation is of primary importance for maintaining appropriate calibration of sensorimotor maps over both short and long timescales. A large body of work has focused on how sensory prediction error (SPE), the difference between predicted and actual sensory feedback, drives sensorimotor adaptation (Shadmehr, Smith, & Krakauer, 2010). In addition to

sensorimotor adaptation, there is growing awareness of how reward-based learning contributes to motor control. While several recent studies have shown that rewarding successful actions alone is sufficient for the learning of perturbations (Izawa & Shadmehr, 2011; Therrien, Wolpert, & Bastian, 2016, 2018), little is known about how rewards impact implicit adaptation. Thus, a central question remains as to how learning systems tuned to SPE versus those tuned to rewards interact during motor tasks.

Despite utilizing very similar task paradigms, initial studies have led to an inconsistent picture of how reward impacts performance in sensorimotor adaptation tasks. For example, in two separate visuomotor rotation studies using similar task paradigms and reward structures, the first study reported no effect of reward on adaptation rates but an enhancement of motor memory due to rewards (Galea, Mallia, Rothwell, & Diedrichsen, 2015), while the second reported a beneficial effect of rewards specifically on adaptation rate (Nikooyan & Ahmed, 2015). In a more recent study, however, manipulation of reward attenuated overall learning (Leow, Marinovic, & Carroll, 2018).

One factor that may contribute to these inconsistencies is highlighted by recent work showing that, even in relatively simple sensorimotor adaptation tasks, overall behavior reflects a combination of explicit and implicit processes (Jordan A. Taylor & Ivry, 2011; Jordan A. Taylor, Krakauer, & Ivry, 2014). Unless the explicit component is directly assayed (Jordan A. Taylor et al., 2014), measures of adaptation can be confounded by explicit aiming. That is, while the SPE is thought to drive adaptation (Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007), participants are often also consciously aware of the perturbation and decide to aim in order to compensate for it, thereby improving task performance. It may be that reward promotes the activation of explicit processes, which can be more flexibly implemented depending on the task demands (Bond & Taylor, 2015). A recent study provides evidence for this hypothesis (Codol, Holland, & Galea, 2017), showing that at least one of the putative effects of reward, the strengthening of motor memories (Shmuelof et al., 2012), is primarily the result of

the re-instantiation of explicit aiming strategies as opposed to a direct modulation of adaptation. As explicit learning is much more flexibly implemented and sensitive to task demands than implicit adaptation (Bond & Taylor, 2015), differential demands on strategies are likely to contribute toward the inconsistent effects reported across studies manipulating reward (Holland, Codol, & Galea, 2018).

A recently developed method, referred to as clamped visual feedback, isolates implicit adaptation from an invariant visual error signal (Morehead, Taylor, Parvin, & Ivry, 2017). During the clamp, the angular trajectory of a feedback cursor is invariant with respect to the target location and thus spatially independent of hand position (Kim, Morehead, Parvin, Moazzezi, & Ivry, 2018; Morehead et al., 2017; Shmuelof et al., 2012; Vandevoorde & Orban de Xivry, 2018; Vaswani et al., 2015). Participants' knowledge of the visual perturbation and instructions to ignore it are intended to prevent any explicit aiming, thus allowing a clean probe of implicit adaptation (Morehead et al., 2017).

Here, we employ the clamp method to better understand how rewards may affect implicit adaptation from SPE, without interference from explicit aiming strategies. In a series of three experiments, the clamp offset was held constant and only the size of the target was manipulated, affecting whether the cursor would hit or miss the target. Thus, we were able to experimentally manipulate both the putative SPE (angular offset of clamp) and the reward (hitting versus not hitting the target). We assume that hitting the target would be intrinsically rewarding (Leow et al., 2018; Xu-Wilson, Zee, & Shadmehr, 2009), even though participants are explicitly aware that hitting the target is independent of their actual performance. Given this assumption, we ask how reward impacts adaptation from a constant SPE.

The results of the first two experiments revealed a strong attenuation of adaptation when the cursor hit the target. Based on these results, in Experiment 3 we evaluated two hypotheses regarding the mechanism by which intrinsic rewards affect adaptation. We hypothesized that either intrinsic reward activates reward-based reinforcement in parallel to SPE-driven adaptation, with movement being the

net result of these two independent processes (Movement Reinforcement model), or intrinsic reward directly modulates adaptation (Adaptation Modulation model). Our results provide support for the latter, although our model-based analyses suggest there may be a mixture of both mechanisms.

RESULTS

In all experiments we used clamped visual feedback, in which the angular trajectory of a feedback cursor is invariant with respect to the target location and thus spatially independent of hand position (Morehead et al., 2017; Fig. 1a). The instructions emphasized that the participant's behavior would not influence the cursor trajectory: They were to ignore this stimulus and always aim directly for the target. This method allows us to isolate implicit learning from an invariant SPE, eliminating potential contributions from strategic changes that might be used to reduce task performance error.

In Experiment 1, we asked whether hitting the target under conditions in which the feedback is not contingent on behavior would modulate adaptation, based on the assumption that this would be intrinsically rewarding. We tested three groups of participants (n=16/group) with a 3.5° clamp offset for 80 cycles (8 targets per cycle). The purpose of this experiment was to examine the effects of three different relationships between the clamp and target while holding the visual error (defined as the center-to-center distance between the cursor and target) constant (Fig. 1b): Hit Target (when the terminal position of the clamped cursor is fully embedded within a 16 mm diameter target), Straddle Target (when roughly half of the cursor falls within a 9.8 mm target, with the remaining part outside the target), Miss Target (when the cursor is fully outside a 6 mm target). As seen in Fig. 1d, hitting the target reduced the overall change in behavior. Statistically, there was a marginal difference on the rate of initial adaptation (one-way ANOVA: $F(2,45)=2.67$, $p=.08$, $\eta^2=.11$; Fig. 1e) and a significant effect on late learning ($F(2,45)=4.44$, $p=.016$, $\eta^2=.17$; Fig. 1f). For the latter measure, the value for the Hit Target group was approximately 35% lower than for the Straddle and Miss Target groups, with post-hoc

comparisons confirming the substantial differences in late learning between the Hit Target and both the Straddle Target (95% CI [-16.13°, -2.34°], t(30)=-2.73, p=.010, d=.97) and Miss Target (95% CI [-16.76°, -2.79°], t(30)=-2.86, p=.008, d=1.01) groups. The learning functions for the Straddle and Miss Target groups were remarkably similar throughout the entire clamp block and reached similar magnitudes of late learning (95% CI [-7.90°, 8.97°], t(30)=.13, p=.898, d=.05).

Interestingly, these results appear qualitatively different to those observed when manipulating the clamp offset. Our previous study using clamped visual feedback demonstrated that varying clamp offset alone results in different early learning rates, but produces the same magnitude of late learning (Kim et al., 2018). The results here in Experiment 1 however, suggest that the intrinsically rewarding feedback associated with hitting the target results in small differences in early learning that are amplified in late learning. Furthermore, the effect of intrinsic reward appears to be categorical, as it was only observed for the condition in which the cursor was fully embedded within the target (Hit Target), and not when the terminal position of the cursor fell partially outside the target (Straddle Target).

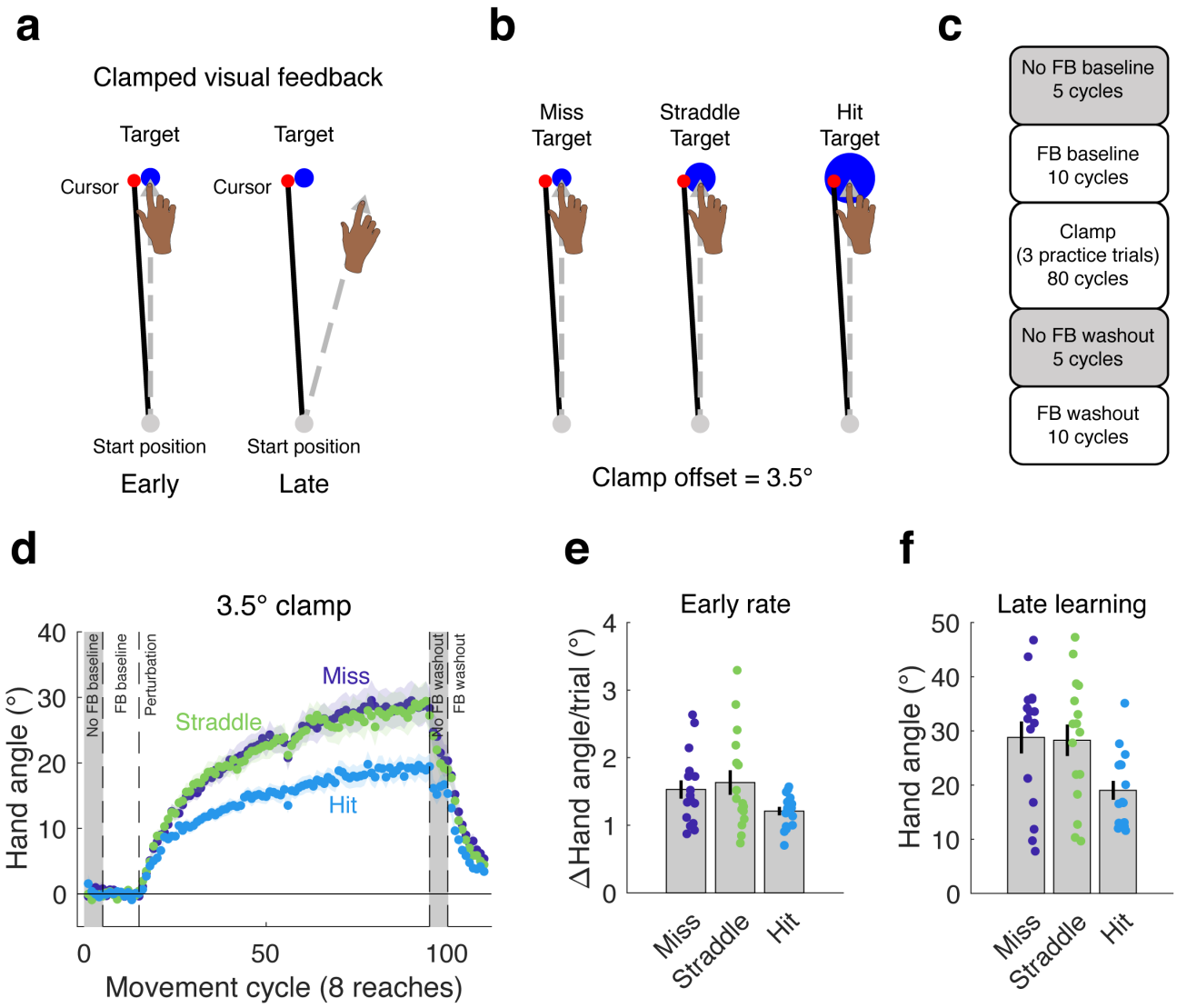


Figure 1 Hitting the target attenuates adaptation from SPE. **(a)** During clamped visual feedback, the angular offset of the cursor feedback is held constant throughout the perturbation block. Despite participants' knowledge of the clamp, this manipulation elicits robust changes in hand angle. **(b)** The clamp offset was equal across all three conditions tested in Experiment 1, with only the target size varying between conditions. **(c)** Block design for experiment. **(d)** There was attenuation of adaptation in the Hit Target condition, observed in the **(e)** early adaptation rate (the per cycle rate of change in hand angle during clamp cycles 3-7), and more dramatically in **(f)** late learning (mean hand angle over last 10 clamp cycles). Dots represent individuals; shading and error bars denote SEM.

Experiment 2

Experiment 2 was designed to extend the results of Experiment 1 in two ways: First, to verify that the effect of hitting a target generalized to other contexts, we changed the size of the clamp offset. We

tested two groups of participants (n=16/group) with a 1.75° clamp offset. For the Hit Target group (Fig. 2a), we used the large 16 mm target, and thus, the cursor was again fully embedded. For the Straddle Target group, we used the small 6 mm diameter target, resulting in an endpoint configuration in which the cursor was approximately half within the target and half outside the target. We did not test a Miss Target condition because having the clamped cursor land fully outside the target would have necessitated an impractically small target (~1.4 mm). Moreover, the results of Experiment 1 indicate that this condition is functionally equivalent to the Straddle Target group. The second methodological change was made to better assess asymptotic adaptation. We increased the number of clamped reaches to each location to 220 (reducing the number of target locations to four to keep the experiment within a 1.5 hour session). This resulted in a nearly three-fold increase in the number of clamped reaches per location.

Consistent with the results of Experiment 1, the Hit Target group showed an attenuated learning function compared to the Straddle Target group (Fig. 2b). Statistically, there was again only a marginal difference in the rate of early adaptation (95% CI [-.52°/cycle, .01°/cycle], $t(30)=-1.96$, $p=.06$, $d=.69$; Fig. 2c), whereas the difference in late learning was quite pronounced (95% CI [-11.38°, -1.25°], $t(30)=-2.54$, $p=.016$, $d=.90$; Fig. 2d). Indeed, the 35% attenuation in asymptote for the Hit Target group compared to the Straddle Target group is approximately equal to that observed in Experiment 1.

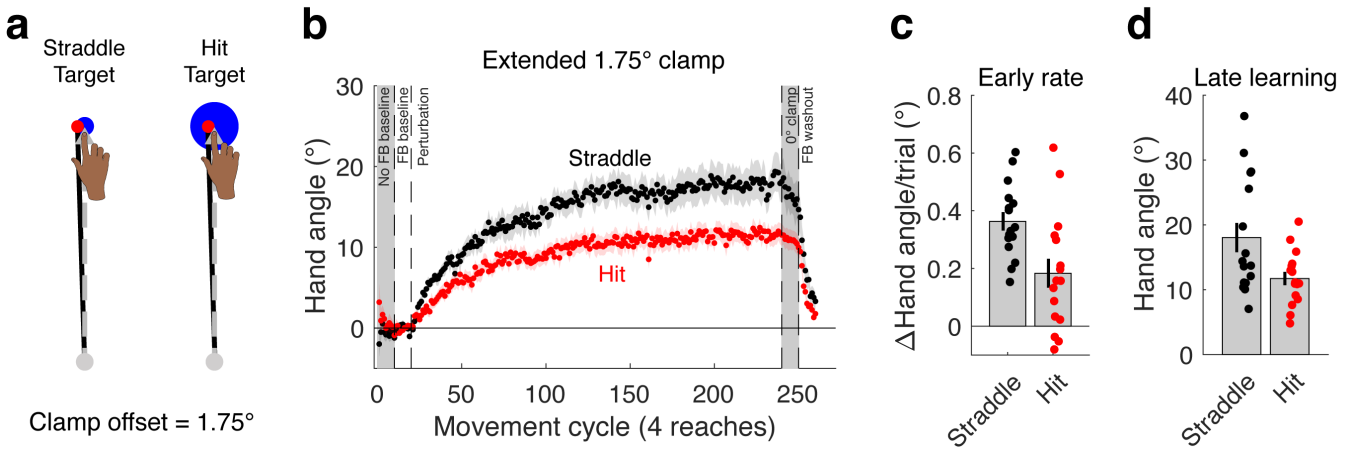


Figure 2 The attenuation of adaptation caused by hitting the target **(a)** generalizes to a different clamp offset and is stable over an extended clamp block **(b)**. There were marginal differences in **(c)** early adaptation rates and **(d)** a significant difference in the magnitude of late learning. Dots represent individuals; shading and error bars denote SEM.

The results of these first two experiments converge in showing that adaptation from SPE is attenuated when the cursor hits the target, relative to conditions in which at least part of the cursor falls outside the target. This effect replicated across two experiments that used different clamp offsets.

Attenuated behavioral changes are not due to differences in motor planning

Although we interpret the attenuation of behavioral change as the effect of an intrinsic reward signal, generated in the Hit Target conditions, there are some alternative explanations for the effect of target size on adaptation. We aim to address these alternatives by analyzing the kinematic data in Experiments 1 and 2.

One alternative is that participants in the Hit Target groups had reduced accuracy demands relative to the other groups, given that they were reaching to a larger target (Soechting, 1984). If the accuracy demands were reduced for these large targets, then the motor command could be more variable, resulting in more variable sensory predictions from a forward model, and thus a weaker SPE (Körding & Wolpert, 2004; see Fig. 3). While we do not have direct measures of planning noise, a reasonable

proxy can be obtained by examining movement variability during unperturbed baseline trials (data from clamped trials would be problematic given the induced change in behavior). If there is substantially more noise in the plan for the larger target, then the variability of hand angles should be higher in this group (Churchland, Afshar, & Shenoy, 2006). In addition, one may expect faster movement times (or peak velocities) and/or reaction times for reaches to the larger target, assuming a speed-accuracy tradeoff (Fitts' law; Fitts, 1992).

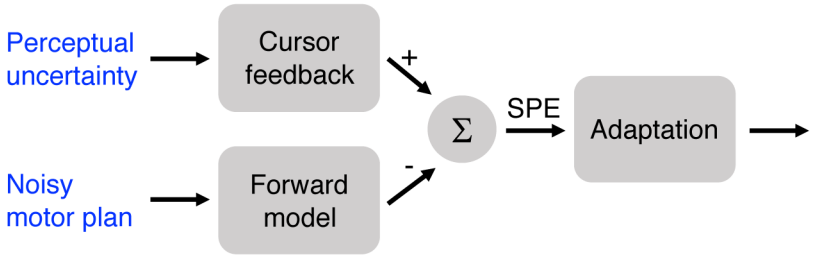


Figure 3 Adaptation could be affected by different factors, such as perceptual uncertainty or greater variability in motor planning. In the case of perceptual uncertainty, the feedback signal is weakened, thus leading to a weaker SPE signal. In the case of noisy motor planning, the forward model prediction would also be more variable and effectively weaken the SPE.

Examination of kinematic and temporal variables did not support this noisy motor plan hypothesis. During baseline trials with veridical feedback, mean spatial variability, measured in terms of hand angle, was actually lower for the group reaching to the larger target (Hit Target group: 3.09° ± .18°; Straddle Target group: 3.56° ± .16°; $t(30)=-1.99$ $p=.056$, $d=0.70$). Further supporting the argument that planning was no different across conditions, neither reaction times (Hit Target: 378 ± 22 ms; Straddle Target: 373 ± 12 ms) nor movement times (Hit Target: 149 ± 8 ms; Straddle Target: 157 ± 8 ms) differed between the groups ($t(30)=-0.183$, $p=.856$, $d=.06$ and $t(30)=0.71$, $p=.484$, $d=.25$, respectively). Qualitatively similar results for baseline behavior were observed in Experiment 1 (see Supplement).

One reason for not observing an effect of target size on accuracy or temporal measures could be due to the constraints of the task. Studies which observe effects of target size on motor planning typically utilize point-to-point movements (Knill, Bondada, & Chhabra, 2011; Soechting, 1984) in which accuracy

requires planning of both movement direction and extent. In our experiments, we utilized shooting movements, thus minimizing demands on the control of movement extent. Endpoint variability is generally larger for movement extent compared to movement direction (Gordon, Ghilardi, & Ghez, 1994). It is possible that participants are near ceiling-level performance in terms of hand angle variability. Another reason for the absence of a speed accuracy trade-off in the current experiment could be that with the clamp method, participants do not receive task performance feedback throughout the experiment.

A second alternative to the intrinsic reward hypothesis is that participants adapted less during Hit Target conditions due to perceptual uncertainty, an idea we test in a control condition in Experiment 3.

Experiment 3

Based on the results of Experiments 1 and 2, we considered two ways in which an intrinsic reward, generated by hitting the target, could attenuate the rate and asymptotic level of learning. First, intrinsic reward could act as a positive reinforcement signal, strengthening the representation of rewarded movements (Shmuelof et al., 2012) (Fig. 4a, Movement Reinforcement model). This would effectively operate as a resistance to the directional change in hand angle induced by SPEs, since the reward would reinforce the executed motor command. By this model, intrinsic reward has no direct effect on the adaptation process, in that reward and error-based learning are operating in parallel, with the final movement being a composite of two different processes. Alternatively, intrinsic reward might directly modulate adaptation, attenuating the trial-to-trial change induced by the SPE (Fig. 4b, Adaptation Modulation model). For example, the reward signal might serve as a gain controller, reducing the rate at which an internal model is updated, attenuating the learning drive.

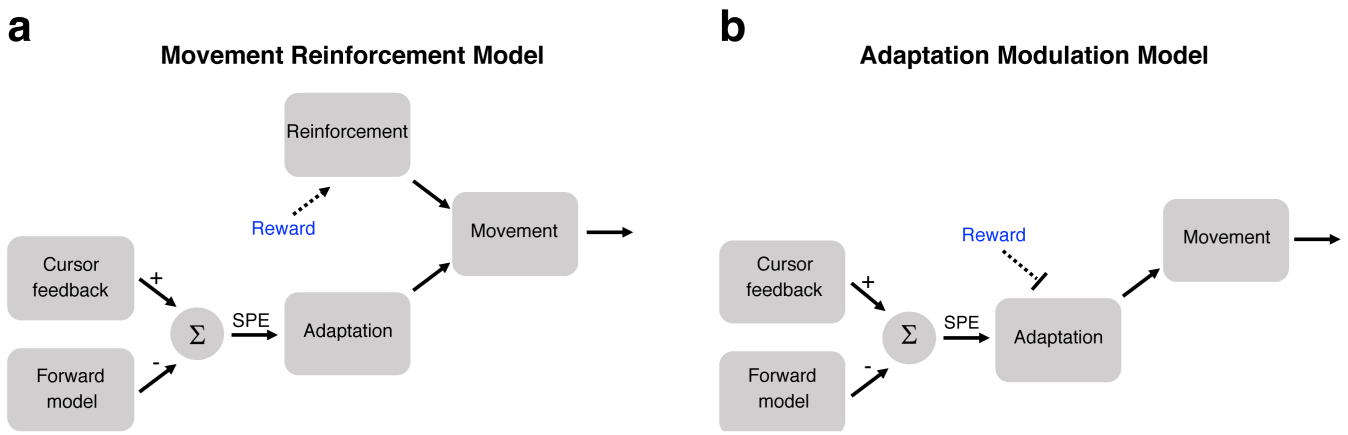


Figure 4 Two models of how intrinsic reward could affect overall learning. **(a)** In the Movement Reinforcement model, reward signals cause reinforcement learning processes to bias future movements towards previously rewarded movements. The adaptation process is sensitive only to SPE and not reward. The overall movement reflects a composite of the two processes. **(b)** In the Adaptation Modulation model, reward directly attenuates adaptation to SPE.

The experimental designa employed in Experiments 1 and 2 cannot distinguish between these two hypotheses because both make similar predictions when the clamp is introduced. In the Movement Reinforcement model, the attenuated asymptote arises because movements are rewarded throughout, including during early learning, biasing future movements towards baseline. The Adaptation Modulation model makes a similar prediction, but here the effect arises because the adaptation system is directly attenuated.

However, a transfer design in which the target size changes after an initial adaptation phase affords an opportunity to contrast the two models. In Experiment 3, we tested a group of participants (n=12) with a 1.75° clamp, using the design depicted in Fig. 5a (Straddle-to-Hit group). In an initial acquisition phase (first 120 clamp cycles), the target was small, such that the clamp always straddled the target. Based on the results of Experiments 1 and 2, we expect to observe a relatively large change in hand angle at the end of this phase. The key test comes during the transfer phase (final 80 clamp cycles), in which the target size is increased such that the invariant clamp now results in a target hit. By the Movement

Reinforcement model, hitting the target will produce an intrinsic reward signal, reinforcing the associated movement. Therefore, there should be no change in performance (hand angle) following transfer: The SPE remains the same, and with the introduction of a reward signal, the executed movements would now be reinforced (Fig. 5b). In contrast, the Adaptation Modulation model assumes that the introduction of the reward signal will directly attenuate the output of the adaptation system. As such, this model predicts a marked decay in hand angles following transfer, relative to the initial asymptote.

In addition to the Straddle-to-Hit group described above, we also tested a second group (n=12) in which the large target (reward) was used in the acquisition phase and the small target (no reward) in the transfer phase (Hit-to-Straddle group). Both models make the same predictions for the Hit-to-Straddle group. At the end of the acquisition phase, there should be a relatively small change in hand angle due to the presence of an intrinsic reward signal. Following transfer, the Movement Reinforcement model predicts that, with the switch to the small target, the intrinsic reward signal will now be absent, weakening the contribution of the reward-based system to the motor output. As such, there should be an increase in hand angle following transfer. The Adaptation Modulation model predicts a similar change in behavior due to the removal of the direct inhibitory effect of the reward system on adaptation following transfer. Although this group in isolation does not discriminate between the models, it does provide a second test of each model, as well as an opportunity to rule out alternative hypotheses for the behavioral effects at transfer. For example, the absence of a change at transfer might be due to reduced sensitivity to the clamp following a long initial acquisition phase. With the Hit-to-Straddle group, both models predict a marked increase in hand angle.

For our analyses, we first examined performance during the acquisition phase (Fig. 5c). Consistent with the results from Experiments 1 and 2, the Hit-to-Straddle Target group adapted slower than the Straddle-to-Hit group (95% CI [-.17°/cycle, -.83°/cycle], $t(22)=-3.15$, $p=.005$, $d=1.29$; Fig. 5d) and

reached a lower asymptote (95% CI [-5.25°, -15.29°], t(22)=-4.24, p=.0003, d=1.73). The reduction at asymptote was approximately 45%.

We next examined performance during the transfer phase where the target size reversed for the two groups. Our primary measure of behavioral change for each subject was the difference in late learning (average hand angle over last 10 cycles) between the end of the acquisition phase and the end of the transfer phase. As seen in Fig. 5c, the two groups showed opposite changes in behavior in the transfer phase, evident by the strong (group x phase) interaction ($F(2,33)=43.1$, $p<10^{-7}$, partial $\eta^2=.72$). The results of a within-subjects t-test showed that the Hit-to-Straddle group showed a marked increase in hand angle following the decrease in target size (95% CI [4.9°, 9.1°], t(11)=7.42, p<.0001, $d_z=2.14$; Fig. 5e), consistent with the predictions for both the Movement Reinforcement and Adaptation Modulation models, assuming that transfer resulted in the removal of the intrinsic reward signal.

The Straddle-to-Hit group's transfer performance provides the critical test of the two hypotheses. Following the switch to the large target, there was a decrease in hand angle. Applying the same statistical test, the mean decrement in hand angle was 5.7° from the final cycles of the training phase to the final cycles of the transfer phase (95% CI [-3.1°, -8.2°], t(11)=-4.84, p=.0005, $d_z=1.40$; Fig. 5e). This result is consistent with the prediction of the Adaptation Modulation model, namely that the introduction of an intrinsic reward signal attenuated the output of the adaptation system. The reduction in hand angle cannot be accounted for by the Movement Reinforcement model. For all participants in both groups, the directional changes in hand angle following transfer were consistent with the predictions of the Adaptation Modulation model (Fig. 5e).

To quantitatively evaluate the Adaptation Modulation model, we simulated the results of the transfer phase of Experiment 3 based on parameters estimated from the acquisition phase of both groups. We fit the data using a single rate state-space model of the following form:

$$x_{n+1} = Ax_n + U(e_n) \qquad \text{[Equation 1]}$$

where *x* represents the motor output on trial *n*, *A* is a retention factor, and *U* represents the update/correction size (or, learning rate) as a function of the error size, *e*. This model is mathematically equivalent to a standard single rate state-space model (Thoroughman & Shadmehr, 2000), with the only modification being the replacement of the error sensitivity term, *B*, with a correction size function. Unlike standard adaptation studies where error size changes over the course of learning, however, *e* is a constant with clamped visual feedback, and therefore *U(e)* can be fit as a single parameter (for further details, see Kim et al. 2018). We refer to this model as the motor correction variant of the standard state space model.

To estimate *A* and *U(e)*, we fit the bootstrapped samples of mean behavior, using only the data from the acquisition phase. The model provided good fits of behavioral change during the acquisition phase (Fig. 5f), with a median r-squared value of .94 (95% CI: [.86, .96]). Parameter estimates for the Hit-to-Straddle group were [.952, .973] for A and [.33°, .62°] for *U(e)* (values represent bootstrapped 95% CIs). For the Straddle-to-Hit group, estimated *A* and *U(e)* values during acquisition were [.939 .970] and [.69° 1.37°], respectively. Using non-parametric permutation tests on the parameter estimates for individual data, reliable differences between groups were observed for *U(e)* (p = .012), but not *A* (p = .802). Thus, the analysis of the parameter estimates indicates that reward modulates the error size-dependent motor correction within the adaptation system, effectively reducing the size of the trial-to-trial correction.

To further test whether the effect of intrinsic reward was better explained by a reduction in learning rate rather than a change in retention, we compared models in which only *U* or *A* were free to vary, asking how well these models fit the bootstrapped samples of the acquisition phase data. For the model in which *U* was a free parameter, we fixed *A* to its median value from the original fits (.96); for the model in which *A* was a free parameter, we fixed *U* to its median value from the original fits (.64). The free *U* model explained, on average, ~8% more of the variance in the data than the free *A* model, and also provided excellent fits of the data (95% CIs for r-squared values: free *U* model [.85, .96]; free *A* model [.61, .94]).

The estimated parameters for each group's acquisition phase data were then used to predict the transfer performance for the other group. That is, parameter estimates from the Hit-to-Straddle group were used to predict the transfer performance of the Straddle-to-Hit group. In a complementary manner, parameter estimates from the Straddle-to-Hit group were used to predict the transfer performance of the Hit-to-Straddle group. We used all 1000 sets of parameter estimates from each group to generate the mean and variance of the predicted behavior (Fig. 5f). During transfer, the model captures the qualitative change in performance for both groups, with an increase in hand angle for the Hit-to-Straddle group and decrease in hand angle for the Straddle-to-Hit group. However, the predictions of the model slightly underestimate the observed rates of change for both groups. We return to this issue in the Discussion; for now, we note that modeling results are consistent with the hypothesis that intrinsic reward directly modulates the adaptation system.

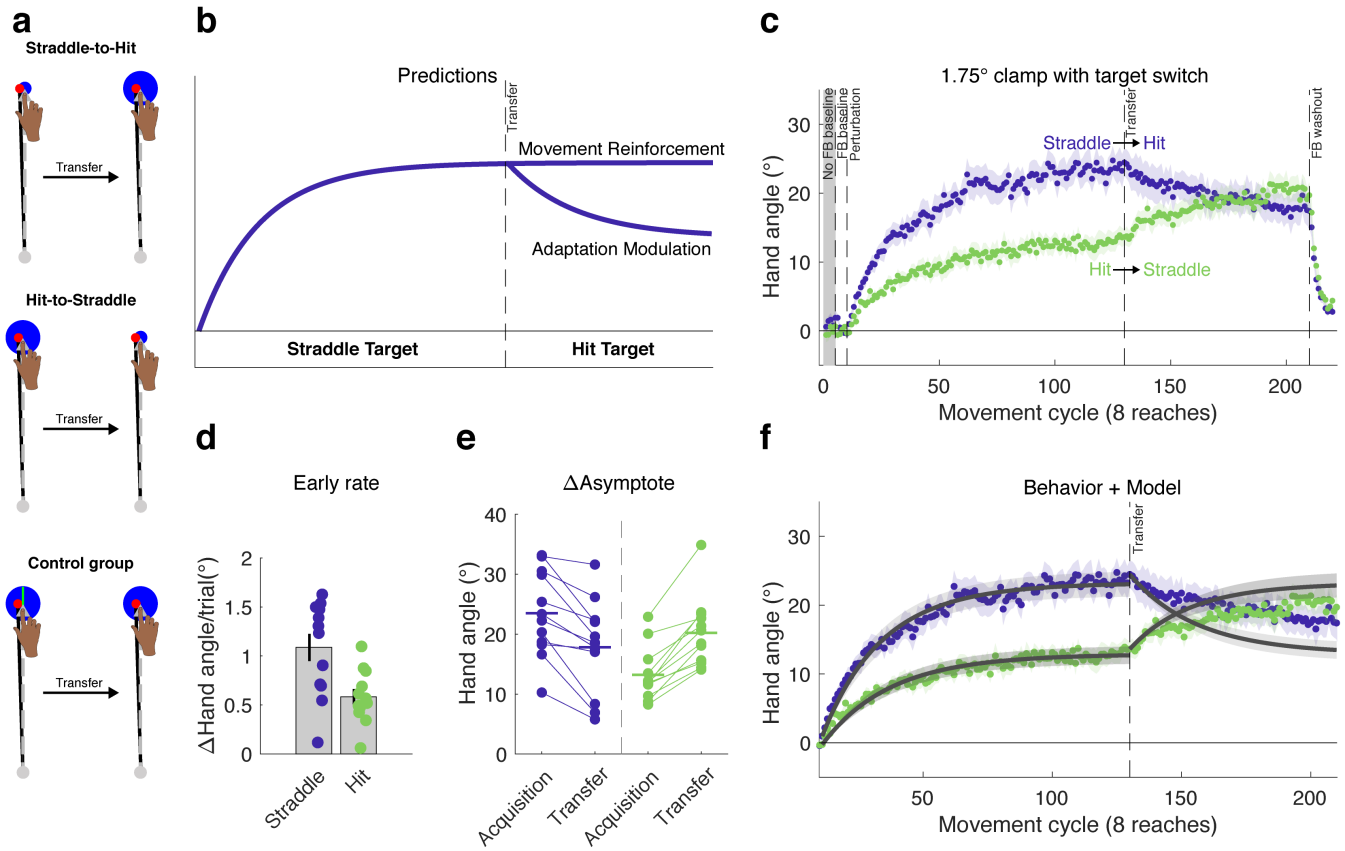


Figure 5 Intrinsic rewards directly modulate the output of the adaptation system. Using a transfer design **(a)**, the two models diverge in their behavioral predictions for the Straddle-to-Hit group following transfer **(b)**. The Movement Reinforcement model predicts a consistent asymptote following transfer, as the learning drive from the SPE is the same and the reward system now reinforces the movement. In contrast, the Adaptation Modulation model predicts an immediate decay in hand angle, given that the intrinsic reward attenuates the output of the adaptation system after transfer. **(c)** The learning functions were consistent with the predictions of the Adaptation Modulation model. The rise in hand angle for the Hit-to-Straddle group is consistent with both models. Note that during the acquisition phase, we again observed differences in the early adaptation rate **(d)** as well as late learning. **(e)** All participants in both groups demonstrated changes in late learning from acquisition to transfer phases that were consistent with Adaptation Modulation model predictions. **(f)** A two-parameter state-space model was able to characterize the learning functions during the acquisition phase. However, the changes in behavior during the transfer phase were somewhat slower than predicted, based on predictions made by switching parameters for both groups. Dots represent individuals; shading and error bars denote SEM.

Control group for testing perceptual uncertainty hypothesis

Across the three experiments, the amount of adaptation induced by clamped visual feedback was attenuated when participants reached to the large target. We considered if this effect could be due, in part, to the differences between the Hit and Straddle/Miss conditions in terms of perceptual uncertainty.

For example, the reliability of the visual error signal might be weaker if the cursor is fully embedded within the target; in the extreme, failure to detect the angular offset might lead to the absence of an SPE on some percentage of the trials.

To evaluate this perceptual uncertainty hypothesis, we tested an additional group in Experiment 3 with a large target, but modified the display such that a bright line, aligned with the target direction, bisected the target (Fig. 5a). With this display, the feedback cursor remained fully embedded in the target, but was clearly off-center. If the attenuation associated with the large target is due to perceptual uncertainty, then the inclusion of the bisecting line should produce an adaptation effect similar to that observed with small targets. Alternatively, if perceptual uncertainty does not play a prominent role in the target size effect, then the adaptation effects would be similar to that observed with large targets.

Consistent with the second hypothesis, performance during the acquisition phase for the group reaching to a bisected target was similar to that of the group reaching to the standard large target (Hit-to-Straddle, see Supplement). Planned pair-wise comparisons showed no significant differences between the two groups (early adapt: 95% CI [-.34°/cycle, .22°/cycle], t(22)=-.47; p=.64,; d=.19; late learning: 95% CI [-7.80° 1.19°], t(22)=-1.52; p=.14; d=.62). In contrast, the group reaching to bisected targets showed slower early adaptation rates (95% CI [-.81°/cycle, -.07°/cycle], t(22)=-2.49, p=.02, d=1.02) and lower magnitudes of late learning (95% CI [-12.58°, -1.35°], t=-2.57, p=0.017, d=1.05) when compared with the group reaching to small targets (Straddle-to-Hit). Given our analysis plan entailed multiple comparisons, we also performed an omnibus one-way ANOVA on the late learning data at the end of the acquisition phase. The effect of group was significant (F(2,33)=9.33, p=.0006, η^2=.36).

During the transfer phase, the target size for the perceptual uncertainty group remained large, but the bisection line was removed. If perceptual uncertainty contributed to the Hit Target effect, we would expect to observe a decrease in hand angle (since uncertainty would increase following transfer). However, following transfer to the non-bisected large target, there was no change in asymptote (95% CI [-.87°, 2.32°], $t(11)=1.0$, $p=.341$, $d_z=.29$). In sum, the results from this control group indicate that the attenuated adaptation observed when the cursor is fully embedded within the target is not due to perceptual uncertainty,

DISCUSSION

The impact of reward on sensorimotor adaptation has been the focus of recent debate and investigation. A number of studies have demonstrated, either through the direct manipulation of reward (Galea et al., 2015; Nikooyan & Ahmed, 2015), or indirectly by varying task outcomes (Leow et al., 2018; Reichenthal, Avraham, Karniel, & Shmuelof, 2016; Jordan A. Taylor & Ivry, 2011), that task success signals can modulate performance changes in sensorimotor adaptation tasks. What remains unclear, however, is how to characterize the interaction of reward-based and error-based learning systems. Based on previous results and modeling work, reward signals have been hypothesized to operate on certain aspects of learning such as consolidation (e.g., Shmuelof et al., 2012 and Galea et al., 2015). Other studies suggest rewards are exploited by learning systems distinct from SPE-driven implicit adaptation (Codol, Holland, & Galea, 2018), with the resulting performance a composite of changes resulting from the independent operation of these different systems (Jordan A. Taylor & Ivry, 2011; Jordan A. Taylor et al., 2014). The interpretation of the results from these studies is complicated by the fact that the experimental tasks conflate different learning processes. In the present study, we sought to avoid this complication by using a new method to study adaptation, one in which performance changes arise implicitly in response to an invariant visual error signal.

Using the visual clamp method (Morehead et al., 2017), we observed a striking difference between conditions in which the final position of the cursor was fully embedded in the target compared to conditions in which the cursor either terminated outside or straddled the target: When the cursor was fully embedded, the rate of learning was reduced and, more strikingly, the asymptotic level of learning was attenuated. Interestingly, the effect of varying the target size was qualitatively different than what we observed in previous studies in which we varied the angular direction of the clamp. In that work, small clamp angles reduced the rate of adaptation (Kim et al., 2018), but, over a large range of values, failed to produce reliable differences in asymptotic levels of learning (Kim et al., 2018; Morehead et al., 2017).

The difference in behavioral change as a function of relative target size was observed across different clamp sizes and did not appear to be because of differences in perceptual sensitivity or motor competence. This was supported by our control analyses, perceptual control experiment, and our finding that the Straddle group in Experiment 1 was similar to the Hit group, suggesting that the effect of target size was categorical. As such, we assume that the effect of target size on behavior arises from the generation of an intrinsic reward signal, one that is generated when the cursor lands fully within the target. In the final experiment, we explored two ways in which an intrinsic reward signal could impact performance. One hypothesis centered on the idea that reward modulates the strength of movement representations associated with task success, a variant of the idea that reward and SPE engage distinct representations and learning systems (Shmuelof et al., 2012). The other hypothesis considered a more direct modulatory impact on the adaptation process. The results showed that the differences in asymptote cannot be attributed solely to strengthening of rewarded movements. Rather, intrinsic reward directly attenuates the operation of the adaptation system.

We recognize that our interpretation of the results rests on the assumption that “hitting” the target with the cursor is intrinsically rewarding (Huang et al., 2011; Leow et al., 2018). If correct, this assumption

437 holds despite the participants' awareness that the angular motion of the cursor is causally unrelated to
438 their behavior. Our earlier work with clamped feedback had shown that adaptation can be driven by a
439 task-irrelevant error signal, the SPE defined by the difference between the cursor and target. Here we
440 see the automatic operation of an intrinsic reward signal. Of course we do not have evidence,
441 independent of the behavior, that hitting a target is rewarding; this might require using methods such as
442 fMRI (Daw, Gershman, Seymour, Dayan, & Dolan, 2011) or pupillometry (Manohar, Finzi, Drew, &
443 Husain, 2017) to assess the presence of well-established signatures of reward.
444
445 State space models have provided a concise computational account of sensorimotor adaptation
446 (Huang, Haith, Mazzoni, & Krakauer, 2011; Smith, Ghazizadeh, & Shadmehr, 2006; Tanaka, Krakauer,
447 & Sejnowski, 2012; Thoroughman & Shadmehr, 2000). In the simplest version, these models entail two
448 parameters, a memory term, *A*, representing the retention of the current state from trial to trial, and a
449 learning rate term, *B*, representing how the state is updated based on the error from the current trial
450 (the *A* and *U(e)* terms in Eq. 1, respectively). Given this framework, we can consider how reward might
451 modulate adaptation. One possibility is that reward modulates retention. This hypothesis is consistent
452 with the results of a recent visuomotor adaptation study comparing groups that either received only
453 cursor feedback or cursor feedback and a monetary reward, scaled to their accuracy. The latter showed
454 greater retention during a washout block in which the feedback was removed (Galea et al., 2015).
455 When the data were fit with the standard state space model, this effect was accounted for by an
456 increase in the retention term, *A*, interpreted as indicating that rewarded movements are better
457 consolidated.
458
459 A retention-based account, however, does not accord well with the current results. If the memory term
460 was larger in conditions with intrinsic reward (i.e., Hit Target conditions), then we should have observed
461 a higher asymptote when the cursor was embedded in the target compared to when it missed (or
462 straddled) the target, since the SPE is invariant and more of the current state is retained from trial to

trial. The behavior went in the opposite direction of this prediction: The Hit Target conditions consistently resulted in lower asymptotic values. Thus, a retention-based account of the intrinsic reward effect would mandate lower values of *A*, a situation in which the memory term results in the adaptation system being resistant to learning from errors. We suspect that the washout results observed in Galea et al. (2015) are not due to a change in the adaptation process, but rather reflect the residual effects of an aiming strategy induced by the reward. That is, the monetary rewards might have reinforced a strategy during the rotation block, and this carried over into the washout block. Indeed, the idea that reward impacts strategic processes has been advanced in studies comparing conditions in which the performance could be enhanced by re-aiming (Codol et al., 2018; Holland et al., 2018).

Alternatively, intrinsic reward could influence overall learning by modulating the learning rate parameter. A priori, one might suppose that reward would enhance learning (Nikooyan & Ahmed, 2015), either by increasing the sensitivity and responsiveness to error, or by promoting exploratory behavior to generate appropriate compensatory strategies. The latter would be a case where the learning rate parameter encompasses the effects of both implicit and explicit learning processes, especially relevant in standard adaptation studies where the task outcome is contingent on the participant's behavior and the perturbation is large (Bond & Taylor, 2015; Jordan A. Taylor & Ivry, 2011; Jordan A. Taylor et al., 2014).

In contrast, the clamp method, by eliminating the contribution of strategic processes, allows us to directly examine how reward might influence estimated rates of implicit learning. Here we see that the effect would suggest that reward reduces the learning rate, made salient by the parameter estimates from the acquisition phase of Experiment 3 (see also, Leow et al., 2018). A reduction in the learning rate can be conceptualized as a gain factor attenuating the system's response to error. In terms of the standard state space model, this would translate into reducing the system's sensitivity to error; in the motor correction variant of the state space model, this would translate into reducing the amount of

change induced by an error of a given size. In either conceptualization, the end result is that in the presence of an intrinsic reward signal, the error-dependent drive is reduced.

The hypothesis that reward attenuates the learning rate within the adaptation system provides a parsimonious account of the data from all three experiments. Following the introduction of the clamped feedback, a lower asymptote was observed in Experiments 1-3 when the cursor hit the target. Assuming the memory process is unaffected, the reduced error-dependent drive will result in a lower asymptote. Moreover, the rate of change in behavior, operationalized here by the early learning rate, should also be lower, a pattern evident in all three experiments, although only statistically significant in Experiment 3. Moreover, a change in learning drive can account for the behavioral effects observed in the transfer phase of Experiment 3. The loss of an intrinsic reward signal (Hit-to-Straddle group) would increase the error-dependent learning drive, resulting in an increase in hand angle. Conversely, the introduction of an intrinsic reward signal (Straddle-to-Hit group) would decrease the learning drive, resulting in a drop in hand angle.

Although the results indicate that reward directly modulates adaptation, the observed changes in behavior during the transfer phase of Experiment 3 were more gradual than predicted based on parameter estimates derived from the initial acquisition phase data. The quantitative predictions here assume that the adaptation system is time-invariant. This assumption may be too rigid; for example, learning parameters may change with increased exposure to a perturbation (Mawase, Shmuelof, Bar-Haim, & Karniel, 2014; Zarahn, Weston, Liang, Mazzoni, & Krakauer, 2008) or change as a function of the context (Herman, Harwood, & Wallman, 2009).

We also recognize that behavioral changes here may reflect the operation of multiple processes (Krakauer & Mazzoni, 2011), and the composite effects of these processes might account for why the observed changes were more gradual than predicted. For example, intrinsic reward may not only

directly modulate adaptation, but may also reinforce an executed movement (Castro, Monsen, & Smith, 2011), a combination of the Movement Reinforcement and Adaptation Modulation models sketched in Figure 4. For example, in the Straddle-to-Hit condition, the introduction of intrinsic reward at transfer would reinforce movements at the initial asymptote, resisting the effect of reduced learning drive.

Studies involving non-human primates and rodents have provided insights into possible neural substrates supporting the interaction of systems involved in reward- and error-based learning. Converging evidence points to a critical role for the cerebellum in sensorimotor adaptation (Butcher et al., 2017; Izawa, Criscimagna-Hemminger, & Shadmehr, 2012; J A Taylor, Klemfuss, & Ivry, 2010; Tseng et al., 2007), including the observation that patients with cerebellar degeneration show a reduced response to visual error clamps (Morehead et al., 2017). Reward-based learning is associated with a more distributed network of cortical and subcortical areas, including a prominent role for dopaminergic signals in the basal ganglia (Schultz, 2015). Neuroanatomical studies have identified di-synaptic reciprocal connections between the basal ganglia and cerebellum (Bostan, Dum, & Strick, 2010), as well as direct connections between the cerebellum and dopaminergic nuclei in the brainstem (Perciavalle, Berretta, & Raffaele, 1989; Watabe-Uchida, Zhu, Ogawa, Vamanrao, & Uchida, 2012). These connections might provide a relatively direct pathway for reward signals to modulate cerebellar activity. Alternatively, or perhaps complementary, recent work has indicated that both simple (Wagner, Kim, Savall, Schnitzer, & Luo, 2017) and complex (Ohmae & Medina, 2015) spike activity in the cerebellum may signal information about rewards or anticipated rewards. This work suggests a more expansive view may be required to understand cerebellar function, one in which error-based learning is modulated by contextual factors. The current study provides a striking example of how intrinsic reward signals, in the form of persistent target hits, may serve as one such contextual factor that can modulate cerebellar-dependent sensorimotor adaptation.

ACKNOWLEDGMENTS

We thank Matthew Hernandez and Wendy Shwe for assistance with data collection. We are also grateful to Maurice Smith, Ryan Morehead, Guy Avraham, and Ian Greenhouse for helpful discussions regarding this work.

Competing Interests

No competing interests, financial or otherwise, are declared by the authors.

REFERENCES

Bond, K. M., & Taylor, J. A. (2015). Flexible explicit but rigid implicit learning in a visuomotor adaptation task. *Journal of Neurophysiology*, *113*(10), 3836–3849. https://doi.org/10.1152/jn.00009.2015

Bostan, A. C., Dum, R. P., & Strick, P. L. (2010). The basal ganglia communicate with the cerebellum. *Proceedings of the National Academy of Sciences*, *107*(18), 8452–8456. https://doi.org/10.1073/pnas.1000496107

Butcher, P. A., Ivry, R. B., Kuo, S.-H., Rydz, D., Krakauer, J. W., & Taylor, J. A. (2017). The cerebellum does more than sensory prediction error-based learning in sensorimotor adaptation tasks. *Journal of Neurophysiology*, *118*(3), 1622–1636. https://doi.org/10.1152/jn.00451.2017

Castro, L. N. G., Monsen, C. B., & Smith, M. A. (2011). The Binding of Learning to Action in Motor Adaptation. *PLOS Computational Biology*, *7*(6), e1002052. https://doi.org/10.1371/journal.pcbi.1002052

Churchland, M. M., Afshar, A., & Shenoy, K. V. (2006). A Central Source of Movement Variability. *Neuron*, *52*(6), 1085–1096. https://doi.org/10.1016/j.neuron.2006.10.034

Codol, O., Holland, P. J., & Galea, J. M. (2017). The relationship between reinforcement and explicit strategies during visuomotor adaptation. *BioRxiv*, 206284.

Codol, O., Holland, P. J., & Galea, J. M. (2018). The relationship between reinforcement and explicit control during visuomotor adaptation. *Scientific Reports*, *8*(1), 9121. https://doi.org/10.1038/s41598-018-27378-1

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Fitts, P. M. (1992). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General*, *121*(3), 262–269. https://doi.org/10.1037/0096-3445.121.3.262

Galea, J. M., Mallia, E., Rothwell, J., & Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nature Neuroscience*, *18*(4), 597–602. https://doi.org/10.1038/nn.3956

Gallivan, J. P., Logan, L., Wolpert, D. M., & Flanagan, J. R. (2016). Parallel specification of competing sensorimotor control policies for alternative action options. *Nature Neuroscience*, *19*(2), 320–326. https://doi.org/10.1038/nn.4214

Gordon, J., Ghilardi, M. F., & Ghez, C. (1994). Accuracy of planar reaching movements - I. Independence of direction and extent variability. *Experimental Brain Research*, *99*(1), 97–111. https://doi.org/10.1007/BF00241415

Hadjiosif, A., & Smith, M. (2013). Savings is restricted to the temporally labile component of motor adaptation. In *Translational and Computational Motor Control (TCMC)*. San Diego, CA.

Holland, P. J., Codol, O., & Galea, J. M. (2018). The contribution of explicit processes to reinforcement-based motor learning. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00901.2017

Huang, V. S., Haith, A., Mazzoni, P., & Krakauer, J. W. (2011). Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. *Neuron*, *70*(4), 787–801. https://doi.org/10.1016/j.neuron.2011.04.012

Huberdeau, D. M., Krakauer, J. W., & Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Curr Opin Neurobiol*, *33*, 71–77. https://doi.org/10.1016/j.conb.2015.03.003

Izawa, J., Criscimagna-Hemminger, S. E., & Shadmehr, R. (2012). Cerebellar contributions to reach adaptation and learning sensory consequences of action. *J Neurosci*, *32*(12), 4230–4239. https://doi.org/10.1523/jneurosci.6353-11.2012

Izawa, J., & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput Biol*, *7*(3), e1002012–e1002012. https://doi.org/10.1371/journal.pcbi.1002012

Kim, H. E., Morehead, J. R., Parvin, D. E., Moazzezi, R., & Ivry, R. B. (2018). Invariant errors reveal limitations in motor correction rather than constraints on error sensitivity. *Communications Biology*, *1*(1), 19. https://doi.org/10.1038/s42003-018-0021-y

Knill, D. C., Bondada, A., & Chhabra, M. (2011). Flexible, Task-Dependent Use of Sensory Feedback to Control Hand Movements. *Journal of Neuroscience*, *31*(4), 1219–1237. https://doi.org/10.1523/JNEUROSCI.3522-09.2011

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Krakauer, J. W. (2005). Adaptation to Visuomotor Transformations: Consolidation, Interference, and Forgetting. *Journal of Neuroscience*, *25*(2), 473–478. https://doi.org/10.1523/JNEUROSCI.4218-04.2005

Krakauer, J. W., & Mazzoni, P. (2011). Human sensorimotor learning: adaptation, skill, and beyond. *Curr Opin Neurobiol*, *21*(4), 636–644. https://doi.org/10.1016/j.conb.2011.06.012

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Leow, L.-A., Marinovic, W., & Carroll, T. (2018). Task errors contribute to implicit remapping in sensorimotor adaptation. *BioRxiv*, 263988.

Manohar, S. G., Finzi, R. D., Drew, D., & Husain, M. (2017). Distinct Motivational Effects of Contingent and Noncontingent Rewards. *Psychological Science*, *28*(7), 1016–1026. https://doi.org/10.1177/0956797617693326

Mawase, F., Shmuelof, L., Bar-Haim, S., & Karniel, A. (2014). Savings in locomotor adaptation explained by changes in learning parameters following initial adaptation. *Journal of Neurophysiology*, *111*(7), 1444–1454. https://doi.org/10.1152/jn.00734.2013

McDougle, S. D., Ivry, R. B., & Taylor, J. A. (2016). Taking Aim at the Cognitive Side of Learning in Sensorimotor Adaptation Tasks. *Trends in Cognitive Sciences*, *20*(7), 535–544. https://doi.org/10.1016/j.tics.2016.05.002

Morehead, J. R., Taylor, J. A., Parvin, D. E., & Ivry, R. B. (2017). Characteristics of Implicit Sensorimotor Adaptation Revealed by Task-irrelevant Clamped Feedback. *Journal of Cognitive Neuroscience*, *29*(6), 1061–1074. https://doi.org/10.1162/jocn_a_01108

Nikooyan, A. A., & Ahmed, A. A. (2015). Reward feedback accelerates motor learning, 633–646. https://doi.org/10.1152/jn.00032.2014

Ohmae, S., & Medina, J. F. (2015). Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. *Nature Neuroscience*, *18*(12), 1798–1803. https://doi.org/10.1038/nn.4167

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Perciavalle, V., Berretta, S., & Raffaele, R. (1989). Projections from the intracerebellar nuclei to the ventral midbrain tegmentum in the rat. *Neuroscience*, *29*(1), 109–119. https://doi.org/10.1016/0306-4522(89)90336-9

Reichenthal, M., Avraham, G., Karniel, A., & Shmuelof, L. (2016). Target size matters: target errors contribute to the generalization of implicit visuomotor learning. *Journal of Neurophysiology*, *116*(2), 411–424. https://doi.org/10.1152/jn.00830.2015

Schultz, W. (2015). Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev*, *95*(3), 853–951. https://doi.org/10.1152/physrev.00023.2014

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annu Rev Neurosci*, *33*, 89–108. https://doi.org/10.1146/annurev-neuro-060909-153135

Shmuelof, L., Huang, V. S., Haith, A. M., Delnicki, R. J., Mazzoni, P., & Krakauer, J. W. (2012). Overcoming Motor “Forgetting” Through Reinforcement Of Learned Actions. *Journal of Neuroscience*, *32*(42), 14617-14621a. https://doi.org/10.1523/JNEUROSCI.2184-12.2012

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting Adaptive Processes with Different Timescales Underlie Short-Term Motor Learning. *PLOS Biology*, *4*(6), e179. https://doi.org/10.1371/journal.pbio.0040179

Soechting, J. F. (1984). Effect of target size on spatial and temporal characteristics of a pointing movement in man. *Experimental Brain Research*, *54*(1), 121–132. https://doi.org/10.1007/BF00235824

Tanaka, H., Krakauer, J. W., & Sejnowski, T. J. (2012). Generalization and Multirate Models of Motor Adaptation. *Neural Computation*, *24*(4), 939–966. https://doi.org/10.1162/NECO_a_00262

Taylor, J A, Klemfuss, N. M., & Ivry, R. B. (2010). An explicit strategy prevails when the cerebellum fails to compute movement errors. *Cerebellum*, *9*(4), 580–586. https://doi.org/10.1007/s12311-010-0201-x

Taylor, Jordan A., & Ivry, R. B. (2011). Flexible Cognitive Strategies during Motor Learning. *PLoS Computational Biology*, *7*(3), e1001096. https://doi.org/10.1371/journal.pcbi.1001096

Taylor, Jordan A., & Ivry, R. B. (2014). Cerebellar and Prefrontal Cortex Contributions to Adaptation, Strategies, and Reinforcement Learning. In *Progress in Brain Research* (Vol. 210, pp. 217–253). Elsevier. https://doi.org/10.1016/B978-0-444-63356-9.00009-1

Taylor, Jordan A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and Implicit Contributions to Learning in a Sensorimotor Adaptation Task. *Journal of Neuroscience*, *34*(8), 3023–3032. https://doi.org/10.1523/JNEUROSCI.3619-13.2014

Therrien, A. S., Wolpert, D. M., & Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*, *139*(1), 101–114. https://doi.org/10.1093/brain/awv329

Therrien, A. S., Wolpert, D. M., & Bastian, A. J. (2018). Increasing Motor Noise Impairs Reinforcement Learning in Healthy Individuals. *Eneuro*, ENEURO.0050-18.2018. https://doi.org/10.1523/ENEURO.0050-18.2018

Thoroughman, K. A., & Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, *407*(6805), 742–747. https://doi.org/10.1038/35037588

Tseng, Y. W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J Neurophysiol*, *98*(1), 54–62. https://doi.org/10.1152/jn.00266.2007

Vandevoorde, K., & Orban de Xivry, J.-J. (2018). Motor adaptation but not internal model recalibration declines with aging. *BioRxiv*. https://doi.org/10.1101/292250

Vaswani, P. A., Shmuelof, L., Haith, A. M., Delnicki, R. J., Huang, V. S., Mazzoni, P., … Krakauer, J. W. (2015). Persistent Residual Errors in Motor Adaptation Tasks: Reversion to Baseline and Exploratory Escape. *Journal of Neuroscience*, *35*(17), 6969–6977. https://doi.org/10.1523/JNEUROSCI.2656-14.2015

Wagner, M. J., Kim, T. H., Savall, J., Schnitzer, M. J., & Luo, L. (2017). Cerebellar granule cells encode the expectation of reward. *Nature*. https://doi.org/10.1038/nature21726

Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A., & Uchida, N. (2012). Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons. *Neuron*, *74*(5), 858–873. https://doi.org/10.1016/j.neuron.2012.03.017

Xu-Wilson, M., Zee, D. S., & Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research*, *196*(4), 475–481. https://doi.org/10.1007/s00221-009-1879-1

Zarahn, E., Weston, G. D., Liang, J., Mazzoni, P., & Krakauer, J. W. (2008). Explaining Savings for Visuomotor Adaptation: Linear Time-Invariant State-Space Models Are Not Sufficient. *Journal of Neurophysiology*, *100*(5), 2537–2548. https://doi.org/10.1152/jn.90529.2008

METHODS

Participants: Healthy, young adults (N=116, 69 females, age = 20.9 ± 2.1 years old) were recruited from the University of California, Berkeley, community. Each participant was tested in only one experiment. All participants were right-handed, as verified with the Edinburgh Handedness Inventory (Oldfield, 1971). Participants provided informed consent and received financial compensation for their participation. The Institutional Review Board at UC Berkeley approved all experimental procedures.

Experimental Apparatus: The participant was seated at a custom-made tabletop housing an LCD screen (53.2 cm by 30 cm, ASUS), mounted 27 cm above a digitizing tablet (49.3 cm by 32.7 cm, Intuos 4XL; Wacom, Vancouver, WA). The participant made reaching movements by sliding a modified air hockey "paddle" containing an embedded stylus. The position of the stylus was recorded by the tablet at 200 Hz. The experimental software was custom written in Matlab, using the Psychtoolbox extensions[26].

Reaching Task: Center-out planar reaching movements were performed from the center of the workspace to targets positioned at a radial distance of 8 cm. Direct vision of the hand was occluded by the monitor, and the lights were extinguished in the room to minimize peripheral vision of the arm. The start location and target location were indicated by white and blue circles, respectively (start circle: 6 mm in diameter; target: either 6, 9.8 or 16 mm depending on condition).

To initiate each trial, the participant moved the digitizing stylus into the start location. The position of the stylus was indicated by a white feedback cursor (3.5 mm diameter). Once the start location was maintained for 500 ms, the target appeared. For Experiments 1 and 3, the target could appear at one of 8 locations, placed in 45° increments around a virtual circle (0°, 45°, 95°, 135°, 180°, 225°, 270°, 315°). For Experiment 2, the target could appear at one of four locations placed in 90° increments around a virtual circle (45°, 135°, 225°, 315°). We reduced the number of targets from 8 to 4 in this experiment in

order to increase the overall number of training cycles with the clamp, while keeping the experiment under 1.5 hours, and so that participants would reach a stable asymptote. Participants were instructed to accurately and rapidly "slice" through the target, without needing to stop at the target location. Visual feedback, when presented, was provided during the reach until the movement amplitude exceeded 8 cm. As described below, the feedback either matched the position of the stylus (veridical) or followed a fixed path (clamped). If the movement was not completed within 300 ms, the words “too slow” were generated by the sound system of the computer.

After the hand crossed the target ring, endpoint cursor feedback was provided for 50 ms either at the position in which the hand crossed the virtual target ring (veridical feedback) or at a fixed distance determined by the size of the clamp. During the return movement, the feedback cursor reappeared when the participant’s hand was within 1 cm of the start.

Experimental Feedback Conditions: Across the experimental session, there were three types of visual feedback. On no-feedback trials, the cursor disappeared when the participant‘s hand left the start circle and only reappeared at the end of the return movement. On veridical feedback trials, the cursor matched the position of the stylus during the 8 cm outbound segment of the reach. On clamped feedback trials, the feedback followed a path that was fixed along a specific hand angle. The radial distance of the cursor from the start location was still based on the radial extent of the participant's hand during the 8 cm outbound segment, but the angular position was fixed relative to the target (i.e., independent of the angular position of the hand).

The primary instructions to the participant remained the same across the experimental session: Specifically, that they were to reach directly towards the visual target. Prior to the introduction of the clamped feedback trials, participants were briefed about the feedback manipulation. They were informed that the position of the cursor would now follow a fixed trajectory and that the angular position

would be independent of their movement. They were explicitly instructed to ignore the cursor and continue to reach directly to the target. Participants also performed three instructed trials with the clamp perturbation on. During these practice trials, a target appeared at the 90 deg location (straight ahead), and the experimenter instructed the participant to first “reach straight to the left” (ie, 180 deg). For the second practice trial, the participant was instructed to “reach straight to the right” (0 deg). For the last trial, the participant was instructed to “reach straight down (towards your torso)” (ie, 270 deg). The purpose of these trials was to familiarize the participant with the exact clamp condition they were about to experience. Following these three practice trials, the experimenter confirmed with the participant they understood now what was meant by clamped visual feedback. These practice trials were removed from future analyses.

The same instructions in abbreviated form (“Ignore the cursor and move your hand directly to the target location”) were repeated verbally and with onscreen text at every block break during the clamp perturbation. Participants were debriefed at the end of the experiment and asked whether they ever intentionally tried to reach to locations other than the target. All subjects reported aiming to the target throughout the experiment.

We counterbalanced clockwise and counterclockwise clamp offsets within each group for all three experiments.

Experiment 1

Participants (n=48, 16/group) were randomly assigned to one of three groups, each training with a 3.5° clamp but differing only in terms of the size of the target: 6mm, 9.8, or 16 mm diameter. These sizes were chosen so that at an 8 cm radial distance the clamped cursor would be adjacent to the target without making any contact (Target Miss group), straddling the target by being roughly half inside and half outside the target (Straddle Target group), or fully embedded within the target (Hit Target group).

The Euclidean distance for this clamp offset, measured from the centers of cursor and target, was 4.9 mm.

The session began with two baseline blocks, the first comprised of 5 movement cycles (40 total reaches to 8 targets) without visual feedback and the second comprised of 10 cycles with a veridical cursor displaying hand position. The experimenter then informed the participant that the visual feedback would no longer be veridical and would now be clamped at a fixed angle from the target location. Immediately following these general instructions, the experimenter continued providing instructions for the three practice trials which immediately followed (see Experimental Feedback Conditions). After the practice trials and confirming the participant's understanding of the task, the clamp block ensued for a total of 80 cycles. A short break (<1 min), as well as a reminder of the task instructions, was provided after 40 cycles (i.e., at the halfway point of this block). Immediately following the perturbation block, there were two washout blocks, first a 5 cycle block in which there was no visual feedback, followed by 10 cycles with veridical visual feedback. These blocks were preceded by instructions regarding the change in experimental condition and participants were reminded to always aim for the target and to attempt to slice through it with their hand.

Experiment 2

In Experiment 2 we assessed adaptation over an extended number of clamped visual feedback trials. The purpose of extending the perturbation block was to ensure that participants reached asymptotic levels of learning. In order to achieve a greater number of training cycles, we reduced the number of target locations within the set from 8 to 4.

Participants (n=32, 16/group) trained with a 1.75° clamp (2.4 mm distance between target and cursor centers) and were assigned to either a small (Straddle) or large (Hit) target condition. The session started with two baseline blocks, 10 cycles (40 reaches) without visual feedback and then 10 cycles

with veridical feedback. Following 3 practice trials with the clamp, the number of cycles in the clamped visual feedback block was nearly tripled from that of Experiment 1 to 220 cycles, with breaks provided after every 70 cycles. Following 220 cycles of training with a 1.75° clamp, there were two washout blocks, first a 10 cycle block in which there was a 0° clamp, followed by 10 cycles with veridical visual feedback. Prior to washout, participants were again instructed to always aim directly to the target.

Experiment 3

We assume that with the large targets, an intrinsic reward is generated when the cursor lands within the target. This could serve as a positive reinforcement signal, strengthening the representation of rewarded movements, and operating as a resistance to the learning drive associated with an SPE (Movement Reinforcement model). Alternatively, intrinsic reward may directly modulate the output of the adaptation system (Adaptation Modulation model). As a test of these hypotheses, we tested two main groups (n=12/group) in Experiment 3, using a 1.75° clamp in a transfer design. The session started with two baseline blocks, 5 cycles (40 reaches) without visual feedback and then 5 cycles with veridical feedback. After the baseline blocks, clamp instructions and three practice trials were provided to all participants. The first clamp block lasted 120 cycles, with participants training with either a small or large target. Following the first 120 cycles, the target sizes were reversed for the next 80 (Straddle-to-Hit or Hit-to-Straddle conditions). Our main predictions focused on the transfer phase, comparing the behavior to the predictions of both Movement Reinforcement and Adaptation Modulation models. Breaks of < 1 min were provided after every 35 cycles of training. On the break preceding the transfer (15 cycles before target switch), participants were told that everything would continue on as before, except that the target size would change at some point during the block. The purpose of staggering the break with the transfer was to mitigate any change in adaptation due to temporal decay that could result from a break in training (Hadjiosif & Smith, 2013).

Control group

A third group (n=12) was added to test whether the attenuation of adaptation in the large target condition was due to perceptual uncertainty. Here, the block structure was identical to the first two groups. We used a modified large target (16mm), one which had a bright green bisecting line through the middle, aligned with the target direction. The clamped cursor always fell within one half of the target (either clockwise or counter-clockwise depending on the condition), thus providing a clear indication that the cursor was off center. At the transfer, the bisecting line was removed and participants trained for 80 cycles with the standard large target.

Data Analysis

All statistical analyses and modeling were performed using Matlab 2015b and the Statistics Toolbox. The primary dependent variable in all experiments was hand angle at peak radial velocity, defined by the angle of the hand relative to the target at the time of peak radial velocity (i.e., angle between lines connecting start position to target and start position to hand). Throughout the text, we refer to this variable as hand angle. Additional analyses were performed using hand angle at “endpoint” (angle of the hand as it crossed the invisible target ring) rather than peak radial velocity. The results were essentially identical for the two dependent variables; as such, we only report the results of the analyses using peak radial velocity.

Outlier responses were removed from the analyses. For the sole purpose of identifying outliers, the Matlab “smooth” function was used to calculate a moving average (using a 5-trial window) of the hand angle data for each target location. Outliers were trials in which the observed hand angle was greater than 90° or deviated by more than 3 standard deviations from the moving average. In total, less than 0.8% of trials overall were removed, and the most trials removed for any individual across all three experiments was 2%.

Individual baseline biases for each target location were subtracted from all data. Biases were defined as the average hand angles across cycles 2-10 (Experiments 1 and 2) or 2-5 (Experiment 3) of the feedback baseline block. These same cycles were used to calculate mean RTs, MTs, and movement variability (SD). To calculate each participant's baseline RT or MT, we took the average of median values at each target location. To calculate each participant's movement variability, we took the average of the standard deviations of hand angles at each target location.

In order to pool all of the data and to aid visualization, we flipped the hand angles for all participants clamped in the counterclockwise direction.

For Experiments 1 and 3, movement cycles consisted of 8 consecutive reaches (1 reach/target); for Experiment 2, we only used four targets, thus a movement cycle consisted of 4 consecutive reaches (1 reach/target). Early adaptation rate was quantified by averaging the hand angle values over cycles 3-7 of the clamp, and dividing by the number of cycles (i.e., 5) to get an estimate of the per trial rate of change in hand angle. We opted to use this measure of early adaptation rather than obtain parameter estimates from exponential fits since the latter approach gives considerable weight to the asymptotic phase of performance and, therefore would be less sensitive to early differences in rate. This would be especially problematic in Experiment 2, which utilized 220 clamp cycles. Asymptotic adaptation was defined as the last 10 cycles within a clamp block. In Experiment 1, the aftereffect was quantified by using the data from the first no-feedback cycle following the last clamp cycle. We also performed a secondary analysis of early adaptation rates using cycles 2-11 (Krakauer, 2005), rather than 3-7. Results from using this alternate metric were consistent with the reported analyses (i.e., slower rates for Hit Target groups), only they resulted in larger effect sizes due to the gradually increasing divergence of learning functions.

All t-tests were two-tailed. Posthoc pairwise comparisons following significant ANOVAs were performed using two-tailed t-tests. Cohen's d, eta squared (η^2), partial eta squared (for mixed model ANOVA), and d_z (for within-subjects design) values are provided as standardized measures of effect size (Lakens, 2013). Values in main text are reported as 95% CIs in brackets and mean ± SEM.

Modeling

For our model fitting and simulation procedures we applied standard bootstrapping techniques, constructing group-averaged hand angle data 1000 times by randomly resampling with replacement from the pool of participants within each group. Using Matlab's *fmincon* function, we started with ten different initial sets of parameter values and estimated the retention and learning parameters which minimized the least squared error between the bootstrapped data and model output (x_n).

We also fit the model to the acquisition phase data of each participant in Experiment 3 in order to compare parameter values between groups using a non-parametric permutation test. We first calculated our two test statistics, the average difference in *A* values and the average difference in *U* values between groups. Then, we randomly shuffled the group assignments using 10000 Monte Carlo simulations in order to create the null distributions for mean *A* and *U* parameter values, separately. We then calculated exact p-values by summing the proportion of each respective null distribution that was at least as or more extreme than our test statistic values (i.e., using 2-sided tests).

No statistical methods were used to predetermine sample sizes. The chosen sample sizes were based on our previous study using the clamp method (Kim et al., 2018; Morehead et al., 2017), as well as prior psychophysical studies of human sensorimotor learning (Galea et al., 2015; Gallivan, Logan, Wolpert, & Flanagan, 2016; Huang et al., 2011; Vaswani et al., 2015).

SUPPLEMENTAL INFORMATION

Kinematic variables

Experiment 1

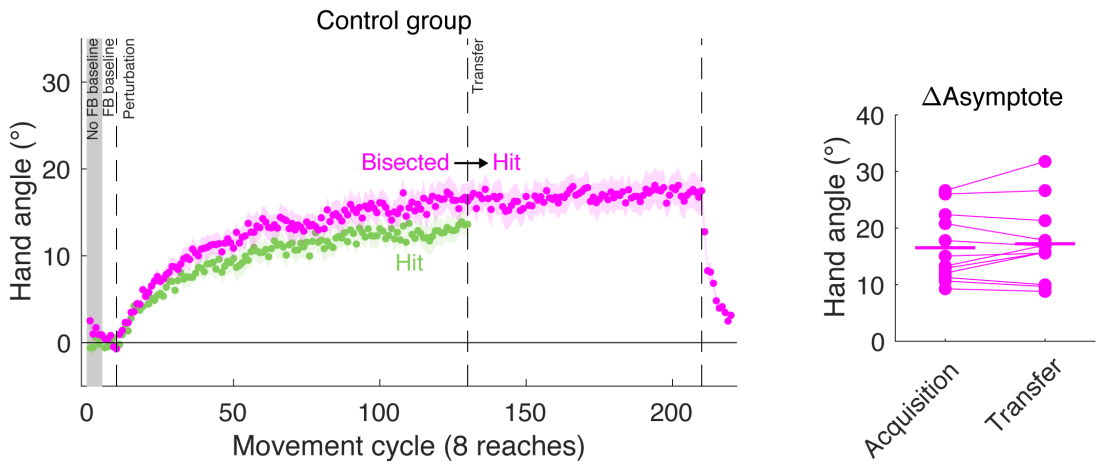
Average movement variability across the eight targets during cycles 2-10 of the veridical feedback baseline block were not different between groups (variability: F(2,45)=0.2.32, p=.110, η^2=.093). Movement times across groups were not different (F(2,45)=2.19, p=.123, η^2=.089). However, we did observe a difference in baseline RTs (F(2,45)=4.48, p=.017, η^2=.166), with post hoc Tukey-Kramer tests confirming that the large target (Hit) group had faster RTs (325 $\pm$ 7 ms) than the small target (Miss) group (387 $\pm$ 22 ms). The medium target (Straddle) group's RTs (362 $\pm$ 12 ms) were not reliably different from either group. This baseline difference in RTs was only observed in this experiment (see below), and there was no correlation between baseline RT and late learning for the large target group (r = .09, p = .73), suggesting that RTs are not associated with the magnitude of adaptation.

Experiment 2

Baseline movement variability was marginally less for the group reaching to the larger target (Hit Target group: 3.09° $\pm$ .18°; Straddle Target group: 3.56° $\pm$ .16°; t(30)=-1.99 p=.056, d=0.70). Further supporting the argument that planning was no different across conditions, neither reaction times nor movement times differed between the groups (t(30)=-0.183, p=.856, d=.06 and t(30)=0.71, p=.484, d=.25, respectively).

Experiment 3

Baseline movement variability was not different across all three groups, including the control group trained with the bisected target (F(2,33)=1.38, p=.267, η^2=.077). Also, no differences across groups were observed for either RTs (F(2,33)=1.51, p=.236, η^2=.0084) or MTs (F(2,33)=.46, p=.634, η^2=.027).

Control group from Experiment 3

The behavior of the control group reaching to a large target bisected by a line (aligned with the target direction) during the acquisition phase is shown in magenta. The change in hand angle was not significantly different than that observed for the group that was tested with the large target in the acquisition phase of Experiment 3 (replotted here in green), suggesting that perceptual uncertainty did not make a substantive contribution to the effects of hitting the target. We omitted the transfer behavior of the large target group as this was when the large target was replaced with the small target.