Adaptation is sometimes viewed as a process in which the nervous system learns to predict and cancel effects of a novel environment, returning movements to near baseline (unperturbed) conditions. An alternate view is that cancellation is not the goal of adaptation. Rather, the goal is to maximize performance in that environment. If performance criteria are well defined, theory allows one to predict the reoptimized trajectory. For example, if velocity-dependent forces perturb the hand perpendicular to the direction of a reaching movement, the best reach plan is not a straight line but a curved path that appears to overcompensate for the forces. If this environment is stochastic (changing from trial to trial), the reoptimized plan should take into account this uncertainty, removing the overcompensation. If the stochastic environment is zero-mean, peak velocities should increase to allow for more time to approach the target. Finally, if one is reaching through a via-point, the optimum plan in a zero-mean deterministic environment is a smooth movement but in a zero-mean stochastic environment is a segmented movement. We observed all of these tendencies in how people adapt to novel environments. Therefore, motor control in a novel environment is not a process of perturbation cancellation. Rather, the process resembles reoptimization: through practice in the novel environment, we learn internal models that predict sensory consequences of motor commands. Through reward-based optimization, we use the internal model to search for a better movement plan to minimize implicit motor costs and maximize rewards.

**Key words:** motor learning; motor adaptation; cerebellar damage; ataxia; optimal control; internal model

**Introduction**

Studies of motor adaptation often rely on externally imposed perturbations to induce errors in behavior. For example, in a reaching task, a robotic arm may be used to introduce perturbations. On each trial, “error” is measured as the difference between the observed trajectory on that trial and some average behavior in a baseline condition (before perturbations were imposed). Implicit in many works is the idea that adaptation proceeds by reducing such errors. However, this approach makes the fundamental assumption that the baseline movements are somehow the optimal movements in all conditions. This assumption is demonstrably false. For example, when a spring-like force makes it so that the path of minimum resistance between two points is not a straight line but a curved path, people adapt to the novel dynamics by reaching along that curved path (Uno et al., 1989; Chib et al., 2006). Therefore, at least in these extreme cases, motor error as classically defined does not drive adaptation.

The above example highlights the idea that the purpose of our movements is, at least to a first approximation, to acquire rewarding states (e.g., reach the end point accurately) at a minimum cost. Each environment has its own cost and reward structure. The trajectory or feedback response that was optimum in one environment is unlikely to remain optimum in the new environment (Wang et al., 2001; Burdet et al., 2001; Diedrichsen, 2007; Emken et al., 2007).

Recent advances in optimal control theory (Todorov, 2005) allowed us to revisit the well studied reach adaptation paradigm of force fields and make some theoretical predictions about what the adapted trajectory should look like in each field. In this framework, the problem is to maximize performance. To do so, one of the required steps is to identify (i.e., build a model of) the novel environment so one can accurately predict the sensory consequences of motor commands. A second required step is to use this internal model to find the best movement plan.

As we explored the theory, we found that it made rather interesting predictions in conditions in which the force field was stochastic. Stochastic behavior of an environment introduced uncertainty in the internal model, and the controller that attempted to maximize performance took this uncertainty into account as it generated movement plans. An intuitive example is lifting a cup of hot coffee in which a lid obscures the amount of liquid. If one cannot see the amount of liquid, one is uncertain about its mass. Lifting and drinking from this cup will tend to be slower, particularly as it reaches our mouth.

In the force-field task, one can introduce uncertainty by making the environment stochastic. To maximize performance (reach to the target in time), the theory takes into account this uncertainty and reoptimizes the reach plan. We asked whether adaptation proceeded by returning trajectories toward a baseline or whether adaptation more resembled a process of reoptimization.
Materials and Methods

Our volunteers were healthy right-handed individuals [26.5 ± 5.2 years old (mean ± SD)]. Protocols were approved by the Johns Hopkins School of Medicine Institutional Review Board, and all subjects signed a consent form. Volunteers sat on a chair in front of a robotic arm and held its handle (Hwang et al., 2003) that housed a light-emitting diode (LED). A white screen was positioned immediately above the horizontal plane of the robot/arm on which an overhead projector (refresh rate, 70 Hz; EP739; Optoma, Milpitas, CA) painted the screen. In experiment 1, the projector displayed a cursor to represent hand position, and the LED was off. In all other experiments, the handle LED was on, and it represented hand position. Movements were made only to a single direction (90°, straight away from the body along a line perpendicular to the frontal plane), along the midline of the subject’s body.

Experiment 1. In this experiment, we repeated the standard force-field adaptation paradigm (Shadmehr and Mussa-Ivaldi, 1994). Subjects (n = 28) trained for 3 consecutive days, practicing reaching in a single direction. They were provided with a target at 9 cm (target was a 5 mm square) and were rewarded (via a target explosion) if they completed the movement in 450 ± 50 ms. Timing feedback was provided in the form of a blue-colored target if the movement was slower than required. After completion of the movement, the robot pulled the hand back to the starting position.

On day 1, the experiment started with a familiarization block of 150 trials without any force perturbations (null). This was followed by four blocks of field training. The second and third days each consisted of four blocks of field training (each block was 150 trials). Overall, subjects performed ∼2400 field trials.

Let us label the forces produced by the robot as $f = D \dot{x}$, where $x$ is hand velocity in Cartesian coordinates. On each trial, $D$ was drawn from a normal distribution such that $D = \bar{D} + \delta \bar{x} = \bar{D} + \hat{D} \delta$, where $\bar{D}$ is the mean of the distribution (see below) and $\delta$ is a normally distributed scalar random variable with zero mean and variance $\sigma^2$. The subjects were divided in four groups. For the first and second groups, the variance of the field was zero (i.e., the field did not vary from trial to trial): for group $1 (n = 7)$, $\bar{D} = [0, 13; -13, 0]$ Ns/m [a clockwise (CW) perturbation]; for group $2 (n = 7)$, $\bar{D} = [0, -13; 13, 0]$ Ns/m [a counterclockwise (CCW) perturbation]. Group 3 ($n = 7$) and group 4 ($n = 7$) practiced in a field with the same mean as groups 1 and 2, respectively, but with $\sigma = 0.3$. Groups 3 and 4 experienced an additional block of 50 movements at the end of the third day. The variance of the field was set to zero during this last block. The data from this last block allowed us to compare the behavior of the subjects from groups 1 and 2 with groups 3 and 4 in precisely the same environment.

Experiment 2. In this experiment, we simply presented a field that had a zero mean but a non-zero variance. The distance between the starting and the target was 18 cm. The target was projected as a box, the size of which was 5 × 5 mm. The movement direction was the same as that in experiment 1. Subjects ($n = 18$) were instructed to complete their reach within 600 ± 50 ms. A score indicating the number of successful trials was displayed on the screen. To encourage performance, we paid the subjects based on their score.

This experiment was composed of eight blocks of trials (each block was 150 trials) on a single day. The first block was a familiarization block with no force perturbations. This was followed by seven field training blocks. The force field was defined as $f = D \delta \bar{x}$, where $\bar{D} = [0, 13; -13, 0]$ Nm/s and $\delta$ is a normally distributed random variable with zero mean and variance $\sigma^2$. The subjects were divided in two groups. In the small variance group, for the first four blocks of field training, $\sigma = 0.3$. In the large variance group, for the first four blocks of field training, $\sigma = 0.6$. For the remaining three blocks, $\sigma = 0$.

Experiment 3. To further test how variance in the environment affected movement planning, we considered a reaching task through a via-point. The via-point target and the final target were positioned at a distance of 9 and 18 cm from the starting position, respectively. Each target was 5 × 5 mm. The starting, via-point, and final targets were positioned on a straight line, requiring a movement at 90°. The subjects ($n = 22$) were instructed to reach by passing through the via-point target at $t = 400 ± 50$ ms and then stop at the final target no later than $t = 1.0$ s. If both timing constraints were met, then both targets exploded at the completion of the movement. Otherwise, at the completion of the movement, the subject was provided timing feedback for the via-point with arrows and for the final target by color. No timing feedback was provided during the movement. A score indicating the number of explosions was constantly displayed on the screen. To encourage performance, we paid the subjects based on their score.

This experiment was composed of eight blocks of trials (each block was 150 trials) on a single day. The first block was a familiarization block with no force perturbations. This was followed by seven field training blocks. The force field was defined as $f = D \delta \bar{x}$, where $\bar{D} = [0, 13; -13, 0]$ Nm/s and $\delta$ is a normally distributed random variable with zero mean and variance $\sigma^2$. The subjects were divided in two groups. In the small variance group, for the first four blocks of field training, $\sigma = 0.3$. In the large variance group, for the first four blocks of field training, $\sigma = 0.6$. For the remaining three blocks, $\sigma = 0$.

Movement analysis. Movement initiation was defined as the time when the hand velocity crossed the threshold of 3 cm/s. Hand paths were aligned at the starting position. In the first experiment, we computed overcompensation by forming a difference trajectory between the null and field conditions: for each subject, we computed the average hand path over the last 50 trials of field training and then subtracted the x-position (axis perpendicular to the direction of the target) of this trajectory from the x-position of the subject’s own average hand path in the last 50 trials of the null field. To compute how changes in field variance affected hand speed in experiments 2 and 3, we computed an average speed profile from the mean hand path of the last 50 trials of each block. The speed profiles were normalized with the peak speed in the first familiarization block to remove the effect of personal maximum speed biases on the average speed profile.

Modeling and simulations. We used stochastic optimal feedback control (OFC) to model reaching (Todorov and Jordan, 2002; Todorov, 2005). In this framework, the trajectory of a reach is determined by three components: an optimal controller that generates motor commands, an internal model that predicts the sensory consequences of those commands, and a motor plant/environment that reacts to those sensory consequences. Noise is signal dependent with an SD that grows with the size of the motor commands (Harris and Wolpert, 1998; Jones et al., 2002; van Beers et al., 2004). Our theoretical work here is novel only in the sense that it considers the problem of optimal control in the context of uncertainty about the internal model. To tackle this problem, we extended the approach introduced by Todorov (2005). We found that the problem of model uncertainty was a dual to the problem of control with signal-dependent noise. The mathematics that we used to solve the uncertainty problem is very similar to those used by Todorov (2005) to solve the signal-dependent noise problem. Here, we only outline the procedures and leave the derivations for the supplemental material (available at www.jneurosci.org).

Consider a linear dynamical system: $x_{t+1} = Ax_t + Bu_t$, Here, $x_t$ is the state of the system at time $t$, $u_t$ is the control signal input to the system, and the matrices $A$ and $B$ are the dynamics of the system. Previous studies have used deterministic model parameters $A$ and $B$, leaving no room for representation of the learner’s uncertainty about these parameters. Here, we represent the model parameter $A$ as a stochastic variable, leading to the following equation:

$$x_{t+1} = (A + V_t)x_t + Bu_t = Ax_t + Bu_t + V_t x$$

where $V_t$ is a Gaussian random variable with mean zero and variance $Q_t$. One can see that the uncertainty in parameter $A$ is state-dependent noise. We derived the following optimal feedback controller and optimal state estimator with model noise:

Dynamics: $x_{t+1} = Ax_t + Bu_t + \xi + \sum_{i=1}^{t} e_i/C_{xi}$

(1)

Observation: $y_t = Hx_t + \omega_t$

(2)

Cost: $0 \leq x_t^T Q x_t + u_t^T R u_t$

(3)
where $y_i$ is the observation made by the system, $H$ is the observation matrix, and $e_i$, $e_n$, and $w_o$ are Gaussian random variables with mean 0 and variance 1 representing the additive and multiplicative state variability and the measurement noise, respectively. $C_i$ is the scaling matrices for the state- and control-dependent noise for each noise source. $Q_i$ is the weight matrix of state cost, and $R$ is the weight matrix of motor cost. $x_i$ is the actual state of the system that is not available to the controller. The controller only has an estimate of the state $\hat{x}_i$ available through the state estimation process. For analytical tractability, the state is assumed to be updated according to a linear recursive filter: $\hat{x}_{i+1} = A\hat{x}_i + Bw_i + K_i(y_i - H\hat{x}_i)$, where $K_i$ is the Kalman gain. The optimal control policy is of the following form:

$$u_i = -L_i\hat{x}_i,$$

where $L_i$ is the time varying feedback-gain matrix that determines the controller’s response.

Optimal control provides closed-form solutions for only linear dynamical systems. We therefore modeled the arm for the single direction of movement as a point mass in Cartesian coordinates. The components of state were $x(t) = [p_x(t), p_y(t), p_z(t), f_x(t), f_y(t), f_z(t), T_x, T_y, T_z]$, where $p$ is hand position, $f$ is force, and $T$ is target position. The cost function was as follows:

$$w_x(u_x'^2 + u_y'^2) \quad 0 \leq t < T$$

$$w_y((p_x - T_x)^2 + (p_y - T_y)^2)$$

$$+ w_y(p_x'^2 + p_y'^2) + w_y(u_x'^2 + u_y'^2) \quad MT_e \leq t < MT + MT_H,$$

where $MT_e$ is the desired movement time and $MT_H$ is the time interval after movement completion for which the controller is supposed to hold position at the target.

Whereas in the mathematics we could solve the problem only for the case in which variance was a measure of within-trial noise in the parameter $D$, because of safety concerns, in our experiments we held the noise constant during a trial and only changed it from trial to trial.

The force-field parameters used in the simulations were the same as those used for the experiments. We found that even with extensive training, subjects learn only ~80% of the field. For example, in channel trials in which we and others have measured the forces that subjects produce, the force trajectory is at most 80–82% of the imposed field (Scheidt et al., 2000; Hwang et al., 2006; Smith et al., 2006). To account for this, in the simulations, the environment produced forces that were identical to those produced by our robot, but adaptation of the subject was modeled as an internal model that predicted a fraction of these forces $\hat{D} = \alpha D + \gamma D$, where $\alpha$ is the fraction and $\gamma$ is a normally distributed random variable with 0 mean and variance $\sigma^2$. For simulations of the via-point task, we set $\hat{D} = \gamma D$ (because the mean of the field was zero), and the state was extended to hold the via-point positions $TV_x$ and $TV_y$. The cost for the via-point task simulations were as follows:

$$w_x(u_x'^2 + u_y'^2) \quad 0 \leq t < MT$$

$$w_y((p_x - TV_x)^2 + (p_y - TV_y)^2)$$

$$+ w_y(p_x'^2 + p_y'^2) + w_y(u_x'^2 + u_y'^2) \quad t = MT_e,$$

$$+ w_y(p_x'^2 + p_y'^2) + w_y(u_x'^2 + u_y'^2) \quad MT_e \leq t < MT + MT_H,$$

where $MT_e$ is the “via-point time.”

**Results**

**Model predictions: reaching in a deterministic or stochastic field**

In the null field, the optimal control policy to move a mass to a target in a given amount of time is a straight-line trajectory (Fig. 1a, dashed-line trajectory) with a bell-shaped velocity profile. However, if the mass is moving in a velocity-dependent curl field that pushes it perpendicular to its direction of movement, then the best policy is a slightly curved movement (Fig. 1a, trajectory marked by “I”) that overcompensates for the initial curl forces.

![Figure 1](image-url)
That is, if the field pushes the hand to the right, the optimum policy is a trajectory that is to the left of baseline.

To see the rationale for this, we plotted the forces produced by the optimal controller and compared it to forces that must be produced if the mass is to move along a straight-line, minimum-jerk trajectory (Fig. 1b). The optimal controller produced less total force (Fig. 1b, $\int f^2 dt$) because by overcompensating early into the movement, when speeds were small, it could rely on the environmental forces to bring the mass back toward the target. Therefore, the curved, apparently overcompensating trajectory actually produced smaller total forces than a straight trajectory.

We arrived at this result by assuming that the learner had formed a perfect model of the force field. If the learner’s model predicted <100% of the effects of the field, then the combined effect of the controller and the environment are a more complicated trajectory. For example, suppose that the actual field is $\alpha \dot{x} + \beta \ddot{x}$, and the learner’s estimate is $\hat{\alpha} \dot{x} + \hat{\beta} \ddot{x}$ for $0 \leq \alpha \leq 1$. When $\alpha$ is small (e.g., 0.2), the field overpowers the controller and pushes the mass in the direction of its perturbation (Fig. 1a). Because the controller acquires a more accurate forward model ($\alpha$ becomes closer to 1), the trajectory becomes “S” shaped, displaying an apparent overcompensation. Previous work suggests that with training, subjects acquire a model accuracy of $\sim 0.8$ (Scheidt et al., 2000; Hwang et al., 2006; Smith et al., 2006). That is, in channel trials in which force output is quantified, the peak is $\sim 80\%$ of the field.

We produced these results using a linear, point-mass model of dynamics. We wondered whether overcompensation was also present for a more realistic nonlinear model of the two-link arm. In this case, closed-form solutions are not possible, but Li and Todorov (2007) have provided tools that aid the search process. We simulated reaching to various directions using a deterministic, nonlinear model of dynamics and again found that the controller produced an overcompensation in all directions (see supplemental material, available at www.jneurosci.org).

The above results were in a condition in which the task dynamics were invariant from trial to trial. How should planning change when one is uncertain of the strength of the force field? That is, what is the best way to perform the task if the field is stochastic? We returned to the linear model of dynamics and derived the closed-form solution for this stochastic optimal control problem (see supplemental material, available at www.jneurosci.org). We found that the overcompensation in the controller was a function of its uncertainty: as the variance of the field increased, the controller produced smaller amounts of overcompensation and had larger errors in the direction of the field (Fig. 1c). Furthermore, the simulations uncovered an interesting prediction: as field variance increased, the optimum plan no longer had a bell-shaped speed profile. Rather, the speed profile became skewed with a peak that was larger (Fig. 1c), slowing the hand as it approached the goal (i.e., the target).

In summary, theory predicted that in a deterministic curl field, the optimal trajectory is a slightly curved hand path that appears to overcompensate for the forces. In a stochastic field, the optimum trajectory loses its overcompensation tendencies, peak velocities become larger, and the timing of the peak shifts earlier in time, allowing the hand to approach the target more slowly.

**Experiment 1: deterministic versus stochastic curl fields**

In experiment 1, subjects experienced either a CW (groups 1 and 3) or a CCW (groups 2 and 4) velocity-dependent curl field. In groups 1 and 2, the field was constant from trial to trial. In groups 3 and 4, the field had the same mean as in the constant group but had a non-zero variance. Trajectories in the constant field did not return to the straight paths recorded in the null condition. Rather, they tended to show an overcompensation. Data from two subjects during various stages of adaptation are shown in Figure 2a. The hand paths tended to overcompensate early into the movement and then slightly undercompensate as the hand approached the target, resulting in S-shaped paths. This basic result was reported before (Thoroughman and Shadmehr, 2000). The only novelty here is that we see that this shape was attained on the first day of training, and was maintained for the duration of the 3 d experiment.

By the end of training on day 1, the success rates had reached levels observed in the null condition (Fig. 2b), and rates tended to improve with more days of training (repeated-measures ANOVA for the two groups and 3 d; effect of day: $F_{(2,2)} = 5.9, p = 0.009$; effect of group: $F_{(1,12)} = 2.76, p = 0.12$; interaction, $F_{(2,28)} = 1.67, p = 0.21$). We quantified overcompensation as the within-subject maximum perpendicular displacement from their null trajectory (Fig. 2c). Overcompensation would imply a negative measure for the CW group and a positive measure for the CCW group. The timing of this measure was early in the movement: in the CW
group, at 149, 139, and 154 ms (SD for each day, ~40 ms); and in the CCW group, at 176, 180, and 179 ms (SD, ~22 ms). For the CCW group, this measure was significantly greater than zero for all 3 d (paired t test; df = 6; p < 0.05 for each day). For the CW group, this measure was significantly less than zero for all 3 d (paired t test; df = 6; p < 0.05 for each day). Importantly, when we examined performance of each subject on each day, we found a significant correlation between overcompensation and success rates (CW field: r = 0.51, p < 0.02; CCW field: r = 0.61, p < 0.005). Therefore, both groups had hand trajectories that appeared to overcompensate for the field. As the overcompensation increased, there was a tendency for performance to improve.

Theory had predicted that overcompensation should disappear and velocity peaks should rise as the field acquires stochastic properties. We trained a new set of volunteers (groups 3 and 4) in a field that had the same mean as in groups 1 and 2, respectively, but with a non-zero variance. Figure 3a shows the average hand trajectories in the last 50 trials of each day for subjects who trained in the deterministic ($\sigma_d$) or stochastic ($\sigma_s$) CW fields. In the stochastic field, movements gradually lost their overcompensation (Fig. 3b) (overcompensation in the $\sigma_s$ vs $\sigma_d$ group on day 3; p < 0.05; df = 12). There was one additional data point for the last block on the last day of training for the $\sigma_d$ group. During this block, subjects from the $\sigma_s$ group were exposed to a zero variance field. That is, in this block of 50 trials, the environments for the $\sigma_s$ and $\sigma_d$ groups were identical. Despite this, the overcompensation remained significantly smaller for the stochastic group (p < 0.05; df = 12). Because the $\sigma_s$ groups kept the success rate high in the test block, the change of the trajectory is not a result of interference from the force perturbations.

Another prediction of the theory was that hand trajectories in the $\sigma_s$ group should show larger errors in the direction of the force perturbations (Fig. 1c): we quantified the interaction between the early overcompensation and the late undercompensation as the within-subject difference in the signed area enclosed between the hand trajectory in the training session and the hand trajectory in null session (Fig. 3c, inset, A1–A2). Simulations had predicted that with increased field variance, this parameter should become more positive. We observed this tendency in our subjects: for the parameter A1–A2, the differences between $\sigma_s$ and $\sigma_d$ were significantly different in day 3 (p < 0.02; df = 12), as was the difference between $\sigma_s$ (day 3) and $\sigma_d$ (day 3: test) (p < 0.02). As overcompensation declined, performance in the $\sigma_s$ group improved (Fig. 3d).

Finally, the model had predicted that the stochastic and deterministic groups would show different speed profiles (Fig. 1c). Figure 3f shows the speed profiles for the last training block on each day of training, and Figure 3e shows the peak speed distribution on each day. The speed profiles were normalized with respect to the individual peak speed in the last 50 trials of the null distribution on each day. The speed profiles were normalized with respect to the individual peak speed in the last 50 trials of the null condition on each day. The speed profiles were normalized with respect to the individual peak speed in the last 50 trials of the null block on day 1. By day 3, the $\sigma_s$ group displayed a hand speed that was skewed and had a higher peak value (p < 0.05; df = 12). Even in the same environment (day 3 test), the peak speed was significantly larger in $\sigma_s$ versus $\sigma_d$ (p < 0.01; df = 12).

The same tendencies were observed when subjects were trained in a stochastic CCW field (Fig. 4): overcompensation disappeared (Fig. 4a,b), the measure A1–A2 became more negative (Fig. 4c) as performance rates improved (Fig. 4d), and peak speeds increased (Fig. 4e). For example, the average within-subject correlation between overcompensation and success rates was $-0.52$ (p < 0.01). That is, people responded to the increased field variance by eliminating their overcompensation and pro-

![Figure 3. Reaching in a stochastic curl field $f = \Gamma \hat{x}$. On each trial, $\Gamma$ was drawn from a normal distribution such that $f = \Gamma \hat{x} = \Gamma \xi$, where $\xi = [0, 13; 0, 0]$, $\mu \approx 1 \text{N/m}$, and $\sigma = 0.3$. The data for the deterministic and stochastic groups are labeled as $\sigma_d$ and $\sigma_s$. a, Mean hand path (averaged for the last 50 trials of training and across subjects) for the two groups. The dashed line represents the across-group average hand path in the null condition. In the stochastic field, the overcompensation disappeared (trajectories are mean ± SE). b, Overcompensation of the hand path as measured with respect to the null trajectory. The measure refers to the maximum perpendicular displacement to the left of the null trajectory. Test, An extra set of 50 trials that is included in day 3, in which field variance was reset to zero. c, A measure of the entire shape of the hand paths in the two groups. A1–A2 areas are shown on the right. The dotted line is the trajectory in the null field. d, Success rates (arrival at target in time) for the two groups. Null, The last 50 trials in the null field; First, the first 50 trials in the force field; Day 1, Day 2, and Day 3, the last 50 trials of training on each day. e, f, Hand speed corresponding to the hand paths shown in a, normalized to the peak of hand speed in the null condition. In a zero variance field, speeds returned to near baseline, whereas in the high variance field, speeds remained elevated. All error bars are SEM.](https://example.com/figure3.png)
Reducing an increasing peak speed that skewed the profile and slowed the hand as it approached the target.

**Experiment 2: deterministic versus stochastic zero-mean fields**

The reduced overcompensation in the high-variance field was consistent with the optimal policy, but it may have been attributable to another cause: if the field is more variable, people may learn it less well (Donchin et al., 2003). Although this would not explain the increased speed in the \( \sigma_2 \) group, it can account for the reduced overcompensation. This motivated us to test the theory further. In experiment 2, the task was to reach to a target at 18 cm. Importantly, the field was now zero-mean with zero, small, or large variance. Because the mean of the field was zero in all blocks, this helped ensure that any differences that we might see regarding control policies should be attributable to the variance of the field and not a bias in learning of the mean.

Volunteers \((n = 18)\) were separated into two groups: a group that experienced a field with small variability \( \sigma_0 \) and a group that experienced a field with larger variability \( \sigma_L \). They began with a familiarization set (150 trials, no forces), followed by four sets of training in a zero-mean stochastic field, followed by three sets of a zero-mean deterministic field \( \sigma_0 \). Figure 5 shows the mean hand paths and speed during the last 50 trials of each condition. The hand paths were essentially straight in the \( \sigma_0 \) condition. With increased variance, there was a small tendency for hand trajectory to curve to the right of the null, but this tendency did not reach significance (maximum displacement from null, \( p > 0.1 \) for both groups; note that the difference in scales on the x- and y-axes). However, as the theory had predicted, peak speeds gradually increased when the field became variable (Fig. 5b,c). Furthermore, the timing of the peak speed shifted earlier in the trajectory of the hand: from 243 ms in the baseline (zero-variance) condition to 231 and 209 ms in the \( \sigma_0 \) and \( \sigma_L \) conditions, respectively (within-subject change across all subjects: main effect of condition, \( p = 0.005 \); for the \( \sigma_2 \) group: main effect of condition, \( p = 0.012 \)). A more detailed view of these changes in peak speeds is shown in Figure 5c. Increased variance produced a gradual increase in peak speed. Return to zero variance resulted in a gradual reduction in peak speed.

In summary, when we considered an environment in which the mean perturbation was zero, people responded to the unpredictability of the environment by gradually increasing their peak speeds and skewing the speed profile to reduce speed as the hand approached the target.

**Experiment 3: via-point in a stochastic zero-mean field**

The theory explained that increased uncertainty should make one more cautious as the movement approaches the target. In our previous examples, the target was always at the end of the movement. If one has a target in the middle of a movement, then increased uncertainty should make the movement through that target change as uncertainty increases. We tested this idea in experiment 3. Here, the task was to reach to an end point target (at 18 cm) by passing through a via-point (at 9 cm), as illustrated in Figure 6a. The field was zero-mean with either a zero or a non-zero variance. The experiment began with a few blocks of no forces. For another few blocks, a field was introduced that, from trial to trial, had zero mean but non-zero variance. Intuitively, we expected that the optimal trajectory in a low-uncertainty field should be a straight line with a bell-shaped velocity profile. As field uncertainty increased, the movement should slow down as it
approaches the via-point, effectively producing a segmented movement.

Figure 6a shows the predictions of the theory for various levels of field variance $\sigma_0$, $0$, $\sigma_s = 0.3$, and $\sigma_L = 0.6$. The predicted hand paths remained straight, unaffected by the different levels of variance (Fig. 6a). However, as the field variance increased, the controller separated the two movements into “segments,” showing a dip in velocity as the mass approached the via-point. To test these predictions, volunteers ($n = 22$) were divided into two groups. They began with a familiarization set (150 trials, no forces), followed by four sets of training in a zero mean but variable force field $\sigma_0$. The task was always straight. However, the speed profiles showed a segmentation of the movement with increased variance, slowing the hand as it approached the via-point. Figure 6b shows the mean hand paths and speed during the last 50 trials of each condition. The hand paths were essentially straight in the $\sigma_0$ condition with a speed profile that had a single peak. With increased variance, we did not detect a significant difference in the mean of hand paths across subjects (maximum displacement from null, $p > 0.1$), but the speed showed two peaks, suggesting of a segmentation of the reach. Indeed, the speed at via-point (0.4 s) in the $\sigma_0$ condition was significantly higher than the $\sigma_L$ and $\sigma_S$ conditions ($p < 0.01$ in each case).

Discussion

If the purpose of a movement is to acquire a rewarding state at a minimum cost (Todorov and Jordan, 2002), then the idea that the brain computes a desired movement trajectory and that trajectory remains invariant with respect to environmental dynamics is untenable. Rather, a broad implication of OFC theory is that when the environment changes, the learner performs two computations simultaneously: (1) finds a more accurate model of how motor commands produce changes in sensory states; and (2) uses that model to find a better movement plan that reoptimizes performance. Here, we performed a number of experiments to test this idea.

In experiment 1, we revisited the well studied reaching task in which velocity-dependent forces act perpendicular to the direction of motion. Thoroughman and Shadmehr (2000) reported that with adaptation, hand paths curved out beyond the baseline...
trajectory, suggesting an overcompensation in the forces that subjects produced. They interpreted those results as a characteristic of the basis functions with which the brain might be approximating the force field (Donchin et al., 2003; Wainscott et al., 2005). However, experiments that measured hand forces in channel trials found that the maximum force was, at most, 80% of the field (Scheidt et al., 2000, 2001; Smith et al., 2006). How could one produce hand paths that appeared to overcompensate for the field, yet only produce a fraction of the field forces? Here, OFC solved this puzzle. It explained that both the curved hand path and the undercompensation of the peak force were signatures of minimizing total motor costs of the reach.

We next considered an environment in which the dynamics were stochastic. Intuitively, the idea is that if we do not know the amount of coffee in a cup (or how hot it may be), we lift it and drink from it differently than if we are certain of its contents. For example, Chhabra and Jacobs (2006) found that when motor noise was artificially increased, subjects learned to alter their control policy in stabilizing a tool. Inspired by these results, we slightly extended the mathematics of OFC (see supplemental material, available at www.jneurosci.org) to make predictions about how reach plans should change when the learner is uncertain about the dynamics of the environment. If the force field is stochastic, theory predicted that overcompensation should disappear and peak speeds should increase. We observed both tendencies.

In experiment 2, we tested the theory further by considering an environment that, on average, had no perturbations but that had a variance that could be zero, small, or large. Theory predicted that as field variance increased, the trajectory should shift from a bell-shaped speed profile to one that showed a larger peak earlier in the movement, allowing more time to control the limb near the target. We observed this tendency.

Finally, in experiment 3, we considered a task in which one of the goals was to go through a via-point that was positioned along a straight line to the final target. As field variance increased, theory predicted that one should slow down as the hand approaches the via-point. That is, movements should exhibit a single peak in their speed profiles when field variance was zero, but multiple peaks as the field variance increased. Indeed, when faced with a zero variance field, movements had a single peak. With increased variance, subjects segmented their movements into two submovements.

The idea that movement planning should depend on the dynamics of the task seems intuitive. For example, Fitts (1954) observed that changing the weight of a pen affected how people planned their reaching movements: to maintain accuracy, people moved the heavier pen more slowly than the lighter pen. In reach adaptation experiments, however, movement time is often constrained by the experimenter, and so we had thought that a fully adapted trajectory would return to baseline conditions (Shadmehr and Mussa-Ivaldi, 1994). Our results here reject the notion of an invariant desired trajectory. Instead, the results are consistent with a drive to optimize movements in terms of their motor costs and accuracy (Todorov and Jordan, 2002).

In the theory, uncertainty about the magnitude of a velocity-dependent field corresponds to a velocity-dependent noise. Given this noise, one should minimize speed at the task-relevant areas: at the via-point and at the end point. In the simple reach task, when subjects were uncertain about the field, they reached with skewed speed profiles that had a higher peak and a longer tail, resulting in slower speeds near the target. In the via-point task, the increased uncertainty resulted in movements that had a segmented appearance, slowing at the via-point.

Previous models of movement planning successfully explained smoothness of reaching and eye movements using costs such as end point variance (Harris and Wolpert, 1998, 2006) or change in muscle forces or torques (Uno et al., 1989). These two models are closely related. For example, minimum torque change is equal to minimum motor command when the time constant of muscle activation is embedded in the model. The minimum variance is equal to the minimum weighted sum of motor commands when signal-dependent noise is assumed. This notion was supported by the simulations of Thorup and Jordan (2007), in which large overcompensation was predicted by both minimum torque change and minimum variance models in the force-field task. Because these models all minimize the sum of motor commands, the typical OFC problem that has motor costs is closely related to these two models. However, OFC is a more appropriate framework for simulations in motor control because it not only allows one to consider feedback (e.g., noise in sensory measurements), but also because it allows one to consider uncertainty associated with internal models that predict the sensory feedback.

Despite this, our model is a diagram. It represented the hand as a point mass, the costs and rewards as quadratic functions, and uncertainty as state-dependent noise when, in fact, the field was noisy from trial to trial, not within a trial. These are symptoms of our desire to solve mathematical problems analytically. Does the data match the predictions because of some fundamental truth in the model, or because of some unexplained quirk in the unmodeled dynamics? We approached this question in two ways. First, we tried to minimize the influence of unmodeled dynamics by considering environments that, on any given trial, had zero expected value. Second, we tested the same model in separate experiments in which environments had different means and the goals of the tasks were near (experiment 1), far (experiment 2), or both (experiment 3) in time or space. The consistency of the observations is some reassurance that reoptimization is a better model of motor control than perturbation cancellation toward a desired trajectory.

There are significant limitations in our current abilities to apply the theory to biological movements. For example, the theory faces significant hurdles when we consider that there are multiple feedback loops in the biological motor control system, namely the state-dependent response of muscles, spinal reflex pathways, as well as the long-loop pathways. In a more realistic setting, it is unclear what is meant by a motor command and a motor cost. Therefore, the best that we can currently claim is that our experimental results are difficult to explain with the notion of an invariant desired trajectory, but qualitatively in agreement with the theory.

We have implied that with training, people learn a forward model of the task and then use that model to form a better movement plan (Hwang and Shadmehr, 2005). However, it is possible to form optimum policies from the reward prediction errors on each trial without forming an explicit forward model. In our view, such an approach would be inconsistent with the large body of data from experiments in generalization (Conditt et al., 1997).

The cerebellum appears to be the key structure for computing a forward model: cerebellar agenesis produces a striking deficit in the ability to predict and compensate for consequences of one’s own motor commands (Nowak et al., 2007), cerebellar damage impairs the ability to adapt reaching (Maschke et al., 2004; Smith and Shadmehr, 2005) and throwing movements (Martin et al., 1996), and reversible disruption of cerebellar output pathways to the cortex produces within-subject impairments in reach adaptation (Chen et al., 2006). Assuming that the cerebellum is crucial
for forming a more accurate forward model, how does the brain use this model to reoptimize movements?

Because the search for a better movement plan (or control policy) is a problem that depends on costs and rewards of the task, and dopamine appears to be a crucial neurotransmitter that responds to reward prediction errors (Schultz et al., 1997) and reward uncertainty (Fiorillo et al., 2003), it is possible that the process of finding an optimal control policy depends on the basal ganglia. Recently, Mazzoni et al. (2007) demonstrated that the slowness of movements in Parkinson’s disease may be understood in terms of an imbalance in the cost function of an optimal controller. They suggested that in these patients, the motor costs relative to expected rewards had become unusually large. Basal ganglia patients typically show some ability to adapt to force fields (Krebs et al., 2001; Smith and Shadmehr, 2005). However, a strong prediction of the current theory is that they will be impaired in reoptimizing their movements.

In summary, our results support the hypothesis that that control of action proceeds via two related pathways: on the one hand, adaptation produces a more accurate estimate of the sensory consequences of the motor commands (i.e., learn an accurate forward model), and on the other hand, our brain searches for a better movement plan so to minimize an implicit motor cost and maximize rewards (i.e., find an optimum controller).

References


Here we provide details for the simulation results that were presented in the main text. We begin with the optimal control problem with the linear model of dynamics and consider the issue of model uncertainty. We analyze the robustness of the results (over-compensation, speed changes, segmentation) with respect to parameter values. Next, we consider the effect of motor noise. In the final section, we turn our attention to nonlinear models of reach dynamics and examine the robustness of some of the results in a more realistic model of the arm.

1. Optimal control with model noise

In a typical control problem, one attempts to achieve a behavioral goal based on the information that one has regarding the constraints of the task. In optimal control, behavioral goals are represented as costs, and the constraints are represented as a model of the forward dynamics of the task, i.e., a model of how motor commands produce changes in the states of the system. We wanted to allow the controller to have a degree of uncertainty about its forward model and assess how this uncertainty affected movement planning.

In general, we can think of uncertainty as a measure of variance about the mean of a parameter. If the system is linear, then model parameters multiply states of the system to predict future states, and therefore this parameter variability would produce signal-dependent state noise. That is, a noise with a standard deviation that grows linearly with the size of the state vector. To solve this kind of optimal control problem, we were guided by the approach taken by Todorov (2005) in solving a related problem, where the dynamics of the system were affected by signal dependent motor noise. Our insight was to view model parameter uncertainty as a signal-dependent state noise, that is, a dual to the signal-dependent motor noise.

Suppose that we have a linear system with \( x_t \) as its state vector (position, velocity, etc.) at time \( t \), \( u_t \) as the motor command vector composed of elements \( u_i^{(t)} \), corrupted by the noise vector \( \varphi_t \), \( \epsilon \) as a scalar, Gaussian random variable with zero mean and variance 1, \( \epsilon \sim N(0,1) \), and \( A \) and \( B \) as matrices that describe dynamics of the system:
\( x_{t+1} = Ax_t + B(\mathbf{u}_t + \varphi_t) \)

\[ \varphi_t = \begin{bmatrix} c_1 u^{(1)}_t & \epsilon^{(1)}_t \\ 0 & 0 \\ c_2 u^{(2)}_t & \epsilon^{(2)}_t \\ 0 & 0 \end{bmatrix} \]

For example, the noise that affects \( u^{(1)}_t \) is mean zero with a standard deviation that grows with a slope \( c_1 \) as a function of \( u^{(1)}_t \). Following Todorov (2005), it is convenient to rewrite this noise as follows:

\[ C_1 = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 \end{bmatrix} \quad C_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 \end{bmatrix} \]

\[ \varphi_t = \sum_i C_i \mathbf{u}_t \epsilon^{(i)}_t \]

so that the variance of the motor noise grows as a function of the motor input:

\[ \text{var}[\varphi_t] = \sum_i C_i \mathbf{u}_t \text{var}[\epsilon^{(i)}_t] \mathbf{u}_t^T C_i^T = \sum_i C_i \mathbf{u}_t \mathbf{u}_t^T C_i^T \]

This allows one to rewrite the system dynamics:

\[ x_{t+1} = Ax_t + B \mathbf{u}_t + \sum_i \epsilon^{(i)}_t \mathbf{C}_i \mathbf{u}_t \]

If we assume that the controller makes an observation \( \mathbf{y}_t \) at time \( t \), and has the goal of minimizing a cost:

**Observation:** \( \mathbf{y}_t = H \mathbf{x}_t + \omega_t \) \hspace{1cm} (0.1)

**Cost per step:** \( \mathbf{x}_t^T Q \mathbf{x}_t + \mathbf{u}_t^T R \mathbf{u}_t \) \hspace{1cm} (0.2)

where \( \omega_t \sim N(0, \Omega) \), and the matrices \( Q \) and \( R \) are symmetric positive definite matrices, then Todorov’s (2005) method provides a closed-form solution to this constrained optimization problem.

In our scenario, we have the additional problem that we are uncertain about the parameter \( A \). We express this uncertainty as:

\[ x_{t+1} = (A + \gamma_t V) x_t + B \mathbf{u}_t + \sum_i \epsilon^{(i)}_t \mathbf{C}_i \mathbf{u}_t \]

where \( \gamma_t \) is a Gaussian scalar random variable with mean 0 and standard deviation 1, and \( V \) is a scaling parameter matrix which scales the variance of the model parameter uncertainty. This effectively produces a system where the dynamics have both a signal dependent motor noise and a signal dependent sensory noise.

In our simulations, we assume that the dynamics were of the following general form:
Dynamics: $\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \mathbf{z}_t + \sum_{i=1}^{c(t)} \mathbf{C}_i \mathbf{u}_t + \sum_{i=1}^{k(t)} \gamma_i C_i \mathbf{x}_t$ \hspace{1cm} (0.3)

where $\mathbf{z}_t \sim N(0, \Omega^z)$ is an additive vector of Gaussian noise, $\mathbf{e}_t^{(i)}$ and $\gamma_i$ are independent scalar normal random variables, and $\mathbf{C}_i$ and $C_i$ are constant matrices. The initial state $\mathbf{x}_0$ had multivariate normal distribution with mean $\hat{x}_0$ and covariance $\Sigma_0$. The objective of the optimal controller was to find the optimal policy $\mathbf{u}_t$ which minimized the expected cumulative cost $E\left(\sum_{t=1}^{T} (\mathbf{x}_t^{T}\mathbf{Q}\mathbf{x}_t + \mathbf{u}_t^{T}R\mathbf{u}_t)\right)$.

We computed the optimal control policy for a pre-calculated Kalman gain $K_t$ as:

$$\mathbf{u}_t = -L_t \hat{x}_t$$

$$L_t = \left( R + B^T S_{x_{t+1}}^e B + \sum_i \mathbf{C}_i^T \left( S_{x_{t+1}}^e + S_{x_{t+1}}^c \right) \mathbf{C}_i \right)^{-1} B^T S_{x_{t+1}}^e A.$$

$$S_{x_{t+1}}^e = \mathbf{Q}_x + A^T S_{x_{t+1}}^e (A - BL_t) + \sum_i \mathbf{C}_i^T \left( S_{x_{t+1}}^e + S_{x_{t+1}}^c \right) \mathbf{C}_i; \hspace{1cm} S_{x_{t+1}}^c = \mathbf{Q}_c$$ \hspace{1cm} (0.4)

$$S_{x_{t+1}}^e = A^T S_{x_{t+1}}^e BL_t + (A - K_t H)^T S_{t_{x+1}}^c (A - K_t H); \hspace{1cm} S_{n_{t+1}}^c = 0$$

$$s_t = Tr \left( S_{x_{t+1}}^e \Omega^\xi + S_{x_{t+1}}^c \left( \Omega^\xi + \Omega^\eta + K_t \Omega^\eta K_t^T \right) \right) + s_{t+1}; \hspace{1cm} s_n = 0.$$

The matrix $L_t$ is the time-varying feedback gain, and $S_{x_{t+1}}^e$, $S_{x_{t+1}}^c$ are the parameters required to calculate the optimal cost-to-go function at any time step $t$. $Tr$ is the trace operator. The state estimate is updated by using a modified Kalman filter which takes into account the multiplicative noise. For a given feedback gain matrix $L_t$, the corresponding optimal Kalman filter is calculated in a forward pass through time:

$$\dot{\hat{x}}_{t+1} = (A - BL_t) \hat{x}_t + K_t \left( y_t - H\hat{x}_t \right) + \eta_t$$

$$K_t = A \Sigma_t^e H^T \left( H\Sigma_t^e H^T + \Omega^\eta \right)^{-1}$$

$$\Sigma_t^e = \Omega^\xi + \Omega^\eta + \sum_i \left( \mathbf{C}_i^T (\Sigma_i^e + \Sigma_i^{\xi e} + 2\Sigma_i^{\xi e}) \mathbf{C}_i + C_i L_t \Sigma_i^e L_t^T C_i^T \right) + (A - K_t H) \Sigma_t^e A^T$$

$$\Sigma_t^e = \Sigma_t$$ \hspace{1cm} (0.5)

$$\Sigma_t^{\xi e} = (A - BL_t) \Sigma_t^{\xi e} (A - BL_t)^T + K_t H \Sigma_t^e A^T + (A - BL_t) \Sigma_t^{\xi e} H^T K_t^T + K_t H \Sigma_t^{\xi e} (A - BL_t)^T + \Omega^\eta$$

$$\Sigma_t^{\xi e} = \Sigma_t^{\xi e}$$

$$\Sigma_t^{\xi e} = (A - BL_t) \Sigma_t^{\xi e} (A - K_t H)^T - \Omega^\eta$$

The matrices $K_t$, $\Sigma_t^e = E[\mathbf{e}_t \mathbf{e}_t^T]$, $\Sigma_t^\xi = E[\mathbf{\hat{e}}_t \mathbf{\hat{e}}_t^T]$ and $\Sigma_t^{\xi e} = E[\mathbf{\hat{e}}_t \mathbf{\hat{e}}_t^T]$ are the optimal Kalman gain and the covariance matrices for the random variables $\mathbf{e}_t = \mathbf{x}_t - \hat{x}_t$ and $\mathbf{\hat{e}}_t$. 


The details of the derivation are provided at the end of this document.

2. Dynamics of the linear system model of reach control

Much of the data in the main manuscript was based on a model of control for a point mass in a force field. This section provides the details of that model. The inertia in Cartesian coordinates was $M = \begin{bmatrix} 4.0 & 0 \\ 0 & 1.5 \end{bmatrix} \text{ (kg)}$. The state was defined as $\mathbf{x}(t) = [p_x(t), \dot{p}_x(t), p_y(t), \dot{p}_y(t), f_x(t), f_y(t), T_x, T_y]$, where $p_x(t), p_y(t)$ and $f_x(t), f_y(t)$ and $T_x, T_y$ are the hand position, forces produced by the arm, and target position along the x and y axes respectively. We modeled the relationship between forces $f_x(t), f_y(t)$ and the motor commands $u_x(t), u_y(t)$ as a first order linear system with a 120ms time constant.

We tested the optimal control policies predicted by the model in a viscous curl force field. This kind of field is velocity dependent and the forces produced on the hand are given by the relation $\mathbf{f} = D\mathbf{v}$, where $D = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$ and $\mathbf{v}$ is a hand velocity vector. We discretized the system dynamics with a time step of $\Delta t = 10\text{ms}$. The evolution of different components of the state for this discretized system is as follows:

$$\begin{bmatrix} p_x(t + \Delta t) \\ p_y(t + \Delta t) \end{bmatrix} = \begin{bmatrix} p_x(t) \\ p_y(t) \end{bmatrix} + \Delta t \begin{bmatrix} \dot{p}_x(t) \\ \dot{p}_y(t) \end{bmatrix}$$

$$\begin{bmatrix} \dot{p}_x(t + \Delta t) \\ \dot{p}_y(t + \Delta t) \end{bmatrix} = \begin{bmatrix} \dot{p}_x(t) \\ \dot{p}_y(t) \end{bmatrix} + \Delta t \begin{bmatrix} \frac{d_{11}}{m_1} & \frac{d_{12}}{m_2} \\ \frac{d_{21}}{m_1} & \frac{d_{22}}{m_2} \end{bmatrix} \begin{bmatrix} \dot{p}_x(t) \\ \dot{p}_y(t) \end{bmatrix} + \Delta t \left[ \begin{bmatrix} f_x(t) \\ f_y(t) \end{bmatrix} + \begin{bmatrix} \frac{d_{11}}{m_1} & \frac{d_{12}}{m_2} \\ \frac{d_{21}}{m_1} & \frac{d_{22}}{m_2} \end{bmatrix} \end{bmatrix} \right] + \Delta t \cdot \gamma$$

$$\begin{bmatrix} f_x(t + \Delta t) \\ f_y(t + \Delta t) \end{bmatrix} = \begin{bmatrix} 1 - \frac{\Delta t}{\tau} & 0 \\ 0 & 1 - \frac{\Delta t}{\tau} \end{bmatrix} \begin{bmatrix} f_x(t) \\ f_y(t) \end{bmatrix} + \Delta t \begin{bmatrix} u_x(t) \\ u_y(t) \end{bmatrix}$$

where $\gamma$ is a mean zero variance $\sigma^2$ Gaussian noise. The discretized system dynamics can be transformed into the form of Eqn. (0.3) using the matrices:
The sensory feedback matrix was formulated so that the controller was able to observe hand positions, velocities and the target positions over the course of the movement. Hence, 

\[
A = \begin{bmatrix}
A_1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

\[
A_T = \begin{bmatrix}
1 & \Delta t & 0 & 0 & 0 & 0 \\
0 & 1 + \Delta t \frac{d_{11}}{m_1} & \Delta t \frac{d_{12}}{m_1} & \Delta t \frac{1}{m_1} & 0 \\
0 & 0 & 1 & \Delta t & 0 & 0 \\
0 & \Delta t \frac{d_{11}}{m_2} & 0 & 1 + \Delta t \frac{d_{22}}{m_2} & 0 & \Delta t \frac{1}{m_2} \\
0 & 0 & 0 & 0 & 1 - \frac{\Delta t}{\tau} & 0 \\
0 & 0 & 0 & 0 & 1 - \frac{\Delta t}{\tau} & 0 \\
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
\frac{\Delta t}{\tau} & 0 \\
0 & \frac{\Delta t}{\tau} \\
0 & 0 \\
\end{bmatrix}
\]

\[
C_i = \begin{bmatrix}
C_1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

\[
C_{di} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & \sigma \sqrt{\Delta t} \frac{d_{11}}{m_1} & 0 & \sigma \sqrt{\Delta t} \frac{d_{12}}{m_1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & \sigma \sqrt{\Delta t} \frac{d_{21}}{m_2} & 0 & \sigma \sqrt{\Delta t} \frac{d_{22}}{m_2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

\[
\bar{C} = [0_{n \times 2}]
\]

\[
\bar{C}^T = [0_{n \times 2}]
\]

The control (or motor) cost penalty matrix \( R \) was set to a constant value throughout the course of the movement. The parameter \( w_r \) determines the weight of the control cost. We set \( R = w_r I_{2 \times 2} \) for \( 0 \leq t < T + T_H \), where \( T \) is the maximum movement completion time and \( T_H \) is the time for which the arm was supposed to hold position at the target after movement completion.

The 'state cost' penalty matrix was formulated so that the state cost was zero before the movement completion time \( T \) and was increased in a step to a high value for the time period from \( T \) to \( T_H \). This formulation provided the controller with the maximum flexibility to search for the optimal policy since the controller was penalized only for not being at the target after the movement completion time \( T \) without
imposing any constraints on the trajectory followed by the controller to reach the target. 

\[ Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -w_p & 0 \\ 0 & w_r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_p & 0 & 0 & 0 & -w_p \\ 0 & 0 & 0 & w_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_f & 0 & 0 \\ -w_p & 0 & 0 & 0 & 0 & w_p & 0 \\ 0 & 0 & -w_p & 0 & 0 & 0 & w_p \end{bmatrix} \]

for \( T \leq t < T + T_H \)

\[ \Omega^z = 0 \\
\Omega^w = 0 . \]

For simplicity, the variance of the additive Gaussian noises was set to zero.

The cost parameters and the time constraints used for the simulations in the manuscript were:

\[ w_r = 10^{-8}, \ w_c = 0.01, \ w_f = 0, \ w_p = 5, \ T = \frac{0.45}{\Delta t}, \ T_H = \frac{0.05}{\Delta t} . \]  

Sensitivity of the simulations to these parameters are discussed below.

3. Effect of bias of the forward model on the optimal control policy: sensitivity analysis

This section describes how we simulated the effects of incomplete adaptation of the internal model on the average behavior of the controller for the point mass system. This section also considers the robustness of the results by varying components of the cost function.

The estimate of the force field parameter available to the controller through the internal model was \( \hat{D} = \alpha \tilde{D} \). Scaling parameter \( \alpha \) denotes accuracy of internal model corresponding to force field. The simulation results shown in Fig. S1 correspond to a clockwise viscous curl force field where

\[ \tilde{D} = \begin{bmatrix} 0 & 13 \\ -13 & 0 \end{bmatrix} . \]  

The simulations were done for three different levels of accuracy of the internal model \( \alpha = 1, \ 0.8 \) and \( 0.6 \). We also explored the effects of various levels of the position cost parameter \( w_p \) on the simulation results. We found that a completely adapted internal model \( \alpha = 1 \) led to over-compensation. However, as the adaptation level decreased \( \alpha = 0.8 \) and \( 0.6 \), the over-compensation at the early part of the movement became smaller as the under-compensation near the
target increased.

As the position cost parameter $w_p$ decreased, the trajectory did not complete the reaching to target position. However, the over-compensation was a characteristics of all tested values of $w_p$.

4. Effect of uncertainty of the internal model: sensitivity analysis

In this simulation, the internal model was defined as $\hat{D} = \alpha \bar{D} + \gamma_t \bar{D}$, where $\gamma_t$ is a zero mean Gaussian random variable with variance $\sigma^2$. Figure S2A shows how the variance of this noise and position cost $w_p$ interact. In the simulations, we had $\alpha = 0.8$. The results plotted by blue, red and green lines denote different amounts of uncertainty $\sigma = 0.1$, $\sigma = 0.2$ and $\sigma = 0.3$ respectively. When position costs were small ($w_p = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$), the mass did not reach the target and peak velocity was small for all $\sigma$. When this cost was larger ($w_p = 10^0, 10^1, 10^2, 10^3$), the mass reached the target and there were clear differences in peak velocity between $\sigma = 0$, 0.2 and 0.3 such that higher noise caused higher peak velocity. In all cases, the peak velocity was higher with higher noise.

Figure S2B shows the effect of noise for different level of $w_r$ (control cost). When this weight was too large ($w_r = 10^{-6}, 10^{-5}$), the mass did not reach the target. When this weight was smaller ($w_r = 10^{-12}, 10^{-11}, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}$), the mass reached the target and there were clear differences in peak velocity between $\sigma = 0.1$, 0.2 and 0.3 so that higher noise caused higher peak velocity. In general, the prediction of increased peak speed with increased uncertainty was a robust result of the simulations.

5. Effects of model uncertainty in the via point task: sensitivity analysis

The task was to arrive at the target before certain maximum time $T$ and stay there for a hold time of $T_H$ while passing through the via-point (a location along a straight line between the starting position and the target) at a specific via-point time $T_V$. These simulations were performed for an unbiased viscous curl force field, i.e. the components of the matrix $D$ are all zero. Hence, any differences in the optimal control policy are purely due to different levels of uncertainties without being confounded with incomplete learning of the force field parameter $D$. The state for the via-point task was modified to include the location of the via-point. All other components of the state were the same.
\[ x(t) = \left[ p_x(t), \dot{p}_x(t), p_y(t), \dot{p}_y(t), f_x(t), f_y(t), TV_x, TV_y, T_x, T_y \right] \]

where \( TV_x, TV_y \) are x and y co-ordinates of the via point location. The system dynamics matrices were modified to incorporate the via-point location in the state vector. We made small modifications to the matrices so that the via-point location remained constant over the course of the movement as shown below. All other components of these matrices remain the same as in the simple reaching task.

\[
A = \begin{bmatrix} A_d & 0 \\ 0 & I_{4 \times 4} \end{bmatrix}, \quad \quad B = \begin{bmatrix} \frac{\Delta t}{\tau} & 0 \\ 0 & \frac{\Delta t}{\tau} \\ 0 & \frac{\Delta t}{\tau} \end{bmatrix}, \quad \quad C_1 = \begin{bmatrix} C_{d1} & 0 \\ 0 & I_{4 \times 4} \end{bmatrix}, \quad \quad \tilde{C} = [0_{10 \times 2}] 
\]

where \( A_d \) and \( C_{d1} \) were the same as Eqn 0.7.

The sensory feedback matrix \( H \) was changed so that the observation \( y \) included the via-point location.

\[
H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \end{bmatrix} 
\]

The control cost penalty matrix \( R \) was the same as in previous section. However, we had to modify the ‘state cost’ penalty matrix to penalize the controller for not being at the via-point at the via-point time. The variance of the additive Gaussian noises \( (\Omega^x, \Omega^y) \) was set to zero similar to the simulations for the reaching task.

The cost parameters and the time constraints used for the via-point simulations in the manuscript were

\[
w_r = 10^{-8}, \quad w_v = 0.01, \quad w_p = 3, \quad w_{pv} = 1, \quad T = \frac{1.0}{\Delta t}, \quad T_r = \frac{0.4}{\Delta t}, \quad T_H = \frac{0.2}{\Delta t}. \]

To account for the unbiased variable force field, we assumed an internal model which had the model of force field with parameters,

\[
\hat{D} = \gamma_t D, \quad \text{where} \quad D = \begin{bmatrix} 0 & 13 \\ -13 & 0 \end{bmatrix}, \quad \gamma_t \quad \text{is a Gaussian random variable with mean 0 and variance} \quad \sigma. \quad \text{We}
\]
tested $\sigma=0$, 0.2, or 0.3.

Figure S3A shows the effect of noise for different values of $w_p$. The hand path was a straight line for all amplitudes of noise and all costs in all simulations of the via-point task (data not shown). However, with increased uncertainty (noise variance), the speed decreased at the time when hand passed the via-point. This characteristic was consistent for all $w_p$. As shown in Fig. S3B, the characteristics of the decreasing speed at the via point changed drastically by the change of weight of via point $w_{pv}$. When $w_{pv}$ was very small ($w_{pv}=10^{-3}, 10^{-2}, 10^{-1}$), there was effectively no strong cost associated with going through the via-point, and the speed profile did not show segmentation. However, as the constraint for going through the via-point increased in importance, the speed profile became bimodal. Fig. S3C shows trajectories for different control costs $w_r$. When $w_r$ was large ($w_r=10^{-6}, 10^{-5}, 10^{-4}$), the mass did not reach the target, indicating that optimal controller focused more on minimizing motor cost than the cost for reach. Thus, the change of speed at via point was not significant for these conditions. Instead, when $w_r$ was small ($w_r=10^{-11}, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}$), there was a significant change of speed at the via-point, producing segmentation. We concluded that in the via point task, increased uncertainty consistently encouraged segmentation.

6. Effect of motor noise

Noise in the motor commands is an important component of the computational models of the motor system (Harris and Wolpert, 1998; Todorov and Jordan, 2002). We excluded motor command dependent noise in the simulations shown in the manuscript to highlight the effects of model parameter uncertainty on the optimal control policy. In this section we address the issue of the effect of different levels of motor command dependent noise on the model predictions. The parameter $c$ lets us control the variance of the signal dependent motor noise in the system.

$$C = \begin{bmatrix}
0 & 0 & 0 & 0 & \sqrt{\Delta t}c & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sqrt{\Delta t}c & 0 & 0
\end{bmatrix}^T$$

for the point-to-point reach task and
\[ C = \begin{bmatrix} 0 & 0 & 0 & 0 & \sqrt{\Delta t} c & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{\Delta t} c & 0 & 0 & 0 \end{bmatrix}^T \text{ for the via-point task.} \]

Figure S4A shows the simulation results for different amounts of motor command noise (i.e. \( c \)). All other parameters were kept constant at the values used for the simulations shown in the manuscript. All simulations were repeated for three different levels of model parameter uncertainty (\( \sigma = 0, 0.2 \) and 0.3, corresponding to the blue, red, and green lines in this figure). The mass failed to reach the end point and the effects of model parameter uncertainty on the speed profile were unclear for high values of motor noise (\( c > 0.08 \)). On the other hand, the characteristics of trajectory with respect to the model parameter uncertainty were consistent with the results shown in the manuscript for low values of motor noise (\( c < 0.08 \)): an increase in model parameter uncertainty made over compensation decrease and peak velocity increase. Large amounts of motor noise caused slowness of movement, indicating the effect of motor noise is similar to effect of motor cost.

Figure S4B shows the results for different levels of motor noise in the via-point task. The results show a segmented movement as long as the motor noise is low to moderate (\( c < 0.08 \)). With very large motor noise, the segmentation disappears. However, in all cases the increase uncertainty causes the speed profiles to become skewed with a sharper rise as the movement initiates.

In the representative paper concerning motor noise (Todorov and Jordan, 2002), the scaling parameter was set as 0.04. In experiments on arm muscles, the scaling parameter of signal dependent noise is thought to be around 0.05 (Hamilton et al., 2004). Our simulations also found clear effects of model uncertainty when motor noise amplitude was around 0.06 or smaller.

7. **Solving the optimal control problem with model uncertainty**

In this section, we drive the solution to the optimal control problem with model uncertainty.

Consider a linear dynamical system with state \( \mathbf{x}_t \in \mathbb{R}^n \), control signal \( \mathbf{u}_t \in \mathbb{R}^p \), feedback \( \mathbf{y}_t \in \mathbb{R}^q \), in discrete time \( t \):

- **Dynamics**
  \[ \mathbf{x}_{t+1} = A \mathbf{x}_t + B \mathbf{u}_t + \xi_t + \sum_{i=1}^c \gamma_i^t \mathbf{C}_i \mathbf{x}_t + \sum_{i=1}^c \epsilon^t \mathbf{C}_i \mathbf{u}_t \]  
  (0.8)

- **Feedback**
  \[ \mathbf{y}_t = H \mathbf{x}_t + \omega_t \]  
  (0.9)

- **Cost per step**
  \[ \mathbf{x}_t^T Q_x \mathbf{x}_t + \mathbf{u}_t^T R \mathbf{u}_t \]  
  (0.10)

The state estimate of the dynamic system available to the controller is assumed to be updated according to a linear recursive filter for analytical tractability.

\[ \tilde{\mathbf{x}}_{t+1} = A \tilde{\mathbf{x}}_t + B \mathbf{u}_t + K_t (\mathbf{y}_t - H \tilde{\mathbf{x}}_t) + \eta_t \]

We define the estimation error as \( \mathbf{e}_t = \mathbf{x}_t - \tilde{\mathbf{x}}_t \). We can show through induction that the optimal
cost-to-go function or the cost expected to accumulate under the optimal control law after a time step $t$ has the quadratic form.

$$v_t(x_t, \hat{x}_t) = x_t^T S^e_t x_t + (x_t - \hat{x}_t)^T S^e_t (x_t - \hat{x}_t) + s_t = x_t^T S^e_t x_t + e_t^T S^e_t e_t + s_t \quad (0.11)$$

At the final time $t = n$, the optimal cost-to-go function is simply the final state cost $x_n^T Q_n x_n$, and so $v_n$ is in the assumed form with $S^e_n = Q_n$, $S^e_n = 0$, $s_n$.

Consider the optimal control policy denoted by $u_t = \pi(\hat{x}_t)$. Let $v_{t+1}^\pi(x_{t+1}, \hat{x}_{t+1})$ be the cost-to-go function corresponding to the optimal control law. Since this control law is optimal for all time points $t+1, ..., n$, we have $v_{t+1}^\pi = v_{t+1}$, so that the cost-to-go function $v_t^\pi$ satisfies the Bellman equation:

$$v_t^\pi(x_t, \hat{x}_t) = x_t^T Q_n x_t + \pi(\hat{x}_t)^T R \pi(\hat{x}_t) + E[v_{t+1}^\pi(x_{t+1}, \hat{x}_{t+1}) | x_t, \hat{x}_t, \pi] \quad (0.12)$$

Now, the stochastic dynamics of the variables of interest can be written as

$$x_{t+1} = Ax_t + Bu_t + \xi_t + \sum_{i=1}^c \gamma_i C_i x_t + \sum_{i=1}^c \xi_i C_i u_t$$

$$e_{t+1} = (A - K_t H_t) e_t + \xi_t - K_t \omega_t - \eta_t + \sum_{i=1}^c (\gamma_i C_i x_t + \xi_i C_i u_t) \quad (0.13)$$

Then the conditional means and co variances of these random variables of interest are

$$E[x_{t+1} | x_t, \hat{x}_t, \pi] = Ax_t + B \pi(\hat{x}_t)$$

$$E[e_{t+1} | x_t, \hat{x}_t, \pi] = (A - K_t H_t) e_t$$

$$\text{Cov}[x_{t+1} | x_t, \hat{x}_t, \pi] = \Omega^x + \sum_{i=1}^c (C_i x_t X_i T + C_i \pi(\hat{x}_t)) \pi(\hat{x}_t)^T C_i^T$$

$$\text{Cov}[e_{t+1} | x_t, \hat{x}_t, \pi] = \Omega^x + \sum_{i=1}^c (C_i x_t X_i T + C_i \pi(\hat{x}_t)) \pi(\hat{x}_t)^T C_i^T + \Omega^e + K_t \Omega^e K_t^T \quad (0.15)$$

Using the expected values and the covariances we just found and the relations from Eqn 0.11 and 0.12, we get an expression for the cost-to-go function

$$v_t^\pi(x_t, \hat{x}_t) = x_t^T \left( Q_t + A T S_{t+1}^T A + \sum_{i=1}^c C_i T (S^e_{t+1} + S^e_{t+1}) C_i \right) x_t + s_{t+1} + e_t^T \left( A - K_t H_t \right)^T S^e_{t+1} (A - K_t H_t) e_t + \text{Tr}(M_i) + \pi(\hat{x}_t)^T \left( R + B T S_{t+1}^T B + \sum_{i=1}^c C_i T (S^e_{t+1} + S^e_{t+1}) C_i \right) \pi(\hat{x}_t) + 2 \pi(\hat{x}_t)^T B T S_{t+1}^T A x_t$$

Writing the expression in a compact form, we have
\[ v^*(x_t, \hat{x}_t) = x^T_t \left( Q_t + A^T S_{t+1}^x A + C_t \right) x_t + e^T_t \left( A - K_t H \right)^T S_{t+1}^e \left( A - K_t H \right) e_t \]
\[ + s_{t+1} + \text{Tr} \left( M_t \right) + \pi \left( \hat{x}_t \right)^T \left( R + B^T S_{t+1}^x B + \bar{C}_t \right) \pi \left( \hat{x}_t \right) \]
\[ + 2\pi \left( \hat{x}_t \right)^T B^T S_{t+1}^e A x_t \]  
(0.17)

Where

\[ C_t = \sum_i C_{i}^T (S_{t+1}^x + S_{t+1}^e) C_i \]
\[ \bar{C}_t = \sum_i \bar{C}_{i}^T (S_{t+1}^x + S_{t+1}^e) \bar{C}_i \]
\[ M_t = S_{t+1}^x \Omega^x + S_{t+1}^e \left( \Omega^x + \Omega^y + K_t \Omega^s K_t^T \right) \]
(0.18)

The cost-to-go function, however, is a function of the true state \( x_t \), which is not available to the controller. The only thing available to the controller is the state estimate \( \hat{x}_t \). So, we take the expected value of the cost-to-go function over the true state and minimize it with respect to the control policy \( \pi \). Since \( E[x_t \mid \hat{x}_t] = \hat{x}_t \), we have

\[ E[v^*(x_t, \hat{x}_t) \mid \hat{x}_t] = \text{const} + \pi(\hat{x}_t)^T (R + B^T S_{t+1}^x B + \bar{C}_t) \pi(\hat{x}_t) + 2\pi(\hat{x}_t)^T B^T S_{t+1}^e A x_t \]
(0.19)

Thus optimal control law at time point \( t \) is

\[ u_t = \pi(\hat{x}_t) = -L_t \hat{x}_t; \quad L_t = (R + B^T S_{t+1}^x B + \bar{C}_t)^{-1} B^T S_{t+1}^e A. \]
(0.20)

Substituting \( u_t - L_t \hat{x}_t \) in Eqn 0.16

\[ v^x(x_t, \hat{x}_t) = x^T_t \left( Q_t + A^T S_{t+1}^x (A - BL_t) + C_t \right) x_t + \text{Tr} \left( M_t \right) + s_{t+1} \]
\[ + e^T_t \left( A^T S_{t+1}^e BL_t + (A - K_t H)^T S_{t+1}^e (A - K_t H) \right) e_t \]  
(0.21)

Comparing this equation with Eqn 0.11, we can summarize the optimal control law with the following equations:

\[ u_t = -L_t \hat{x}_t \]
\[ L_t = \left( R + B^T S_{t+1}^x B + \bar{C}_t \right)^{-1} B^T S_{t+1}^e A. \]
\[ S_t^x = Q_t + A^T S_{t+1}^x (A - BL_t) + C_t; \quad S_t^e = \Omega^x \]
\[ S_{t+1}^e = A^T S_{t+1}^x BL_t + (A - K_t H)^T S_{t+1}^e (A - K_t H); \quad S_{t+1}^e = 0 \]
\[ s_t = \text{Tr} \left( M_t \right) + s_{t+1}; \quad s_n = 0. \]
(0.22)
Thus, we showed that the cost-to-go function remains in the assumed quadratic form shown in Eqn 0.11 for any time step t given that it is true for the time step t+1, completing the induction proof. The next step is to calculate the optimal Kalman gains for the control policy we just calculated. According to the assumption in the previous section of the Kalman gains not being functions of \( x_t \) and \( \hat{x}_t \), we need to minimize the unconditional expectation of the cost-to-go function \( v_{t+1} \) with respect to \( K_t \) to calculate the optimal Kalman gains.

The terms in \( E[v_{t+1}(x_{t+1}, \hat{x}_{t+1}, L_t)] \) that depend on \( K_t \) are

\[
e_t^T \left( A - K_t H \right)^T S_{t+1}^e \left( A - K_t H \right) e_t + Tr\left(S_{t+1}^e K_t \Omega^o K_t^T\right) (0.23)
\]

Defining the unconditional covariances \( \Sigma_i^e = E[e_i e_i^T] \), \( \Sigma_i^x = E[x_i x_i^T] \) and \( \Sigma_i^{xe} = E[x_i e_i^T] \), the unconditional expectation of the \( K_t \)-dependent expression above becomes

\[
a(K_t) = Tr\left(\left( A - K_t H \right) \Sigma_i^e \left( A - K_t H \right)^T + K_t \Omega^o K_t^T \right) S_{t+1}^e \quad (0.24)
\]

Optimal \( K_t \) must satisfy

\[
\frac{\partial a(K_t)}{\partial K_t} = 0.
\]

\[
\frac{\partial a(K_t)}{\partial K_t} = 2S_{t+1}^e \left(K_t \left( H \Sigma_i^e H^T + \Omega^o \right) - A \Sigma_i^e H^T \right) \quad (0.25)
\]

Setting the derivative to zero and solving for \( K_t \).

\[
K_t = A \Sigma_i^e H^T (H \Sigma_i^e H^T + \Omega^o)^{-1} \quad (0.26)
\]

We found an expression to calculate the optimal Kalman gains. But, we still need to find the unconditional covariances \( \Sigma_i^e, \Sigma_i^{\hat{x}} \) and \( \Sigma_i^{xe} \). Since the variables \( x, \hat{x} \) and \( e \) are deterministically related, we can calculate the covariance of the third variable given that we know the covariance for two variables. Given the covariance of \( \hat{x} \) and \( e \), the covariance of \( x \) is given by

\[
\Sigma_{i+1}^x = E\left[(e + \hat{x})(e + \hat{x})^T\right] = \Sigma_i^e + \Sigma_i^x + \Sigma_i^{xe} + \Sigma_i^{xT}
\]

Now, using the expressions for \( \hat{x}_{t+1} \) and \( e_{t+1} \), we can calculate the expressions for the unconditional covariances shown below

\[
\Sigma_{i+1}^e = (A - K_t H) \Sigma_i^e (A - K_t H)^T + \Omega^{\hat{x}} + \Omega^o + K_t \Omega^o K_t^T + \sum_i \left(C_i \left( \Sigma_i^e + \Sigma_i^x + \Sigma_i^{xe} + \Sigma_i^{xT}\right) C_i^T + \overline{C}_i L_i \Sigma_i^x L_i^T \overline{C}_i^T\right)
\]

\[
\Sigma_{i+1}^x = (A - BL_i) \Sigma_i^x (A - BL_i)^T + K_t \left( H \Sigma_i^e H^T + \Omega^o\right) K_t^T + \left(A - BL_i\right) \Sigma_i^{xe} K_t^T + K_t H \Sigma_i^{xe} (A - BL_i)^T + \Omega^o
\]

\[
+ (A - BL_i) \Sigma_i^{xe} H^T K_t + K_t H \Sigma_i^{xe} (A - BL_i)^T + \Omega^o
\]
\[
\Sigma_{t+1}^{x_e} = (A - BL_t)\Sigma_t^{x_e}(A - K_tH)^T + K_tH\Sigma_t^r(A - K_tH)^T - K_t\Omega^{e}\Sigma_t^{eT} - \Omega^g
\]  

(0.27)

Substituting the expression for \( K_t \) from Eqn 0.26 in the above expressions for covariances, we can simplify them further to get the following system of equations which lets us calculate the optimal Kalman gains in a forward pass through time

\[
\begin{align*}
\dot{x}_{t+1} &= (A - BL_t)\dot{x}_t + K_t(y_t - H\dot{x}_t) + \eta_t \\
K_t &= A\Sigma_t^eH^T(H\Sigma_t^eH^T + \Omega^o) \\
\Sigma_{t+1}^e &= \Omega^e + \Omega^g + \sum_i \left( C_i\left( \Sigma_t^i + \Sigma_t^{i*} + 2\Sigma_t^{xe} \right) C_i^T + \overline{C}_iL_i\Sigma_i^L L_i^T \overline{C}_i^T \right) + (A - K_tH)\Sigma_t^eA^T; \Sigma_t^i = \Sigma_1 \\
\Sigma_{t+1}^{xe} &= (A - BL_t)\Sigma_t^{xe}(A - BL_t)^T + K_tH\Sigma_t^eA^T + (A - BL_t)\Sigma_t^{xe}H^T K_t^T + K_tH\Sigma_t^{xe}(A - BL_t)^T + \Omega^g; \Sigma_t^{i*} = \hat{x}_t\hat{x}_t^T \\
\Sigma_{t+1}^s &= (A - BL_t)\Sigma_t^s(A - K_tH)^T - \Omega^g
\end{align*}
\]  

(0.28)

8. **Optimal control of non-linear dynamics of reaching**

The solution of optimal feedback controller was derived analytically for Linear Quadratic Gaussian (LQG) system with ‘motor’ and/or ‘state’ signal dependent noise, which guarantees global optimality. An important prediction of this model was the curved trajectories, i.e. the over-compensation. To what extent are these results affected if the system was a more realistic model of the arm? Here we will show that over-compensation is a fundamental property of the system.

The dynamic equation of two link arm is

\[
\ddot{\theta} = M(\theta)^{-1}(\tau_u - \tau_e - a_2 \sin \theta_2 \begin{bmatrix} 2\dot{\theta}_2 \dot{\theta}_2 \dot{\theta}_2 \\ -\dot{\theta}_1 \dot{\theta}_1 \dot{\theta}_1 \end{bmatrix}) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix}
\]

where \([\theta_1, \theta_2]\) are shoulder and elbow joint angle, \(M = \begin{bmatrix} a_1 + 2a_2 \cos \theta_2 & a_3 + a_2 \cos \theta_2 \\ a_3 + a_2 \cos \theta_2 & a_3 \end{bmatrix}\),

\(a_1 = I_1 + I_2 + m_2l_2^2, a_2 = m_2l_2s_2, a_3 = I_2, m_i \) are the link masses (1.4kg, 1kg), \(l_i \) are the link lengths (30cm, 33cm), \(s_i \) are the distances from the joint center to the center of the mass (11cm, 16cm), and \(I_i \) are the moment of inertia (0.025kgm\(^2\), 0.045kgm\(^2\)). \(b_e \) is a parameter of passive joint viscosity and we used \( \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix} \) [Nms/rad]. \(\tau_e\) is a torque due to the external force exerted at the
hand produced by the velocity dependent force field: \( \mathbf{f}_D = D \mathbf{v} \), \( D = \begin{bmatrix} 0 & 13 \\ -13 & 0 \end{bmatrix} \), where \( \mathbf{f}_D, \mathbf{v} \) are the hand force and velocity of the hand. The virtual torque produced by the force field is \( \tau_D = J(\theta)^T D J(\theta) \), where \( J(\theta) = \begin{bmatrix} -l_1 \sin(\theta_1) - l_2 \sin(\theta_1 + \theta_2) & -l_2 \sin(\theta_1 + \theta_2) \\ l_1 \cos(\theta_1) + l_2 \cos(\theta_1 + \theta_2) & l_2 \cos(\theta_1 + \theta_2) \end{bmatrix} \) is the Jacobian matrix of kinematics \( dx = Jd\theta \).

The cost was \( w_p^\prime (\tau_1^2 + \tau_2^2) \quad (0 \leq t < T) \), and
\[
 w_p'(\theta_1 - \theta_1^*)^2 + w_p'(\theta_2 - \theta_2^*)^2 + w_v'(\dot{\theta}_1 + \dot{\theta}_2) \quad (T \leq t < T + T_H) ,
\]
where \( \theta^* \) is target position in joint coordinate system, reach duration \( T \) was 490 msec and simulation time \( T + T_H \) was 500 msec. The weights of costs function were \( w_p^\prime =10, \ w_v^\prime =1 \) and \( w_u^\prime =0.1^k \) respectively. The task was reaching out to one of eight targets. The start position was at \([0, 0.45m]\) in the Cartesian coordinate system (where \(0, 0\) is at the shoulder joint). The target was at 10 cm. The mathematics that we used to solve this problem is the same as that developed by (LI and Todorov, 2006), in which the nonlinear equations of forward dynamics are locally linearized and then solved as a linear optimal feedback control problem.

Figure S5 shows hand paths of optimal trajectory for Null and Force Field conditions (the field was a clockwise curl field). The trajectories were straight when no force was exerted at hand. As we expected, the trajectory curved so that trajectory could over-compensate the applied force, when the force field were exerted at hand. These predictions are very similar to the predictions of point-mass system regarding curvature of the over-compensation. However, the maximum of over-compensation of the simulation of two link arm was about 1cm which was slightly larger than that of observed data (<0.5cm). Thoroughman et al. (2007) had made a similar observation when performing these simulations under minimization of torque change or endpoint variance constraints. There are two possible reasons why amplitude of over-compensation was larger than measured data. First, it is very likely that subjects did not build perfect internal models. As we saw in the simulation of point mass system, the inaccuracy suppressed over-compensation. Independent measures suggest that during force field learning, adaptation is no better than 80% of maximum required force (Smith et al. 2006). Finally, it is possible that subjects had some extent of uncertainty for the force field environment due to noise inherently in the portion of state prediction in the brain. Increased uncertainty also tends to reduce the over-compensation.
References


**Figure S1.** The effect of model bias and endpoint position costs on reach trajectory. Upper row: hand path. Bottom row: speed profile. Each column represents trajectory for a different value of weight of position cost $w_p$. Once endpoint costs are large enough so that the mass reaches the target, the effect of model bias is to reduce the over-compensation. There is no effect of model bias on reach speed.
Figure S2. The effects of model uncertainty and costs on the reach to target task. A. Effects of endpoint costs $w_p$ and model uncertainty. Upper row: hand path. Bottom row: speed profile. Model bias $\alpha$ is 0.8 for all simulations. B. Effects of motor cost $w_r$ and model uncertainty.
Figure S3. The effects of model uncertainty and costs on the via-point task. 

A. Speed profiles for different values of weight of position cost $w_p$. 

B. Speed profiles for different values of weight of via-point cost $w_{pv}$. 

C. Speed profiles for different values of weight of motor cost $w_r$. 
Figure S4. Effects of including signal dependent motor noise $c$ for the reach to target and reach through the via-point simulations. Each line represents result of different amounts of model uncertainty. A. Reach to target simulations. B. Reach through a via-point simulations.
**Figure S5.** Hand paths of simulation of optimal control for a nonlinear model of arm dynamics. Left: Environment has zero forces (Null). Right: Environment has a clockwise curl force field. The small circles are target positions.