Behavioral/Systems/Cognitive

Learning Optimal Adaptation Strategies in Unpredictable Motor Tasks

Daniel A. Braun,[1,2,3,4] **Ad Aertsen,**[1,3] **Daniel M. Wolpert,**[4] **and Carsten Mehring**[1,2]

[1]Bernstein Center for Computational Neuroscience, [2]Institute of Biology I, and [3]Institute of Biology III, Albert Ludwig University, 79104 Freiburg, Germany, and [4]Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom

Picking up an empty milk carton that we believe to be full is a familiar example of adaptive control, because the adaptation process of estimating the carton's weight must proceed simultaneously with the control process of moving the carton to a desired location. Here we show that the motor system initially generates highly variable behavior in such unpredictable tasks but eventually converges to stereotyped patterns of adaptive responses predicted by a simple optimality principle. These results suggest that adaptation can become specifically tuned to identify task-specific parameters in an optimal manner.

Introduction

Flexible motor control is an essential feature of biological organisms that pursue their goals in the face of uncertainty and incomplete knowledge about their environment. It is therefore not surprising that the phenomenon of adaptive behavior pervades the entire animal kingdom from simple habituation to complex reinforcement learning (Reznikova, 2007). Conceptually, learning is naturally understood as an optimization process that leads to efficient motor control. Thus, once learning has taken place and stable motor responses have formed, complex motor behaviors can often be understood by simple optimality principles that trade off attributes such as task success and energy expenditure (Todorov, 2004). In particular, optimal feedback control models have been successful in explaining a wide variety of motor behaviors on multiple levels of analysis (Todorov and Jordan, 2002; Scott, 2004; Diedrichsen, 2007; Guigon et al., 2007; Liu and Todorov, 2007). Optimal control models typically start out with the dynamics of the environment (e.g., dynamics of the arm or a tool) and a performance criterion in the form of a cost function (Stengel, 1994). The optimal control is then defined as a feedback rule that maps the past observations to a future action. This feedback rule minimizes the cost and is usually compared with the control actions chosen by a human or animal controller in an experiment (Loeb et al., 1990; Todorov and Jordan, 2002).

Importantly, optimal feedback control requires knowledge of the environmental dynamics in the form of an internal model. Consider, for example, that we wish to move a milk carton with known weight to a new location. An internal model would predict the future state of the controlled system $\mathbf{x}_{t+1}$ (e.g., future carton and hand position, velocity, etc.) from the current state $\mathbf{x}_t$ and the current action or control $\mathbf{u}_t$ (e.g., a neural control command to the muscles). Mathematically, the internal model can then be compactly represented as a mapping F with $\mathbf{x}_{t+1} = F(\mathbf{x}_t, \mathbf{u}_t)$. Experimentally, such internal models have been shown to play a crucial role in human motor control (Shadmehr and Mussa-Ivaldi, 1994; Wolpert et al., 1995; Wagner and Smith, 2008). However, the question arises whether adaptive behavior in an environment where the dynamics are not completely known can be understood by the same principles. Mathematically, we can formalize an adaptive control problem as a mapping $\mathbf{x}_{t+1} = F(\mathbf{x}_t, \mathbf{u}_t, \mathbf{a})$ with unknown system parameters $\mathbf{a}$ that have to be estimated simultaneously with the control process (Sastry and Bodson, 1989; Åström and Wittenmark, 1995). For example, in the case of a milk carton with an unknown weight, the motor system must adapt its estimate of the carton's weight (the parameter $\mathbf{a}$ in this case), while simultaneously exerting the necessary control to bring the carton to a desired location. This raises a fundamental question as to whether such estimation and control is a generic process operating whenever the motor system faces unpredictable situations or whether the adaptation process itself undergoes a learning phase so as to become tuned to specific environments and tasks in an optimal manner. Here we design a visuomotor learning experiment to test the hypothesis that with experience of an uncertain environment the motor system learns to perform a task-specific, stereotypical adaptation and control within individual movements in a task-optimal manner. In the following we will refer to changes in the control policy that occur within individual movements as "adaptation" to distinguish them from "learning" processes that improve these adaptive responses across trials.

Materials and Methods

Data acquisition. Nineteen healthy naive subjects participated in this study and gave informed consent after approval of the experimental procedures by the Ethics Committee of the Albert Ludwig University, Freiburg. Subjects controlled a cursor (radius 1 cm) on a 17″ TFT computer screen with their arm suspended by means of a long pendulum (4

Received July 2, 2008; revised Jan. 14, 2009; accepted March 16, 2009.

This study was supported in part by the German Federal Ministry of Education and Research (Grant 01GQ0420 to the Bernstein Center for Computational Neuroscience Freiburg), the Böhringer-Ingelheim Fonds, the European project SENSOPAC IST-2005-028056, and the Wellcome Trust. We thank Rolf Johansson for discussions and comments on earlier versions of this manuscript. We thank J. Barwind, U. Förster, and L. Pastewka for assistance with experiments and implementation.

Correspondence should be addressed to Daniel A. Braun at the above addresses. E-mail: dab54@cam.ac.uk.

DOI:10.1523/JNEUROSCI.3075-08.2009

 0270-6474/09/296472-07$15.00/0

m) that was attached to the ceiling. Subjects grabbed on to a handle at the bottom of the pendulum and moved it in the horizontal plane. Movements were recorded by an ultrasonic tracker system (CMS20, Zebris Medical, 300 Hz sampling, 0.085 mm accuracy). The screen displayed eight circular targets (radius 1.6 cm) arranged concentrically around a starting position (center–target distance 8 cm). Subjects were asked to move the cursor swiftly into the designated target and each trial lasted two seconds (therefore in early trials subjects often did not reach the target within the time window).

Experimental procedure. Two groups of subjects underwent two experimental blocks (2000 trials each) in which participants performed reaching movements in an uncertain environment. In both blocks the majority of trials were standard trials. However, on 20% of randomly selected trials a visuomotor perturbation was introduced. Each perturbation trial was always followed by at least one standard trial so that random perturbation trials were interspersed individually among the standard trials. In the first group (rotation group, 10 subjects) the perturbation was always a random visuomotor rotation with a rotation angle drawn from a uniform distribution over {±30°, ±50°, ±70°, ±90°}. Thus, the majority of trials had a normal hand–cursor relation and in visuomotor rotation trials the rotation angle could not be predicted before movement, requiring subjects to adapt online within a single trial to achieve the task. In the second group (target jump group, 9 subjects) the first block of 2000 trials were target jump transformations where the target jumped unpredictably to a rotated position (rotation angles drawn randomly again from a uniform distribution over {±30°, ±50°, ±70°, ±90°}). In target jump trials the jump occurred when then hand had moved 2 cm away from the origin. In the second block of 2000 trials the target jump group also experienced random rotations just like the first group. Thus, all subjects performed 4000 trials in total. We analyzed the first 2000 trials to assess how performance changed as subjects learned to adapt to the task requirements. Performance was assessed as the minimum distance to the target within the 2 s trial period, the magnitude of the second velocity peak, and movement variability. To calculate movement variability each two-dimensional positional trajectory was temporally aligned to the speed threshold of 10 cm/s and then the variance of the *x* and *y* positions were calculated for each time point across the trajectories and subjects (time 0 s corresponds to 200 ms before the speed threshold). The total variance was taken as the sum of the variance in *x* position and *y* position, and the square root of the variance (SD) was plotted. The last 2000 trials of the first group were used for fitting subjects' stationary patterns of adaptation to an optimal adaptive control model.

Adaptive optimal control model. To model adaptation and control we used a linear model of the hand/cursor system and a quadratic cost function to quantify performance (Körding and Wolpert, 2004). Full details of the simulations are provided in the supplemental Methods (available at www.jneurosci.org as supplemental material). As we include the effects of signal-dependent noise on the motor commands (Harris and Wolpert, 1998), the resulting optimal control model belongs to a class of modified linear quadratic-Gaussian systems (Todorov and Jordan, 2002). The equations we used are as follows:

$$\text{State update: } \mathbf{x}_{t+1} = F[\phi]\mathbf{x}_t + G\mathbf{u}_t + \text{signal-dependent noise}$$

$$\text{Observation: } \mathbf{y}_t = H\mathbf{x}_t + \text{additive noise.}$$

The state $\mathbf{x}_t$ represents the state of the hand/cursor system (a point-mass model) and the observation $\mathbf{y}_t$ represents the delayed sensory feedback to the controller. The state update equation depends on the current state (first term), the current motor command (second term), and signal-dependent noise (details in supplemental Methods, available at www.jneurosci.org as supplemental material). The observation equation relates the sensory feedback to the current state $\mathbf{x}_t$ and the additive observation noise. The important novelty here is that the forward model of the system dynamics F depends in a nonlinear way on the rotation parameter ϕ between the hand and cursor position. This parameter is unknown to subjects before each trial and must be estimated online during each movement.

The hand was modeled as a planar point-mass ($m = 1$ kg) with position and velocity vectors given by $\mathbf{p}_t^H$ and $\mathbf{v}_t$, respectively. The cursor position is given by a rotation of the hand position $\mathbf{p}_t^C = D_\phi \mathbf{p}_t^H$, where D_ϕ is the rotation matrix for a rotation of angle ϕ. The two-dimensional control signal $\mathbf{u}_t$ is transformed sequentially through two muscle-like low-pass filters both with time constants of 40 ms to produce a force vector $\mathbf{f}_t$ on the hand (with $\mathbf{g}_t$ representing the output of the first filter)—see (Todorov, 2005) and supplemental material (available at www.jneurosci.org) for details. Thus, the 10-dimensional state vector can be expressed as $[\mathbf{p}_t^C; \mathbf{v}_t; \mathbf{f}_t; \mathbf{g}_t; \mathbf{p}^{\text{Target}}]$, where $\mathbf{p}^{\text{Target}}$ corresponds to the target position in cursor space. Sensory feedback $\mathbf{y}_t$ is given as a noisy observation of the cursor position, hand velocity, and force vector with a feedback delay of 150 ms. In Results, we also compute the angular momentum as the cross product $\mathbf{p}_t^H \times \mathbf{v}_t$ multiplied by the point-mass $m = 1$ kg.

The cost function J can be expressed as follows:

$$\text{Cost } J = \frac{1}{2} E\left[\sum_{t=0}^{\infty} \{\mathbf{x}_t^T Q \mathbf{x}_t + \mathbf{u}_t^T R \mathbf{u}_t\}\right].$$

The matrix Q is designed to punish positional error between cursor and target and high velocities and is parameterized accordingly with two scalar parameters w_p and w_v. The matrix R punishes excessive control signals and was taken as the identity matrix scaled by a parameter r. Since the absolute value of the cost J does not matter for determining the optimal control, i.e., only the ratio between Q and R is important, we set $w_p = 1$. We chose a cost function without a fixed movement time (i.e., an infinite horizon cost function) so the amount of time required for adaptation to reach the target might vary. Such a cost function allows computing the state-dependent optimal policy at each point in time considering the most recent estimate of ϕ. Since the trial duration was relatively long (2 s) this cost function allowed reasonable fits to the data.

The optimal policy of the above control problem is the feedback rule that minimizes the cost function J. Since the parameter ϕ is unknown, this adaptive optimal control problem can only be solved approximately by decomposing it into an estimation problem and a control problem (certainty-equivalence principle). The estimation problem consists of simultaneously estimating the unobserved state $\mathbf{x}_t$ and the unknown parameter ϕ from the observations $\mathbf{y}_t$. This can be achieved by introducing an augmented state $\tilde{\mathbf{x}}_t = [\mathbf{x}_t; \phi_t]$ and using a nonlinear filtering method (e.g., unscented Kalman filter) for the estimation $\hat{\tilde{\mathbf{x}}}_t = [\hat{\mathbf{x}}_t; \hat{\phi}_t]$ in this augmented state space—see supplemental material (available at www.jneurosci.org) for details. To allow the controller to adapt its estimate of ϕ we model the parameter as a random walk with covariance Ω_v, which determines the rate of adaptation within a trial. The optimal control command at every time point can then be computed as a feedback control law $\mathbf{u}_t = -L[\hat{\phi}_t]\hat{\mathbf{x}}_t$, where $L[\hat{\phi}_t]$ is the optimal feedback gain for a given parameter estimate $\hat{\phi}_t$. To allow for the uncertainty of the parameter estimate to affect the control process (noncertainty-equivalence effects), we introduce two additional cautiousness parameters λ_p and λ_v. Based on the models uncertainty in the rotation parameter ϕ, these reduce the gains of the position and velocity components of the feedback thereby slowing down the controller in the face of high uncertainty (equivalent to making the energy component of the cost more important). Importantly, the cautiousness parameters do not introduce a new optimality criterion; rather they provide a heuristic to find an approximation to the optimal solution and are often used in adaptive control theory when faced with an analytically intractable optimal control problem (see supplemental material, available at www.jneurosci.org). Accordingly, the costs achieved by a cautious adaptive controller can be lower than by a noncautious adaptive controller—see supplemental material (available at www.jneurosci.org) for details.

Parameter fit. Some of the parameters of the model were taken from the literature as indicated above. There were six free scalar parameters that were fit to the data, and these are (1) the cost parameters w_v and r, (2) the cautiousness parameters λ_p and λ_v, (3) the adaptation rate Ω_v, and (4) the signal-dependent noise level. We adjusted these parameters to fit the mean trajectory of the 90°-rotation trials (by collapsing the +90° and −90° trials into one angle). These parameter settings were then used to extrapolate behavior to both the standard trials and all other rotation

trials. The reason we chose 90° is that the perturbation has the strongest effect here, and therefore the fit would have the best signal-to-noise ratio to allow us to get the most precise estimates of the parameters. Thus, the issue of overfitting is avoided as the model predictions are evaluated for nonfitted conditions. The fit was to the second 2000 trials when subjects of the rotation group exhibited stationary responses to the visuomotor rotations. Details of the parameter fits can be found in the supplemental material (available at www.jneurosci.org).

Results

To test the hypothesis that the motor system can learn to adapt optimally to specific classes of environments we exposed a first group of participants to a reaching task in which on 20% of the trials a random visuomotor rotation was introduced. Since these random rotations could not be predicted (and were zero mean across all rotations), participants had to adapt to the perturbations online during the movement. This online adaptation is different from online error correction (Diedrichsen et al., 2005), since the rules of the control process—i.e., the "control policy" that maps sensory inputs to motor outputs—has to be modified. Importantly, the modification of the control law is a learning process, whereas online error correction, e.g., to compensate for a target jump, can take place under the same policy without learning a new controller. To enforce online adaptation the vast majority of trials had a standard hand/cursor relationship and only occasional trials were perturbed. Thus, movements typically started out in a straight line to the cursor target because subjects assumed by default a standard mapping between hand and cursor — see Figure 1*A*. However, after a time delay of 100–200 ms into the movement subjects noticed the mismatch between hand and cursor position in random rotation trials and started to modify their movements. This adaptive part of the movement can be seen from the change of direction in the trajectory and the appearance of a second peak in the speed profile (Fig. 1*C*).

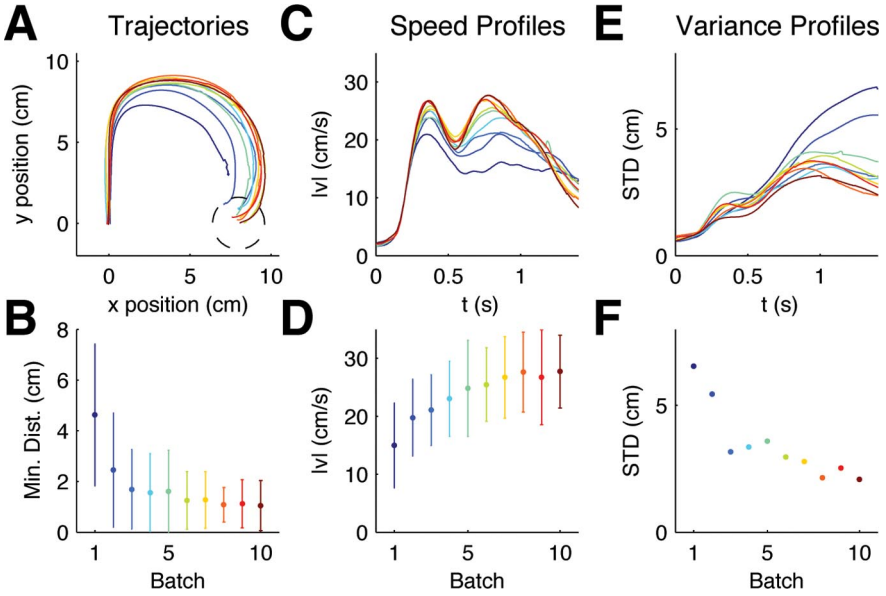

Figure 1. Evolution of within-trial adaptive behavior for random rotation trials. ***A***, Mean hand trajectories for ±90° rotation trials in the first 10 batches averaged over trials and subjects (each batch consisted of 200 trials, ~5% of which were ±90° rotation trials). The −90° rotation trials have been mirrored about the *y*-axis to allow averaging. Dark blue colors indicate early batches, green colors intermediate batches, red colors indicate later batches. ***B***, The minimum distance to the target averaged for the same trials as A (error bars indicate SD over all trajectories and all subjects). This shows that subjects' performance improves over batches. ***C***, Mean speed profiles for ±90° rotations of the same batches. In early batches, movements are comparatively slow and online adaptation is reflected in a second peak of the speed profile which is initially noisy and unstructured. ***D***, The magnitude of the second peak increases over batches (same format as ***B***). ***E***, SD profiles for ±90° rotation trajectories computed for each trial batch. ***F***, SD of the last 500 ms of movement. Over consecutive batches the variability is reduced in the second part of the movement.

To assess our hypothesis of task-optimal adaptation, we first investigated whether subjects showed any kind of improvement in adapting to the unpredictable perturbations during the movements. Indeed, we found that the adaptation patterns in random rotation trials were very different in early trials compared with the same rotations performed later in the experiment (Fig. 1*B,D,F*). In the beginning, large movement errors occurred more frequently, i.e., subjects often did not manage to reach the target precisely within the prescribed 2 s time window (Fig. 1*B*). The difference in the minimum distance to the target within this allowable time window between the first and last batch of 200 trials was significant ($p < 0.01$, Wilcoxon rank-sum test). In early trials the second peak of the speed profile was barely visible as movements were relatively unstructured and cautious, but in later trials a clear second speed peak emerged (Fig. 1*C*). Early trials also showed high variability in the second part of the movement, whereas in later trials adaptive movements were less variable and therefore more reproducible between subjects (Fig. 1*F*)—the variability in the last 500 ms of the movement in the first batch was significantly larger than in the last batch ($p < 0.01$, *F* test). The color code in Figure 1 indicates that the second part of the movement converged to a stereotyped adaptive response. To test for the possibility that subjects simply became nonspecifically better at feedback control, a second group of participants performed a target jump task for the first 2000 trials. In direct correspondence to the random rotation task 20% of the trials were random target jump trials. Since a target jump does not require learning a new policy but simply an update of the target position in the current control law, we would expect to see no major learning processes in this task. This is indeed what we found. In Figure 2 we show the same features that we evaluated in the random rotation trials to assess over-trial evolution of sensorimotor response patterns.

To test whether the change in behavior over trials might represent an improvement—in the sense of minimizing a cost function—we computed the costs of the experimentally observed trajectories for 90° rotations. We used the inverse system equations to reverse-engineer the state space vector $\mathbf{x}_t$ and the control command $\mathbf{u}_t$ from the experimental trajectories. We then used a quadratic cost function that successfully captured standard movements and computed the costs of all the trajectories of the experiment. We found that the cost of the trajectories with regard to the quadratic cost function decreased over trials (Fig. 3*A*). This shows that the observed change in adaptation can be understood as a cost-optimization process. In contrast to the first group, the second group showed no trend that would indicate learning—there is no significant difference between the minimum distance to the target between the first and the last batch ($p > 0.01$, Wilcoxon

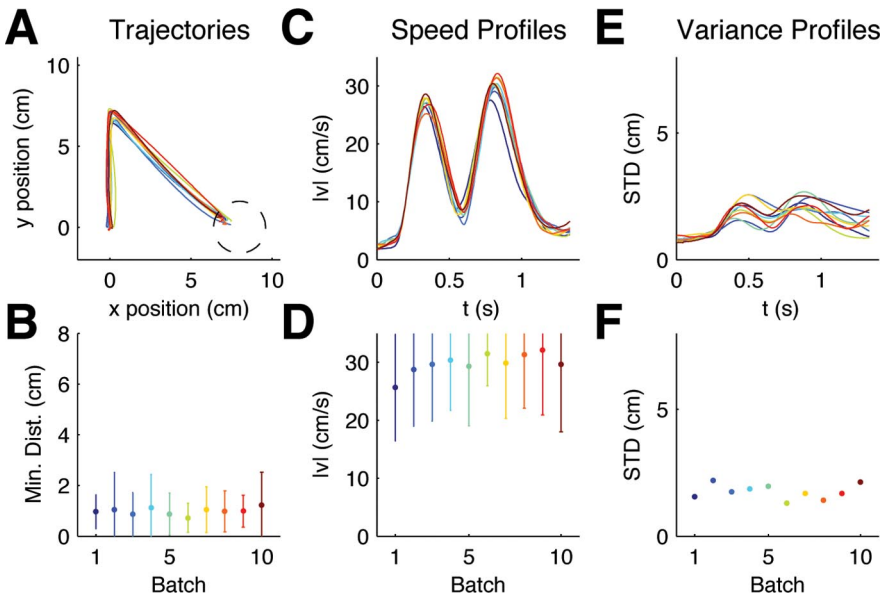

Figure 2. Evolution of motor responses to random target jumps. ***A***, Mean trajectories for ±90° target jumps over batches of 200 trials, 5% of which were ±90° target jump trials. Dark blue colors indicate early batches, red colors indicate later batches. ***B***, The bottom shows that subjects' performance did not significantly improve over trials. Error bars indicate SD over all trials and subjects. ***C***, Mean speed profiles for ±90° target jumps of the same trial batches. A second velocity peak is present right from the start. ***D***, The bottom shows the evolution of the magnitude of the second speed peak. ***E***, SD for ±90° target jumps computed over the same trial batches. Over consecutive batches the variance remains constant. ***F***, SD over the last 500 ms of movement.

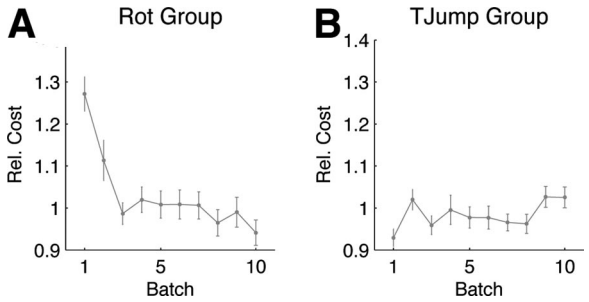

Figure 3. ***A***, Rotation group. Relative cost of subjects' movements in response to ±90° visuomotor rotations. Over trial batches (200 trials) the cost of the adaptive strategy decreases. ***B***, Target jump group. Relative cost of subjects' movements in response to ±90° target jumps. There is no improvement over trials. In both cases the costs have been computed by calculating the control command and the state space vectors from the experimental trajectories by assuming a quadratic cost function. The cost has been normalized to the average cost of the last five trial batches.

rank-sum test). The reverse-engineered cost function for the 90° target jumps was flat over trial batches (Fig. 3*B*).

After the first block of target jump trials, the second group experienced a second block of random rotation trials identical to the second block the first group experienced. If the first group learned a feedback control policy specifically for rotations in the first block of trials then both groups should perform very differently in the second block of trials where both groups experienced random rotation trials. Again this hypothesis was confirmed by our results. The first group that was acquainted with rotations showed a stationary response to unexpected rotations (Fig. 4*A–C*). Performance error, speed profiles, and SD showed no changes over trials (Fig. 5*A–C*). Thus, there was no significant difference between the minimum distance to the target between the first and the last trial batches ($p > 0.01$, Wilcoxon rank-sum test). In contrast the second group initially performed not better than naive subjects; i.e., their performance was the same as the performance of the rotation group in the beginning of the first block (Fig. 4*D–E*). Then, over the first few trial batches this group substantially improved (Fig. 5*D–E*) and the difference in minimum target distance between the first batch and the last are highly significant ($p < 0.01$, Wilcoxon rank-sum test). Therefore, the experience of unpredictable target jumps did not allow for learning an adaptive control policy that is optimized for unpredictable visuomotor rotations.

Finally, we investigated whether the stationary adaptation patterns observed in later trials of the first group could be explained by an adaptive optimal feedback controller that takes the task-specific parameters of a visuomotor rotation explicitly into account. Importantly, a nonadaptive controller that ignores the rotation becomes quickly unstable (Fig. S4). The adaptive optimal controller has to estimate simultaneously the arm and cursor states as well as the hidden "visuomotor rotation"-parameter online (see Materials and Methods). This results in the online estimation of the forward model for the visuomotor transformation. The estimated forward model, in turn, together with the estimated cursor and hand state can be used to compute the optimal control command at every point in time. At the beginning of each trial the forward model estimate of the adaptive controller is initialized to match a standard hand–cursor mapping without a visuomotor rotation (representing the prior, the average of all rotations). Due to feedback delays, any mismatch between actual and expected cursor position can only be detected by the adaptive controller some time into the movement. The observed mismatch can then be used both for the adaptation of the state and parameter estimates and for improved control (supplemental Fig. S3, available at www.jneurosci.org as supplemental material). To test this model quantitatively, we adjusted the parameters of the model to fit the mean trajectory and variance of the 90°-rotation trials and used this parameter set to predict behavior on both the standard and other rotation trials. In the absence of the "cautiousness" parameters which slow down control in the presence of uncertainty about the rotation parameter, the predictions gave hand speeds that were higher than those in our experimental data (supplemental Fig. S5, available at www.jneurosci.org as supplemental material). In the presence of the "cautiousness" parameters not only was the cost of the controller lower, but we also found that the adaptive optimal control model predicted the main characteristics of the paths, speed and angular momentum, as well as the trial-to-trial variability of movements, with high reliability (Fig. 6)—the predictions yielded $r^2 > 0.83$ for all kinematic variables. Both model and experimental trajectories first move straight toward the target and then show adaptive movement corrections after the feedback delay time elapsed. Both model and experiment show a characteristic second peak in the velocity profile, and the model predicts this peak correctly for all rotation angles. Also the trial-by-trial variability is correctly predicted for the different rotations.

Discussion

Our results provide evidence that the motor system converges to task-specific stereotypical adaptive responses in unpredictable motor tasks that require simultaneous adaptation and control. Moreover, we show that such adaptive responses can be explained by adaptive optimal feedback control strategies. Thus, our results provide evidence that the motor system is not only capable of learning nonadaptive optimal control policies (Todorov and Jordan, 2002; Diedrichsen, 2007) but also of learning optimal simultaneous adaptation and control. This shows that the learning process of finding an optimal adaptive strategy can be understood as an optimization process with regard to similar cost criteria as proposed in nonadaptive control tasks (Körding and Wolpert, 2004).

Previous studies have shown that optimal feedback control successfully predicts behavior of subjects that have uncertainty about their environment (e.g., a force-field) that changes randomly from trial to trial (Izawa et al., 2008). However, in these experiments subjects did not have the opportunity to adapt efficiently to the perturbation within single trials. Rather the perturbation was modeled as noise or uncertainty with regard to the internal model. In our experiments subjects also have uncertainty over the internal model, but they have enough time to resolve this uncertainty within the trial and adapt their control policy accordingly. Another recent study (Chen-Harris et al., 2008) has shown that optimal feedback control can be successfully combined with models of motor learning (Donchin et al., 2003; Smith et al., 2006) to understand learning of internal models over the course of many trials. Here we show that learning and control can be understood by optimal control principles within individual trials.

Optimal within-trial adaptation of the control policy during a movement presupposes knowledge of a rotation-specific internal model $\mathbf{x}_{t+1} = F(\mathbf{x}_t, \mathbf{u}_t, a)$, where a denotes the system parameters the motor system is uncertain about (i.e., a rotation-specific parameter). This raises the question of how the nervous system could learn that a is the relevant parameter and that F depends on a in a specific way. In adaptive control theory this is known as the structural learning problem (Sastry and Bodson, 1989; Åström and Wittenmark, 1995) as opposed to the parametric learning problem of estimating a given knowledge of $F(\star, a)$. In our experiments, subjects in the rotation group have a chance to learn the structure of the adaptive control problem (i.e., visuomotor rotations with a varying rotation angle) in the first 2000 trials of the experiment in which they experience random rotations. As previously shown (Braun et al., 2009), such random exposure is apt to induce structural learning and can lead to differential adaptive behavior. Here we explicitly investigate the evolution of structural learning for the online ad-

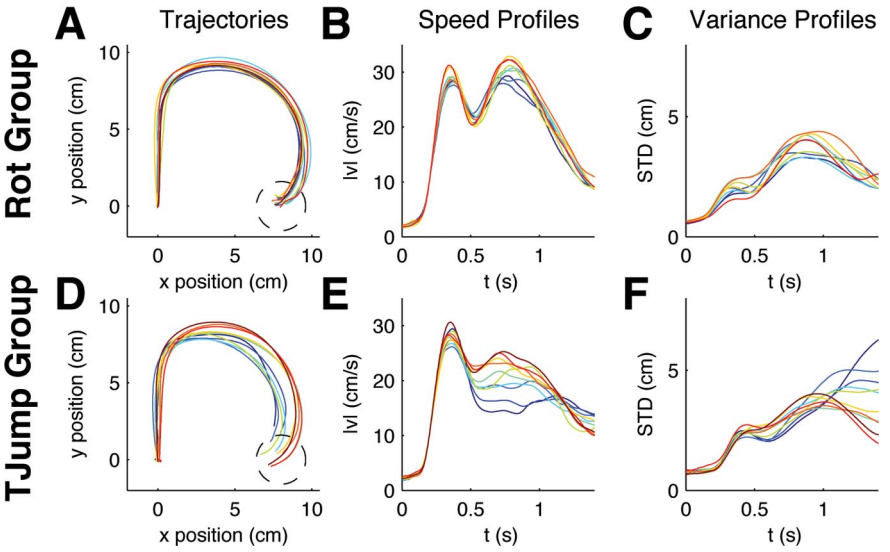

Figure 4. Evolution of within-trial adaptation and control for ±90° random rotations in the second block of 2000 trials. ***A***, Movement trajectories averaged over batches of 200 trials for the group that had experienced unexpected rotation trials already in the previous 2000 trials. Dark blue colors indicate early batches, red colors indicate later batches. This group shows no improvement. ***B***, Speed profiles of the same trial batches. ***C***, SD in the same trials. There is no trend over consecutive batches. ***D***, Average movement trajectories averaged over batches of 200 trials for the group that had experienced unexpected target jump trials in the previous 2000 trials. This group shows learning. ***E***, Speed profiles of the target jump group. ***F***, SD in the same trials. The movement characteristics change over consecutive batches.

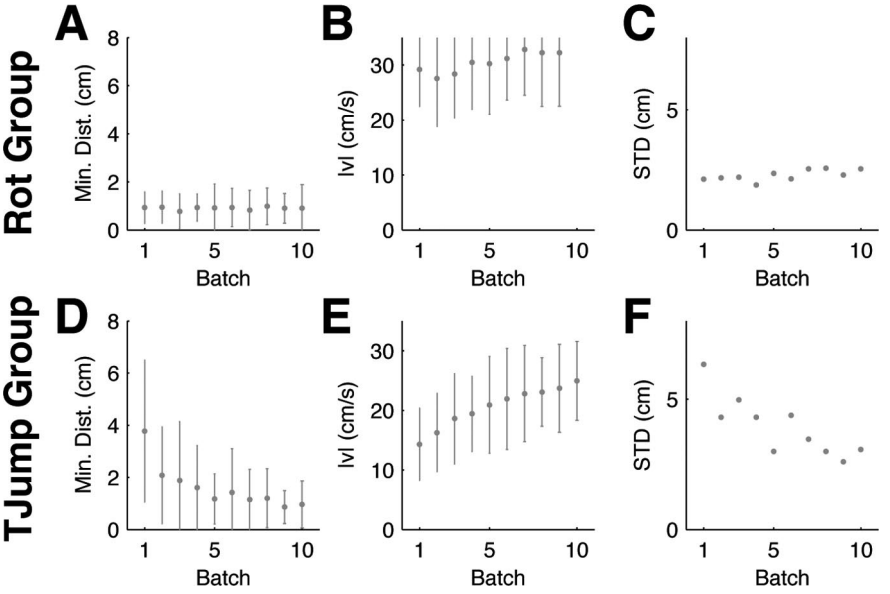

Figure 5. Evolution of within-trial adaptive control for random rotations in the second block of 2000 trials. ***A***, Minimum distance to target in ±90° rotation trials averaged over batches of 200 trials for the group that had experienced unexpected rotation trials already in the previous 2000 trials. This group shows no improvement. Error bars show SD over all trials and subjects. ***B***, Mean magnitude of the second velocity peak over batches of 200 trials for the rotation group. ***C***, SD in the last 500 ms of movement for ±90° rotations computed over the same trial batches for the rotation group. There is no trend over consecutive batches. ***D***, Minimum distance to target in ±90° rotation trials averaged over batches of 200 trials for the group that had experienced unexpected target jump trials in the previous 2000 trials. This group shows a clear improvement. ***E***, Mean magnitude of the second velocity peak over batches of 200 trials for the target jump group. ***F***, SD in the last 500 ms of movement for ±90° rotations computed over the same trial batches for the target jump group. The SD clearly decreases over consecutive batches.

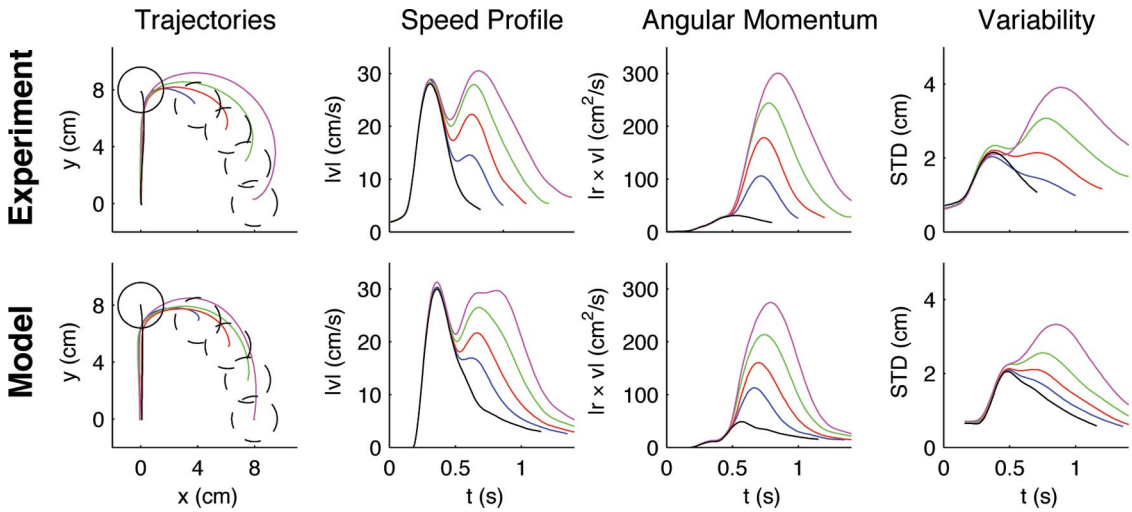

Figure 6. Predictions of the adaptive optimal control model compared with movement data. Averaged experimental hand trajectories (left column), speed profiles (second column), angular momentum (third column), and trajectory variability (right column) for standard trials (black) and rotation trials [±30° (blue), ±50° (red), ±70° (green), ±90° (magenta)]. The second peak in the speed profile and the magnitude of the angular momentum (assuming $m = 1$ kg) reflect the corrective movement of the subjects. Higher rotation angles are associated with higher variability in the movement trajectories in the second part of the movement. The variability was computed over trials and subjects. The trajectories for all eight targets have been rotated to the same standard target and averaged, since model predictions were isotropic. The model consistently reproduces the characteristic features of the experimental curves.

aptation to visuomotor rotations (Fig. 1) and, based on an optimal adaptive feedback control scheme, show that this learning can be indeed understood as an improvement (Fig. 3) leading to optimal adaptive control strategies. It should be noted, however, that learning the rotation structure does not necessarily imply that the brain is learning to adapt literally a single neural parameter, but that exploration for online adaptation should be constrained by structural knowledge leading to more stereotype adaptive behavior. In the latter 2000 trials, when subjects know how to adapt efficiently to rotations, their behavior can be described by a parametric adaptive optimal feedback controller that exploits knowledge of the specific rotation structure.

In the literature there has been an ongoing debate whether corrective movements and multiple velocity peaks indicate discretely initiated submovements (Lee et al., 1997; Fishbach et al., 2007) or whether multimodal velocity profiles are the natural outcome of a continuous control process interacting with the environment (Kawato, 1992; Bhushan and Shadmehr, 1999). Our model predictions are consistent with the second view. Although corrective movements in our experiments are certainly induced by unexpected perturbations, the appearance of corrections and multimodal velocity profiles can be explained by a continuous process of adaptive optimal control.

As already described, online adaptation should not to be confused with online error correction (Diedrichsen et al., 2005). Online correction is, for example, required in the case of an unpredicted target jump. Under this condition the same controller can be used, i.e., the mapping from sensory input to motor output is unaltered. However, unexpectedly changing the hand–cursor relation (e.g., by a visuomotor rotation) requires the computation of adaptive control policies. This becomes intuitively apparent in the degenerate case of 180° rotations, as any correction of a naive controller leads to the opposite of its intended effect. However, it should be noted that the distinction between adaptation and error correction can be blurry in many cases. Strictly speaking, an adaptive control problem is a nonlinear control problem with a hyper-state containing state variables and (unknown) parameters. This means in principle no extra theory of adaptive control is required. In practice, however, there is a well established theory of adaptive control (Sastry and Bodson, 1989; Åström and Wittenmark, 1995) that is built on the (somewhat artificial) distinction between state variables and (unknown) parameters. The two quantities are typically distinct in their properties. In general, the state, for example the position and velocity of the hand, changes rapidly and continuously within a movement. In contrast, other key quantities change discretely, like the identity of a manipulated object, or on a slower timescale, like the mass of the limb. We refer to such discrete or slowly changing quantities as the "parameters" of the movement. Therefore, state variables change on a much faster timescale than system parameters and the latter need to be estimated to allow for control of the state variables. This is exactly the case in our experiments where the parameters (rotation angle) change slowly and discretely from trial to trial, but the state variables (hand position, velocity, etc.) change continuously over time (within a trial). Thus, estimating uncertain parameters can subserve continuous control in an adaptive manner. In summary, our results suggest that the motor system can learn optimal adaptive control strategies to cope with specific uncertain environments.

References

Åström KJ, Wittenmark B (1995) Adaptive control, Ed 2. Reading, MA: Addison-Wesley.

Bhushan N, Shadmehr R (1999) Computational nature of human adaptive control during learning of reaching movements in force fields. Biol Cybern 81:39–60.

Braun DA, Aertsen A, Wolpert DM, Mehring C (2009) Motor task variation induces structural learning. Curr Biol 19:352–357.

Chen-Harris H, Joiner WM, Ethier V, Zee DS, Shadmehr R (2008) Adaptive control of saccades via internal feedback. J Neurosci 28:2804–2813.

Diedrichsen J (2007) Optimal task-dependent changes of bimanual feedback control and adaptation. Curr Biol 17:1675–1679.

Diedrichsen J, Hashambhoy Y, Rane T, Shadmehr R (2005) Neural correlates of reach errors. J Neurosci 25:9919–9931.

Donchin O, Francis JT, Shadmehr R (2003) Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. J Neurosci 23:9032–9045.

Fishbach A, Roy SA, Bastianen C, Miller LE, Houk JC (2007) Deciding when and how to correct a movement: discrete submovements as a decision making process. Exp Brain Res 177:45–63.

Guigon E, Baraduc P, Desmurget M (2007) Computational motor control: redundancy and invariance. J Neurophysiol 97:331–347.

Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. Nature 394:780–784.

Izawa J, Rane T, Donchin O, Shadmehr R (2008) Motor adaptation as a process of reoptimization. J Neurosci 28:2883–2891.

Kawato M (1992) Optimization and learning in neural networks for formation and control of coordinated movement. In: Attention and performance (Meyer D, Kornblum S, eds), pp 821–849. Cambridge, MA: MIT.

Körding KP, Wolpert DM (2004) The loss function of sensorimotor learning. Proc Natl Acad Sci U S A 101:9839–9842.

Lee D, Port NL, Georgopoulos AP (1997) Manual interception of moving targets. II. On-line control of overlapping submovements. Exp Brain Res 116:421–433.

Liu D, Todorov E (2007) Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. J Neurosci 27:9354–9368.

Loeb GE, Levine WS, He J (1990) Understanding sensorimotor feedback through optimal control. Cold Spring Harb Symp Quant Biol 55:791–803.

Reznikova ZI (2007) Animal intelligence: from individual to social cognition. Cambridge, MA: Cambridge UP.

Sastry S, Bodson M (1989) Adaptive control: stability, convergence, and robustness. Englewood Cliffs, NJ: Prentice-Hall Advanced Reference Series.

Scott SH (2004) Optimal feedback control and the neural basis of volitional motor control. Nat Rev Neurosci 5:532–546.

Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive representation of dynamics during learning of a motor task. J Neurosci 14:3208–3224.

Smith MA, Ghazizadeh A, Shadmehr R (2006) Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol 4:e179.

Stengel RF (1994) Optimal control and estimation, revised edition. New York: Dover.

Todorov E (2004) Optimality principles in sensorimotor control. Nat Neurosci 7:907–915.

Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. Neural Comput 17:1084–1108.

Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. Nat Neurosci 5:1226–1235.

Wagner MJ, Smith MA (2008) Shared internal models for feedforward and feedback control. J Neurosci 28:10663–10673.

Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. Science 269:1880–1882.

Supplementary Material

Learning optimal adaptation strategies in unpredictable motor tasks

D.A. Braun, A. Aertsen, D.M. Wolpert, C. Mehring

Contents

1 Adaptive Optimal Control Methods

The general mathematical model underlying the fits and predictions in the main text belongs to a class of modified Linear-Quadratic-Gaussian (LQG) models [Stengel, 1994]. LQG models deal with linear dynamic systems, quadratic cost functions as performance criteria, and Gaussian random variables as noise. Here we consider the following model class:

$$\begin{aligned} \vec{x}_{t+1} &= F[\vec{a}\,]\,\vec{x}_t + G\,\vec{u}_t + \vec{\xi}_t + G\sum_i C_i\,\vec{u}_t\,\sigma_{i,t} \\ \vec{y}_t &= H\,\vec{x}_t + \vec{\chi}_t \\ J &= \frac{1}{2}\mathbb{E}\left[\sum_{t=0}^{\infty}\left\{\vec{x}_t^T\,Q\,\vec{x}_t + \vec{u}_t^T\,R\,\vec{u}_t\right\}\right] \end{aligned} \tag{1}$$

with the following variables

dynamic state	$\vec{x}_t \in \Re^n$
unknown system parameters	$\vec{a} \in \Re^l$
control signal	$\vec{u}_t \in \Re^m$
feedback observation	$\vec{y}_t \in \Re^q$
expected cumulative cost	$J \in \Re$
state cost matrix	$Q = Q^T \geq 0$
control cost matrix	$R = R^T > 0$

Time is discretized in bins of $10ms$. The noise variables $\vec{\xi}_t \in \Re^n$, $\vec{\chi}_t \in \Re^k$, $\sigma_{i,t} \in \Re$ are realizations of independent, zero-mean, Gaussian noise processes with covariance matrices $\mathbb{E}[\vec{\xi}_{t_1}\vec{\xi}_{t_2}^T] = \Omega_\xi\,\delta_{t_1 t_2}$, $\mathbb{E}[\vec{\chi}_{t_1}\vec{\chi}_{t_2}^T] = \Omega_\chi\,\delta_{t_1 t_2}$ and $\mathbb{E}[\sigma_{i_1,t_1}\sigma_{i_2,t_2}] = \delta_{t_1 t_2}\,\delta_{i_1 i_2}$ respectively. The dynamic state $\vec{x}_t$ is a hidden variable that needs to be inferred from feedback observations $\vec{y}_t$. An initial estimate of $\vec{x}_0$ is given by a normal distribution with mean $\hat{\vec{x}}_0$ and covariance P_0^x. Accordingly, an initial estimate of the unknown parameters $\vec{a}$ is given by a normal distribution with mean $\hat{\vec{a}}_0$ and covariance P_0^a. This allows to state the optimal control problem: given $F[\hat{\vec{a}}_0]$, G, H, C_i, Ω_ξ, Ω_χ, P_0^x, P_0^a, R, Q, what is the control law $\vec{u}_t = \vec{\pi}(\hat{\vec{x}}_0, \vec{u}_0, ...\vec{u}_{t-1}, \vec{y}_0, ...\vec{y}_{t-1}, t)$ that minimizes the expected cumulative cost J?

In the absence of multiplicative noise (i.e. $C_i \equiv 0\ \forall i$) and assuming perfect knowledge of all system parameters $\vec{a}$, the posed optimal control problem has a well-known solution [Stengel, 1994]: a Kalman filter estimates the hidden state $\vec{x}_t$ optimally in a least-squares sense and a linear optimal controller maps this estimate $\hat{\vec{x}}_t$ into a control signal $\vec{u}_t$. Several approximative solutions have been suggested in the literature to solve the non-adaptive control problem with multiplicative noise [Moore et al., 1999; Todorov, 2005]. Here we address the optimal control problem with multiplicative noise in the presence of parameter uncertainties.

Unfortunately, adaptive optimal control problems can, in general, neither be solved analytically nor numerically. Therefore, reasonable approximations have to be found that are applicable to broad

classes of problems. In movement neuroscience, usually 'indirect' adaptive control schemes are used, implying that subjects avail themselves of internal models both to predict their environment and to adjust their motor control on the basis of these predictions. Mathematically, this entails the separation of estimation and control processes, i.e. the general proceeding is (1) to identify the system parameters $\vec{a}$ on-line, and (2) to exploit the resulting estimate $\hat{\vec{a}}_t$ by appropriately adjusting the control law $\vec{\pi}$ when computing $\vec{u}_t$.

1.1 The Estimation Problem

To perform system identification on-line in a noisy environment implies solving a *joint filtering problem* [Haykin, 2001], because states and parameters have to be estimated simultaneously. Joint filtering methods are based on the definition of an *augmented* or *joint* state space with the concatenated state vector

$$\vec{\mathfrak{x}}_t = \begin{bmatrix} \vec{x}_t \\ \vec{a}_t \end{bmatrix} \tag{2}$$

Since the unknown parameters are assumed to be constant ($\vec{a}_{t+1} = \vec{a}_t$), system identification can be simply instantiated by letting the parameters do a random walk driven by a process noise[1] $\vec{\nu}_t \sim \mathcal{N}(0, \Omega_\nu)$

$$\vec{a}_{t+1} = \vec{a}_t + \vec{\nu}_t \tag{3}$$

To be compatible with the concatenated state vector, the state transition matrix, the measurement matrix and the process covariance matrix need to be modified for the joint state space

$$\begin{aligned} \tilde{F}[\vec{a}_t] &= \begin{bmatrix} F[\vec{a}_t] & 0 \\ 0 & \mathbb{I}_{l\times l} \cdot \vec{a}_t \end{bmatrix} \\ \tilde{H} &= \begin{bmatrix} H & 0 \end{bmatrix} \\ \tilde{\Omega}_{\tilde{\xi}} &= \begin{bmatrix} \Omega_\xi & 0 \\ 0 & \Omega_\nu \end{bmatrix} \end{aligned}$$

Since adaptive control problems are inherently nonlinear, the standard Kalman filter solution [Kalman, 1960] is not applicable in the augmented state space. A state-of-the-art method for nonlinear estimation problems is *Unscented Kalman filtering* [Haykin, 2001], where the distribution of the random variable is sampled efficiently by carefully chosen *sigma points* that are propagated through the full nonlinearity. The *sigma vectors* of the random variable with mean $\hat{\vec{\mathfrak{x}}} \in \Re^{n+l}$ and covariance $P^{\mathfrak{x}}$ are calculated according to

$$\begin{aligned} \mathcal{X}_0 &= \hat{\vec{\mathfrak{x}}} \\ \mathcal{X}_i &= \hat{\vec{\mathfrak{x}}} + \gamma\,(\sqrt{P^{\mathfrak{x}}})_i \qquad && i = 1, \ldots, n+l \\ \mathcal{X}_i &= \hat{\vec{\mathfrak{x}}} - \gamma\,(\sqrt{P^{\mathfrak{x}}})_{i-n-l} \qquad && i = n+l+1, \ldots, 2(n+l) \end{aligned}$$

[1] The parameter covariance matrix Ω_ν determines the time scale of parameter adaptation. In the present case, Ω_ν is a phenomenological constant that captures the adaptation rate of the brain for a specific learning task. The optimal control problem is posed under the constraint of this given adaptation rate.

with the scaling parameter γ [Julier et al., 1995]. The expression $(\sqrt{P^{\mathfrak{x}}})_i$ denotes the ith column of the matrix square root of $P^{\mathfrak{x}}$ that can be determined, for instance, by the lower-triangular Cholesky factorization. This leads to the following Kalman filter equations:

$$\vec{\hat{\mathfrak{x}}}_t = \vec{\hat{\mathfrak{x}}}_t^{\,-} + K_t\,[\vec{y}_t - \vec{\hat{y}}_t^{\,-}] \tag{4}$$

$$P_t^{\mathfrak{x}} = P_t^{\mathfrak{x}-} - K_t\,P_t^{yy}\,K_t^T \tag{5}$$

with the Kalman gain $K_t = P_t^{xy}\,(P_t^{yy})^{-1}$ and the covariances

$$P_t^{yy} = \sum_{i=0}^{2n} W_i^{(c)} \left(\mathcal{Y}_t^- - \vec{\hat{y}}_t^{\,-}\right)\left(\mathcal{Y}_t^- - \vec{\hat{y}}_t^{\,-}\right)^T + \Omega_\chi \tag{6}$$

$$P_t^{xy} = \sum_{i=0}^{2n} W_i^{(c)} \left(\mathcal{X}_t^- - \vec{\hat{\mathfrak{x}}}_t^{\,-}\right)\left(\mathcal{Y}_t^- - \vec{\hat{y}}_t^{\,-}\right)^T \tag{7}$$

$$P_t^{\mathfrak{x}-} = \sum_{i=0}^{2n} W_i^{(c)} \left(\mathcal{X}_t^- - \vec{\hat{\mathfrak{x}}}_t^{\,-}\right)\left(\mathcal{X}_t^- - \vec{\hat{\mathfrak{x}}}_t^{\,-}\right)^T + \tilde{\Omega}_{\tilde{\xi}} + \sum_i \tilde{G}\,C_i, \vec{u}_t\,\vec{u}_t^T\,C_i^T\,\tilde{G}^T \tag{8}$$

The last summand of equation (8) accounts for higher variability due to multiplicative noise and is derived from a linear approximation scheme following [Moore et al., 1999]. The required sigma points are calculated as

$$\mathcal{X}_t^- = \tilde{F}[\mathcal{X}_{t-1}] + \tilde{G}\,\vec{u}_{t-1} \tag{9}$$

$$\vec{\hat{\mathfrak{x}}}_t^{\,-} = \sum_{i=1}^{2n} W_i^{(m)}\,\mathcal{X}_t^- \tag{10}$$

$$\mathcal{Y}_t^- = \tilde{H}\mathcal{X}_t^- \tag{11}$$

$$\vec{\hat{y}}_t^{\,-} = \sum_{i=1}^{2n} W_i^{(m)}\,\mathcal{Y}_t^- \tag{12}$$

with scaling parameters $W_i^{(m)}$ and $W_i^{(c)}$ [Julier et al., 1995].

1.2 The Control Problem

In general, the posed adaptive optimal control problem will be a *dual control problem*[2] without a straightforward solution [Åström and Wittenmark, 1989]. In the case of partial system observability, it is a common approximation [Bar-Shalom and Tse, 1974; Bar-Shalom, 1981] to decompose the cost function J into a deterministic part J_D (*certainty-equivalent control*), a cautious part J_C (*prudent control*), and a probing part J_P (*explorative control*). Accordingly, the adaptive optimal controller

[2]The dual control problem is conceptually related to the *exploration-exploitation-dilemma* known in reinforcement learning [Sutton and Barto, 1998], since it deals with a similar set of questions: If the parameter uncertainty is not too high, should one act as if there were no uncertainty (*certainty-equivalent control*)? Should one be particularly prudent in an unknown environment (*cautious control*)? Or is it best to be explorative, i.e. invest short-term effort into identifying the unknown parameters and exploit this knowledge subsequently (*probing control*)?

u is designed as a composition of the sub-controllers u^D, u^C and u^P. Then it depends on the characteristics of the specific control problem which sub-controllers dominate and which might be neglected. Especially the design of the sub-controllers u^C and u^P usually follows mere heuristic principles. The design of u^D can be obtained by virtue of the *certainty-equivalence principle*[3]. In the present case, the following certainty-equivalent controller can be derived by applying again the approximation scheme of [Moore et al., 1999] to linearize the multiplicative noise terms

$$\vec{u}_t^D = -L_t[\vec{\hat{a}}_t]\,\vec{x}_t \tag{13}$$

with

$$L_t[\vec{\hat{a}}_t] = \left(R + G^T\,S_t\,G + \sum_i C_i^T\,G^T\,S_t\,G\,C_i\right)^{-1} G^T\,S_t\,F[\vec{\hat{a}}_t] \tag{14}$$

The matrix S_t can be easily computed by solving the pertinent Riccati equation by means of established standard methods

$$S_t = Q + F[\vec{\hat{a}}_t]^T\,S_t\,F[\vec{\hat{a}}_t] - F[\vec{\hat{a}}_t]^T\,S_t\,G\left(R + G^T\,S_t\,G + \sum_i C_i^T\,G^T\,S_t\,G\,C_i\right)^{-1} G^T\,S_t\,F[\vec{\hat{a}}_t] \tag{15}$$

In case of perfect knowledge of system parameters and full state observability (i.e. $\vec{y}_t = \vec{x}_t$), the above solution can be shown analytically to be optimal [Kleinman, 1969]. In case of partial observability, equations (13)-(15) can only be part of an approximative solution [Moore et al., 1999; Todorov, 2005]. In case of parameter uncertainties, the additional difficulty arises that successful system identification in the closed loop cannot be guaranteed generically [Kumar, 1983, 1990; van Schuppen, 1994; Campi and Kumar, 1996; Campi, 1997]. Here, we only considered unknown system parameters in the state transition matrix, but the algorithm is also applicable in the face of general parameter uncertainties provided that questions of stability and closed-loop identification are clarified on a case-to-case basis. These difficulties are omnipresent in adaptive control, simply due to the immense complexity of the topic. Indeed, the vast majority of practical applications in the field that have proven to be very robust lack a thorough mathematical treatment and convergence proof [Åström and Wittenmark, 1989; Sastry and Bodson, 1989]. Here, we compare the performance of the proposed algorithm with other non-adaptive control algorithms considering a multiplicative noise structure (**Fig. S1**).

1.3 Connection to Non-adaptive Optimal Control Models

From a theoretical point of view it seems desirable to design a unified control scheme, where "learning control" equals "standard control" in the absence of parameter uncertainties, and "learning" converges to "standard" over time. The proposed approach in sections (1.1) and (1.2) fulfils this

[3]When neglecting u^C and u^P, the certainty-equivalence principle leads to the control scheme of the *self-tuning regulator* [Åström and Wittenmark, 1989], i.e. the current parameter estimate $\vec{\hat{a}}_t$ is employed for control as if it were the true parameter $\vec{a}$, while the uncertainty P_t^a of the estimate is ignored for control purposes

criterion and is, therefore, consistent. However, the question arises in how far our "standard control" corresponds to non-adaptive optimal control schemes in the literature [Todorov and Jordan, 2002].

In contrast to previous non-adaptive optimal control schemes, we have postulated optimality not for an action sequence on a predefined time interval T, but for an indefinite runtime. In the literature this is known as *infinite horizon control* as opposed to *finite horizon control* with a predefined time window T [Stengel, 1994]. We have chosen this approach, because the finite horizon setting does not allow the implementation of adaptivity in a straightforward manner[4]. Additionally, a noisy infinite horizon model naturally predicts variable movement durations, while variable movement times T in a finite horizon model have to be introduced by deliberately drawing T from a Gaussian distribution. Remarkably, the proposed control architecture is able to reproduce the speed-accuracy trade-off in the presence of multiplicative noise and can account for the speed-target distance relationship as found experimentally (cf. **Fig. S2**). However, it remains to be tested in how far actual motor behavior can be accounted for by time-independent optimal policies, and whether and in which contexts time-dependent policies are indispensible. A straightforward generalization of the present algorithm would be to allow for state-dependent feedback gains (see [Jazwinsky, 1970] for state-dependent Ricatti equation (SDRE) control).

1.4 Arm Model

In the experiment described in the main text, human subjects steered a cursor on a screen to designated targets. Since the hand movement in the experiment was very confined in space ($8cm$), the hand/cursor system is modeled with linear dynamic equations. Following previous studies [Todorov, 2005; Winter, 1990] the hand is modeled as a point mass m with two-dimensional position $\vec{p}^H(t)$ and velocity $\vec{v}^H(t) = \dot{\vec{p}}^H(t)$. The combined action of all muscles on the hand is represented by the force vector $\vec{f}(t)$. The neural control signal $\vec{u}(t)$ is transformed to this force through a second-order muscle-like low-pass filter with time constants τ_1 and τ_2. In every instant of time, the

[4]In the finite horizon setting [Harris and Wolpert, 1998; Todorov and Jordan, 2002] the argument goes that during movement execution there are no explicit constraints apart from avoiding excessive control signals, and only when the target is reached accuracy becomes an issue, i.e. in mathematical terms the cost matrix Q is zero during the movement and takes the value $Q = Q_f$ at the end of the movement. To solve this finite-horizon optimal control problem, the constraint Q_f has to be propagated back through time via the Riccati recursion determining the optimal feedback gain L_t at each point in time. Thus, target-related accuracy requirements ("minimum end-point variance" [Harris and Wolpert, 1998]) shape the optimal trajectory. Obviously, this procedure cannot be carried over to the adaptive case in a straightforward manner, since the Riccati recursion presupposes knowledge of the true system parameters to determine the entire sequence of optimal feedback gains and to backpropagate terminal constraints through time. Finally, going one step back and trying to solve the pertinent Bellman equation under parameter uncertainties is also not an option due to mathematical intractability. In contrast, a stationary feedback controller in an infinite horizon setting easily carries over to the adaptive case by re-computing the "stationary" control law in each time step, thus, considering the most recent parameter estimate. This also implies that such an adaptive controller is applicable on-line.

hand motion is mapped to a cursor motion on a screen by use of a manipulandum. This mapping can either be straightforward, or a rotation ϕ between hand movement and cursor movement can be introduced. Neglecting the dynamics of the frictionless manipulandum, the cursor position $\vec{p}(t)$ is connected to the hand position via a simple rotation operator $\mathscr{D}_\phi$, i.e. $\vec{p}(t) = \mathscr{D}_\phi \vec{p}^{\,H}(t)$. Put together, this yields the following system equations

$$\ddot{\vec{p}}(t) = \frac{1}{m} \mathscr{D}_\phi \vec{f}(t) \tag{16}$$

$$\tau_1 \tau_2 \ddot{\vec{f}}(t) + (\tau_1 + \tau_2)\dot{\vec{f}}(t) + \vec{f}(t) = \vec{u}(t) \tag{17}$$

Equation (17) can be written equivalently as a pair of coupled first-order filters with outputs g and f. This allows to formulate the state space vector $\vec{x} \in \Re^{10}$ as

$$\vec{x}(t) = \left[\begin{array}{cccccccccc} p^x(t) & v^x(t) & f^x(t) & g^x(t) & p_x^{\text{TARGET}} & p^y(t) & v^y(t) & f^y(t) & g^y(t) & p_y^{\text{TARGET}} \end{array} \right]^T$$

where the target location is absorbed in the state vector. When discretizing the above equations with time bin Δ the following system matrices are obtained

$$F[\phi] = \begin{pmatrix} 1 & \Delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{\Delta}{m}\cos(\phi) & 0 & 0 & 0 & 0 & \frac{\Delta}{m}\sin(\phi) & 0 & 0 \\ 0 & 0 & 1 - \frac{\Delta}{\tau_2} & \frac{\Delta}{\tau_2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 - \frac{\Delta}{\tau_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \Delta & 0 & 0 & 0 \\ 0 & 0 & -\frac{\Delta}{m}\sin(\phi) & 0 & 0 & 0 & 1 & \frac{\Delta}{m}\cos(\phi) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \frac{\Delta}{\tau_2} & \frac{\Delta}{\tau_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \frac{\Delta}{\tau_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad G = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \frac{\Delta}{\tau_1} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{\Delta}{\tau_1} \\ 0 & 0 \end{pmatrix}$$

A crucial part of the dynamic equations of the arm model is the multiplicative noise structure [Harris and Wolpert, 1998]. Following [Todorov, 2005], control-dependent noise is generated by multiplying the control signal $\vec{u}_t$ with a stochastic matrix and a scaling parameter Σ_u

$$G\,\Sigma_u \begin{pmatrix} \sigma_t^{(1)} & \sigma_t^{(2)} \\ -\sigma_t^{(2)} & \sigma_t^{(1)} \end{pmatrix} \vec{u}_t$$

Accordingly, the matrices C_i $(i = 1, 2)$ are set to

$$C_1 = \begin{pmatrix} \Sigma_u & 0 \\ 0 & \Sigma_u \end{pmatrix} \qquad C_2 = \begin{pmatrix} 0 & \Sigma_u \\ -\Sigma_u & 0 \end{pmatrix}$$

Feedback is provided by delayed and noisy measurement of position and velocity of the cursor, and proprioception. The system formulation of equation (1) already implied a feedback delay of one time step, since the sensory feedback y_t is received after generation of the control signal u_t.

Including an additional delay of d time steps can be achieved easily by further augmenting the state space as described in the literature [Todorov and Jordan, 2002]. For the present simulations a feedback delay of $150ms$ was assumed. This yields the feedback equation

$$\vec{y}_t = \begin{bmatrix} p^x_{t-d} & v^x_{t-d} & f^x_{t-d} & p^y_{t-d} & v^y_{t-d} & f^y_{t-d} \end{bmatrix}^T + \chi_t$$

When introducing a parameter uncertainty as in (1), long feedback delays can lead to substantial problems in the process of joint estimation, such as instability and oscillations in the parameter estimate. In fact, the joint estimation of states and parameters can only be accomplished if the parameters change on a time-scale well below the delay time. To circumvent these problems we simply iterated the Kalman equations (4)-(5) at every time step t from $t' = 0$ to $t' = t$ by setting $\vec{\hat{a}}_{t'} = \vec{\hat{a}}_t$ and $P^a_{t'} = P^a_t$ for $t' = 0$. This solution is still causal, but makes explicit use of the knowledge that the unknown parameters are constant throughout the control task.

2 Model Fit

For the investigated visuomotor learning task, the cost function J is given by

$$J = \frac{1}{2}\,\mathbb{E}\left[\sum_{t=0}^{\infty}\left\{\vec{x}_t^T Q\,\vec{x}_t + \vec{u}_t^T\,R\,\vec{u}_t\right\}\right] \tag{18}$$

with

$$Q = \begin{pmatrix} w_p^2 & 0 & 0 & 0 & -w_p^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_v^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -w_p^2 & 0 & 0 & 0 & w_p^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_p^2 & 0 & 0 & 0 & -w_p^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & w_v^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -w_p^2 & 0 & 0 & 0 & w_p^2 \end{pmatrix} \qquad R = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}$$

Following equation (13), the certainty-equivalent controller then takes the form

$$\vec{u}_t^D = -L_t[\hat{\phi}_t]\,\vec{\hat{x}}_t \tag{19}$$

where L_t is computed according to equation (14) and the estimates $\vec{\hat{x}}_t$ and $\hat{\phi}_t$ are procured by the Unscented Kalman Filter that operates in the augmented state space

$$\vec{\hat{\mathfrak{x}}}_t = \begin{bmatrix} \vec{\hat{x}}_t \\ \hat{\phi}_t \end{bmatrix} \tag{20}$$

An example of how the parameter estimate evolves within a trial can be seen in **Fig. S3**.

As discussed in the previous section, optimal adaptive controllers generally have to be designed in a problem-specific fashion. To this end, issues of cautious and probing control have to be tackled. In the present case, the probing control problem can safely be neglected, since center-out movements automatically entail better system identification of the rotation parameter $\hat{\phi}$. However, it is intuitively clear that cautiousness (slowing down) is expedient in the presence of very slow parameter identification and high feedback gains. In line with previous work in the engineering sciences [Chakravarty and Moore, 1986; Papadoulis et al., 1987; Papadoulis and Svoronos, 1989], cautiousness is introduced here heuristically by means of an innovation-based "cautious factor". The basic idea is to tune down feedback gains if the parameter innovation is high, i.e. if the current parameter estimate yields poor predictions. The parameter innovation can be obtained by calculating the Robbins-Munro innovation update [Ljung and Sönderström, 1983]

$$I_{t+1}^{\hat{\phi}} = (1-\alpha)\, I_t^{\hat{\phi}} + \alpha\, K_t^{\hat{\phi}} \left[\vec{y}_t - \vec{\hat{y}}_t^-\right] \left[\vec{y}_t - \vec{\hat{y}}_t^-\right]^T (K_t^{\hat{\phi}})^T$$

where $K_t^{\hat{\phi}}$ corresponds to the respective entries of the Kalman gain matrix K_t from the Unscented Kalman Filter working on the augmented state space[5]. An example of the innovation estimator can be seen in **Fig. S3**. Importantly, the parameter innovation $I_t^{\hat{\phi}}$ can be used to adjust the feedback gain L_t. In the present case, the feedback gain is effectively determined by the two cost parameters w_p and w_v. They specify the controller's drive to regulate the position towards the target position, while trying to regulate the velocity to zero. Since the cost function is invariant with regard to a scaling factor (i.e. r can be set arbitrarily), cautiousness can be introduced most generally by means of two effective cost parameters

$$\begin{aligned} \tilde{w}_t^p &= \frac{w_p}{1+\lambda_p\, I_t^{\hat{\phi}}} \\ \tilde{w}_t^v &= w_v \left(1+\lambda_v\, I_t^{\hat{\phi}}\right) \end{aligned}$$

with constants λ_p and λ_v. While the original cost function is still determined by w_p and w_v, the effective cost parameters $\tilde{w}_t^p$ and $\tilde{w}_t^v$ (i.e. the effective cost matrix $\tilde{Q}_t$) can be used to calculate the (approximatively) optimal adaptive feedback gain. The optimal adaptive controller then takes the form

$$\vec{u}_t^{\,opt} = -\tilde{L}_t[\hat{\phi}_t]\, \vec{\hat{x}}_t \tag{21}$$

with $\tilde{L}_t[\hat{\phi}_t]$ from equation (14) calculated on the basis of $\tilde{Q}_t$. In the absence of parameter uncertainty ($I_t^{\hat{\phi}} = 0$) this yields a standard LQG control scheme. For infinitely high innovations one gets $\tilde{w}_t^p \to 0$ and $\tilde{w}_t^v \to \infty$, i.e. the controller halts.

Finally, the model was tested on the experimental movement data. To this end, four effective control parameters and three noise parameters of the model were adjusted to fit the mean trajectory and variance of 90°-transformation trials. The parameters of the arm model were taken from the literature [Todorov, 2005]. The obtained parameter set was then used to predict trajectories,

[5] We set $\alpha = 0.1$

speed profiles, angular speed and variance for all intermediary transformation angles and standard movements.

The parameter set to fit and predict the human movement data was as follows:

Arm Parameters	$\tau_1 = \tau_2 = 40ms$
	$m = 1kg$
Control Parameters	$w_p = 1$
	$w_v = 0.1$
	$r = 0.0001$
	$\lambda_p = 2 \cdot 10^4$
	$\lambda_v = 1 \cdot 10^4$
Noise Parameters	$\Omega_\xi = 0$
	$\Omega_\chi = \Big(0.1 \text{ diag}([1cm \;\; 10cm/s \;\; 100cN \;\; 1cm \;\; 10cm/s \;\; 100cN])\Big)^2$
	$\Omega_\nu = 10^{-7}$
	$\Sigma_u = 0.7$

The average costs appertaining to this parameter set mounted up to $J = 7880 \pm 60$. In contrast, a certainty-equivalent controller ($\lambda_p \equiv \lambda_v \equiv 0$) yields $J^{CE} = 8910 \pm 70$. This clearly shows that for fast movements it is optimal to behave cautiously (cf. **Fig. S5**).

References

[1] Åström KJ, Wittenmark B (1989) Adaptive Control, Addison-Wesley Publishing Company

[2] Bar-Shalom Y (1981) Stochastic dynamic programming: caution and probing, IEEE Transactions on automatic control AC-26:1184-1195

[3] Bar-Shalom Y, Tse E (1974) Dual effect, certainty equivalence, and separation in stochastic control, IEEE Transactions on automatic control AC-19:494-500

[4] Campi MC (1997) Achieving optimality in adaptive control: the "bet on the best" approach, Proc. 36th Conf. on Decision and Control, pp. 4671-4676

[5] Campi MC, Kumar PR (1996) Optimal adaptive control of an LQG system, Proc. 35th Conf. on Decision and Control, pp. 349-353

[6] Chakravarty A, Moore JB (1986) Aircraft flutter suppression via adaptive LQG control, Proc. American Control Conf., pp. 488-493

[7] Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning, Nature 394:780-784

[8] Haykin S (2001) Kalman filtering and neural networks, John Wiley and Sons, Inc., New York

[9] Jazwinsky A (1970) Stochastic processes and filtering theory, New York: Academic Press

[10] Julier SJ, Uhlmann JK, Durrant-Whyte H (1995) A new approach for filtering nonlinear systems, Proc. Am. Control Conference, pp. 1628-1632

[11] Kalman RE (1960) A new approach to linear filtering and prediction problems, Transactions of ASME, Ser. D, Journal of basic engineering 82:34-45

[12] Kleinman D (1969) Optimal stationary control of linear systems with control-dependent noise, IEEE Transactions on automatic control AC-14(6):673-677

[13] Kumar PR (1983) Optimal adaptive control of linear-quadratic-Gaussian systems, SIAM J. Control and Optimization 21(2):163-178

[14] Kumar PR (1990) Convergence of adaptive control schemes using least-squares parameter estimates, IEEE Trans. on Automatic Control, AC-35(5):416-424

[15] Moore JB, Zhou XY, Lim AEB (1999) Discrete time LQG controls with control dependent noise, Systems & Control Letters 36:199-206

[16] Papadoulis AV, Svoronos SA (1989) A MIMO cautious self-tuning controller, AIChE 35:1465-1472

[17] Papadoulis AV, Tsiligiannis CA, Svoronos SA (1987) A cautious self-tuning controller for chemical processes, AIChE 33:401-409

[18] Sastry S, Bodson M (1989) Adaptive control. Stability, convergence, and robustness, Prentice-Hall Information and System Sciences Series

[19] Stengel RF (1994) Optimal control and estimation, Dover Publications

[20] Sutton RS, Barto AG (1998) Reinforcement Learning, MIT Press, Cambridge, Massachusetts

[21] Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system, Neural Comp. 17:1084-1108

[22] Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination, Nat Neurosci. 5:1226-1235

[23] van Schuppen JH (1994) Tuning of Gaussian stochastic control systems, IEEE Trans. on Automatic Control, AC-39:2178-2190

[24] Winter DA (1990) Biomechanics and motor control of human movement, John Wiley and Sons, Inc., New York

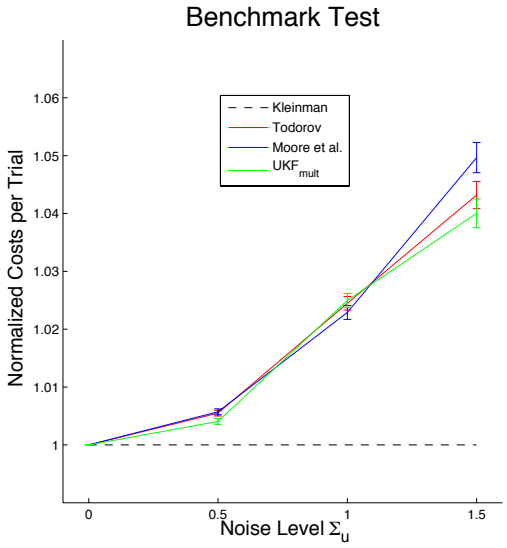

Figure S1. Benchmark Test. The performance of the proposed algorithm ("UKF_{mult}") was measured for different magnitudes of multiplicative noise in a standard center-out reaching task to allow for comparison with existing approximation schemes by [Todorov, 2005] and [Moore et al., 1999]. In the absence of observation noise, Kleinman [Kleinman, 1969] calculated the optimal solution to the posed control problem and, thereby, provided a lower bound. All other algorithms are run in the presence of observation noise and their average costs per trial are normalized by this lower bound. All three algorithms achieve roughly the same performance.

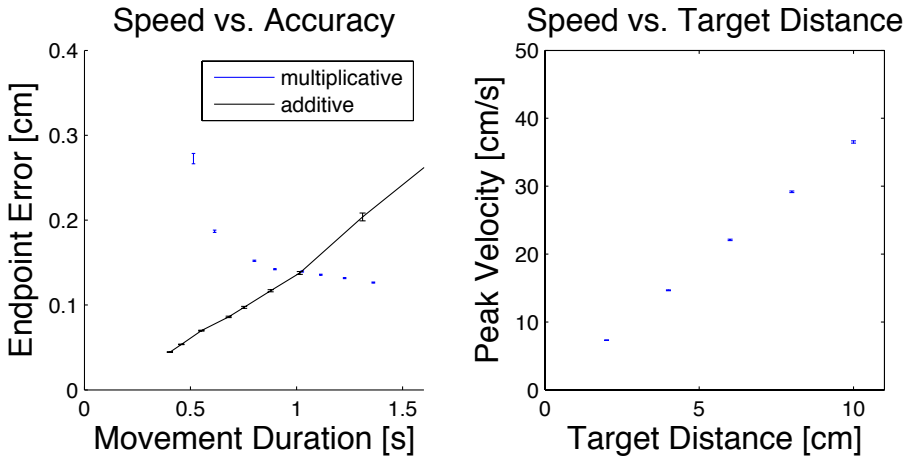

Figure S2. Speed-Accuracy Trade-off. Due to control-dependent multiplicative noise, fast movements entail higher inaccuracy as measured by positional standard deviation once the target is reached. This relationship cannot be explained in an additive noise scenario. Model predictions are in line with the model of Todorov [Todorov, 2005]. (b) Speed vs. Target Distance. The model predicts a linear relationship between target distance and peak velocity as found experimentally. (cf. [Krakauer et al., 2000, Fig. 2D]).

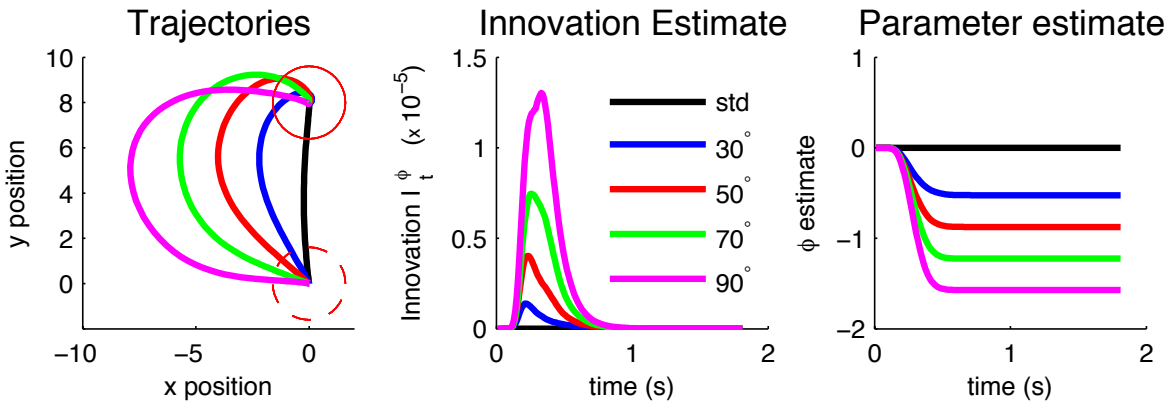

Figure S3. Adaptation of model parameters. The left panel shows trajectories when the controller adapts to different unexpected visuomotor transformations: 0° black, 30° blue, 50° red, 70° green and 90° magenta. The middle panel shows how the innovation estimate evolves within a trial. Due to feedback delay, initially there is no mismatch detected. After the delay time the innovation estimator detects parameter mismatch. Once the correct parameter estimate can be achieved innovations return to zero again. The right panel shows evolution of the parameter estimate within a trial. The different rotation angles are estimated corresponding to different experienced visuomotor rotations.

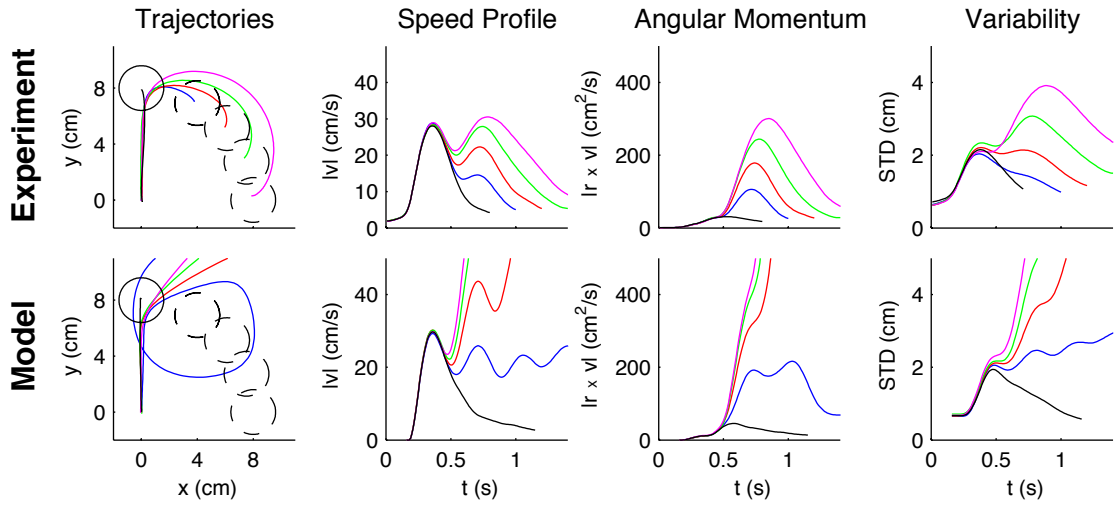

Figure S4. Non-adaptive optimal control model. When the model is not allowed to track the rotation parameter, the controller becomes quickly unstable. The trajectories diverge.

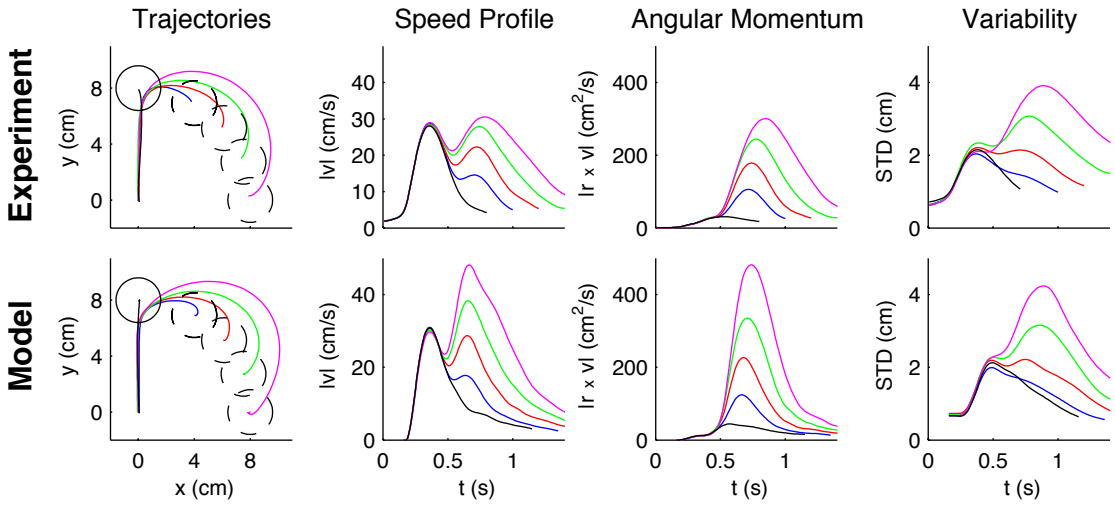

Figure S5. Adaptive controller without "cautiousness". When the cautiousness parameters are set to zero, the controller acts much faster in the second part of the movement, not only leading to higher speeds, but importantly also to higher costs (compare Section 2).