Behavioral/Systems/Cognitive

Motor Adaptation as a Process of Reoptimization

Jun Izawa,[1] **Tushar Rane,**[1] **Opher Donchin,**[2] **and Reza Shadmehr**[1]

[1]Laboratory for Computational Motor Control, Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, and [2]Motor Learning Laboratory, Department of Biomedical Engineering, Ben Gurion University of the Negev, Be'er Sheva 84105, Israel

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 × 5 mm square) and were rewarded (via a target explosion) if they completed their 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 $\mathbf{f} = D\dot{\mathbf{x}}$, where $\dot{x}$ is hand velocity in Cartesian coordinates. On each trial, D was drawn from a normal distribution such that $\mathbf{f} = \bar{D}(1 + \delta)\dot{\mathbf{x}} = \bar{D}\dot{\mathbf{x}} + \bar{D}\delta\dot{\mathbf{x}}$, where $\bar{D}$ is the mean of the distribution (see below) and δ is a normally distributed scalar random variable with zero mean and variance σ^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 position 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 null block with no force perturbations. This was followed by seven field training blocks. The force field was defined as $\mathbf{f} = D\delta\dot{\mathbf{x}}$, where $D = [0, 13; -13, 0]$ Nm/s and δ are a normally distributed random variable with zero mean and variance σ^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 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 $t = 400 \pm$

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 null block with no force perturbations. This was followed by seven field training blocks. The force field was defined as $\mathbf{f} = D\delta\dot{\mathbf{x}}$, where $\bar{D} = [0, 13; -13, 0]$ Nm/s and δ are a normally distributed random variable with zero mean and variance σ^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: $\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t$. Here, $\mathbf{x}_t$ is the state of the system at time t, $\mathbf{u}_t$ 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:

$$\mathbf{x}_{t+1} = (A + V)\mathbf{x}_t + B\mathbf{u}_t = A\mathbf{x}_t + B\mathbf{u}_t + V\mathbf{x}_t,$$

where V is a Gaussian random variable with mean zero and variance Q_v. 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:

$$\text{Dynamics: } \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \xi_t + \sum_{i=1}^{c} \varepsilon_t^i C_t \mathbf{x}_t \qquad (1)$$

$$\text{Observation: } \mathbf{y}_t = H\mathbf{x}_t + \omega_t \qquad (2)$$

$$\text{Cost: } 0 \leq \mathbf{x}_t^T Q_t \mathbf{x}_t + \mathbf{u}_t^T R\mathbf{u}_t \qquad (3)$$

where y_t is the observation made by the system, H is the observation matrix, and ξ_t, ε_t^i, and ω_t 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 ε_t^i. Q_t is the weight matrix of state cost, and R is the weight matrix of motor cost. $\mathbf{x}_t$ is the actual state of the system that is not available to the controller. The controller only has an estimate of the state $\hat{\mathbf{x}}_t$ available through the state estimation process. For analytical tractability, the state is assumed to be updated according to a linear recursive filter: $\hat{\mathbf{x}}_{t+1} = A\hat{\mathbf{x}}_t + B\mathbf{u}_t + K_t (y_t - H\hat{\mathbf{x}}_t)$, where K_t is the Kalman gain. The optimal control policy is of the following form:

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

where L_t 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 $\mathbf{x}(t) = [p_x(t), \dot{p}_x(t), p_y(t), \dot{p}_y(t), f_x(t), f_y(t), T_x, T_y]$, where p is hand position, f is force, and T is target position. The cost function was as follows:

$$w_r(u_x^2 + u_y^2) \qquad 0 \le t < T$$

$$
\begin{aligned}
w_p[(p_x - T_x)^2 &+ (p_y - Y_y)^2] \\
&+ w_v(\dot{p}_x^2 + \dot{p}_y^2) + w_r(u_x^2 + u_y^2) \qquad MT \le t < MT + MT_H
\end{aligned}
$$

where MT 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 $\hat{D} = \alpha\bar{D} + \gamma\bar{D}$, where α is the fraction and γ is a normally distributed random variable with 0 mean and variance σ^2. For simulations of the via-point task, we set $\hat{D} = \gamma\bar{D}$ (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:

$$
\begin{aligned}
w_r(u_x^2 + u_y^2) &\qquad 0 \le t < MT \\
w_{pv}(p_x - TV_x)^2 + w_{pv}(p_y - TV_y)^2 & \\
+ w_r(u_x^2 + u_y^2) &\qquad t = MT_v \\
w_p(p_x - T_x)^2 + w_p(p_y - T_y)^2 & \\
+ w_v(\dot{p}_{x^2} + \dot{p}_{x2}) + w_r(u_x^2 + u_y^2) &\qquad MT \le t < MT + MT_H
\end{aligned}
$$

where MT_v 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 "l") that overcompensates for the initial curl forces.
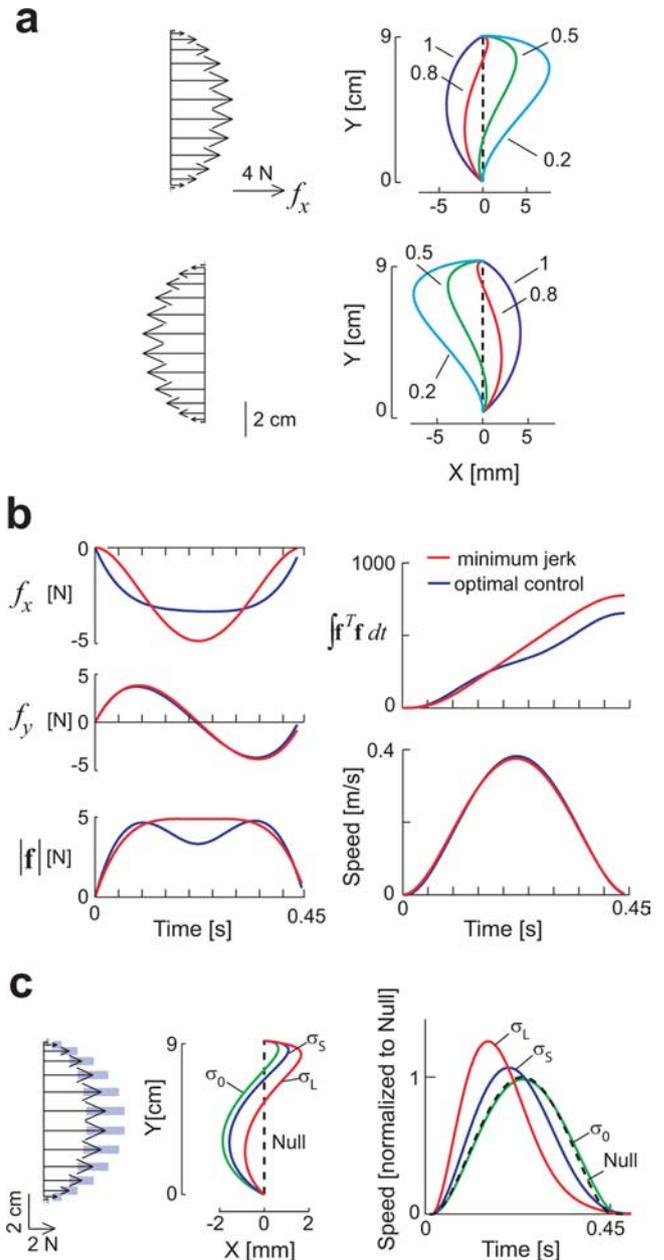

Figure 1. Predictions of OFC theory on mean trajectories of a point mass moving in force fields. The objective is to arrive at the target by 0.45 s, while minimizing motor costs. **a**, When there are no forces acting on the mass, the optimal policy is a straight-line path (dashed line). When a velocity-dependent field pushes the mass to the right, the optimal policy is not a return to the straight path but an apparent overcompensation to the left. The amount of overcompensation depends on the accuracy of one's model of the force field. If the field is described by a viscous matrix, D, this accuracy refers to α, where $\hat{D} = \alpha D$. The labels "1", "0.8", etc. refer to the value of α. The arrows are a schematic representation of the force field during an idealized reach. **b**, The subplots show the forces along the x- and y-direction that the controller produces (f_x and f_y) under an optimal policy (1 in **a**) compared with forces required to move the mass in a minimum-jerk path. The controller overcompensates early into the movement and undercompensates at peak velocity. This results in a total force $\int \mathbf{f}^T \mathbf{f} \, dt$ that is ~16% less than the minimum-jerk (straight line) path. **c**, The optimal policy when the controller's forward model is uncertain about the strength of the field. Left, The shaded area represents the SD of the force vectors that the mass would encounter along its path to the target. Middle, The optimum policy produces less overcompensation as the SD of the field increases from zero (σ_0), to a small ($\sigma_s = 0.2$), to a larger ($\sigma_L = 0.3$) value. Therefore, overcompensation is a good policy only if one is sure that the field will be present. In these simulations, $\alpha = 0.8$. Right, The speed profiles of the optimal policy (normalized to the maximum speed in the null condition) for different field variances. As the field becomes more variable, the optimal plan is a faster start, resulting in a reduced speed as the mass nears the target.
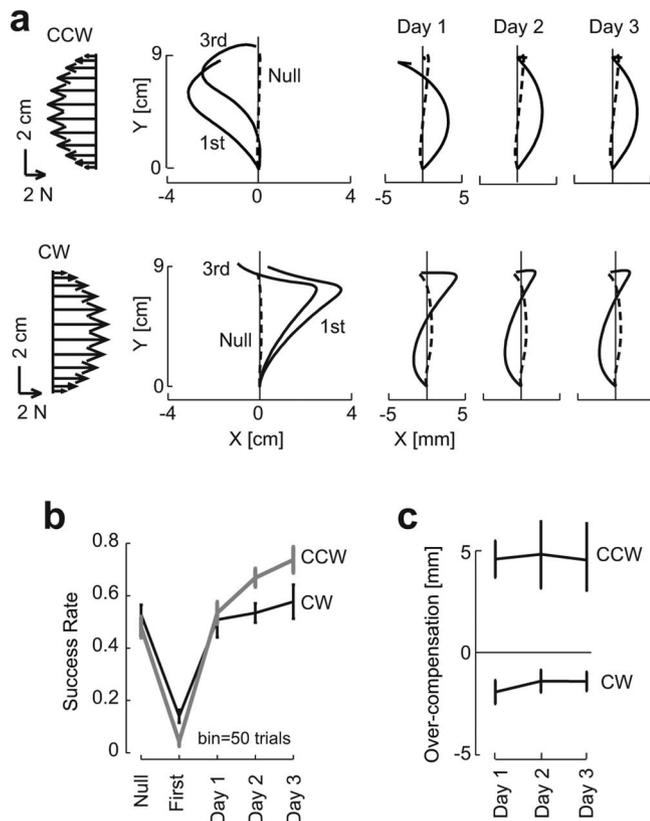
Figure 2. Reaching in a deterministic curl field. Subjects learned to reach in a single direction in a CW or CCW velocity-dependent force field that pushed the hand perpendicular to its direction of motion. The objective was to arrive at the target by 0.45 s. *a*, The arrows are a schematic representation of the force field during an idealized reach. The plots show hand paths of two representative subjects in the CW and CCW groups. The average trajectory in the null set and the first, third, and the average of the final 50 trials on days 1–3 are labeled. The dashed line is the trajectory in the null field, measured on the first day. *b*, Success rates (probability of arriving at target in time). Null, The last 50 trials in the null field; 1st and 3rd, the first and third trials of training on day 1; Day 1, Day 2, and Day 3, the last 50 trials of training on each day. *c*, A measure of overcompensation (maximum difference along the *x*-axis from the path of null trials). For the CCW group, this measure is positive to the left of the null trajectory and zero otherwise. For the CW group, this measure is positive to the right of the null trajectory and zero otherwise. Error bars are SEM.

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. 1*b*). The optimal controller produced less total force (Fig. 1*b*, $\int \mathbf{f}^{\mathrm{T}}\mathbf{f}dt$) 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 represented by $\mathbf{f} = D\dot{\mathbf{x}}$ and the learner's estimate is $\hat{D} = \alpha D$ for $0 \le \alpha \le 1$. When α is small (e.g., 0.2), the field overpowers the controller and pushes the mass in the direction of its perturbation (Fig. 1*a*). Because the controller acquires a more accurate forward model (α 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 ~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 ~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. 1*c*). 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. 1*c*), 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 2*a*. 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. 2*b*), 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. 2*c*). 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 3*a* shows the average hand trajectories in the last 50 trials of each day for subjects who trained in the deterministic (σ_0) or stochastic (σ_L) CW fields. In the stochastic field, movements gradually lost their overcompensation (Fig. 3*b*) (overcompensation in the σ_0 vs σ_L 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 σ_L group. During this block, subjects from the σ_L group were exposed to a zero variance field. That is, in this block of 50 trials, the environments for the σ_0 and σ_L groups were identical. Despite this, the overcompensation remained significantly smaller for the stochastic group ($p < 0.05$; df = 12). Because the σ_L 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 σ_L group should show larger errors in the direction of the force perturbations (Fig. 1*c*): 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. 3*c*, 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 σ_0 and σ_L were significantly different in day 3 ($p < 0.02$; df = 12), as was the difference between σ_0 (day 3) and σ_L (day 3: test) ($p < 0.02$). As overcompensation declined, performance in the σ_L group improved (Fig. 3*d*).

Finally, the model had predicted that the stochastic and deterministic groups would show different speed profiles (Fig. 1*c*). Figure 3*f* shows the speed profiles for the last training block on each day of training, and Figure 3*e* 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 block on day 1. By day 3, the σ_L 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 σ_L versus σ_0 ($p < 0.01$; df = 12).

The same tendencies were observed when subjects were trained in a stochastic CCW field (Fig. 4): overcompensation disappeared (Fig. 4*a*,*b*), the measure A1–A2 became more negative (Fig. 4*c*) as performance rates improved (Fig. 4*d*), and peak speeds increased (Fig. 4*e*). 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-
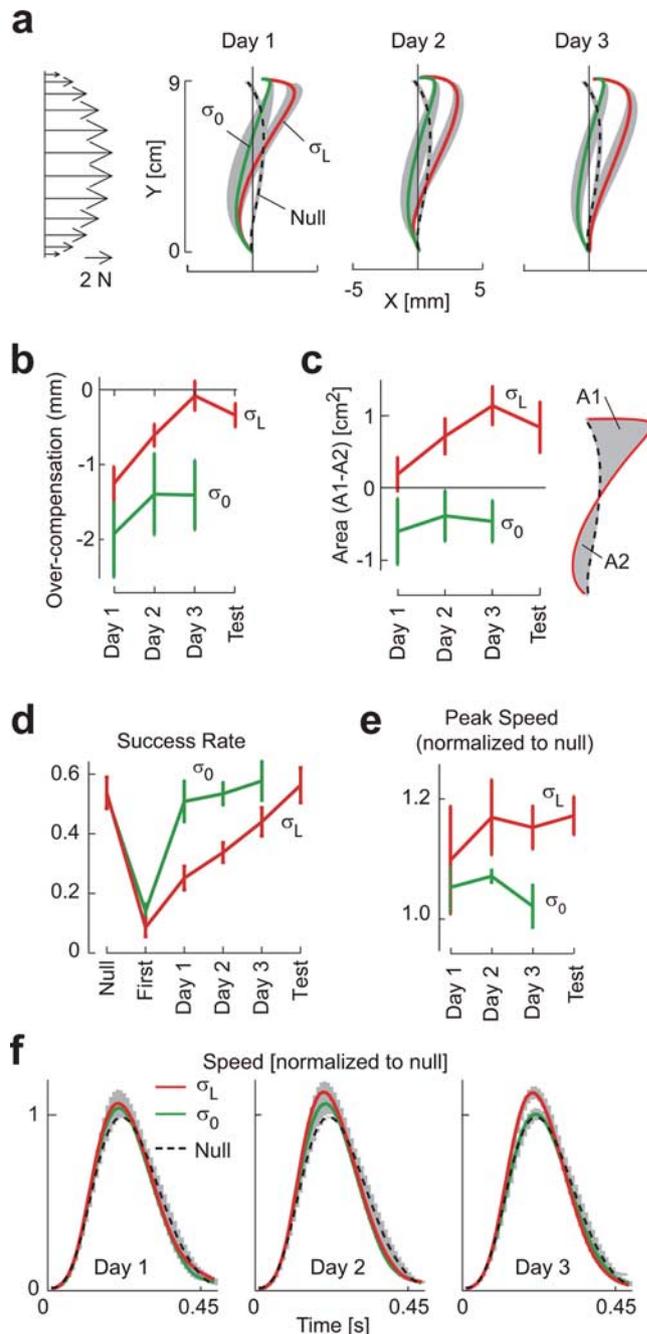

Figure 3. Reaching in a stochastic curl field $\mathbf{f} = D\dot{\mathbf{x}}$. On each trial, D was drawn from a normal distribution such that $\mathbf{f} = \bar{D}\dot{\mathbf{x}} + D\delta\dot{\mathbf{x}}$, where $\bar{D} = [0, 13; -13, 0]$ Nm/s (same value as in the deterministic field) and δ are a normally distributed scalar random variable with zero mean and SD $\sigma = 0.3$. The data for the deterministic and stochastic groups are labeled as σ_0 and σ_L. *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*, Over-compensation 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.
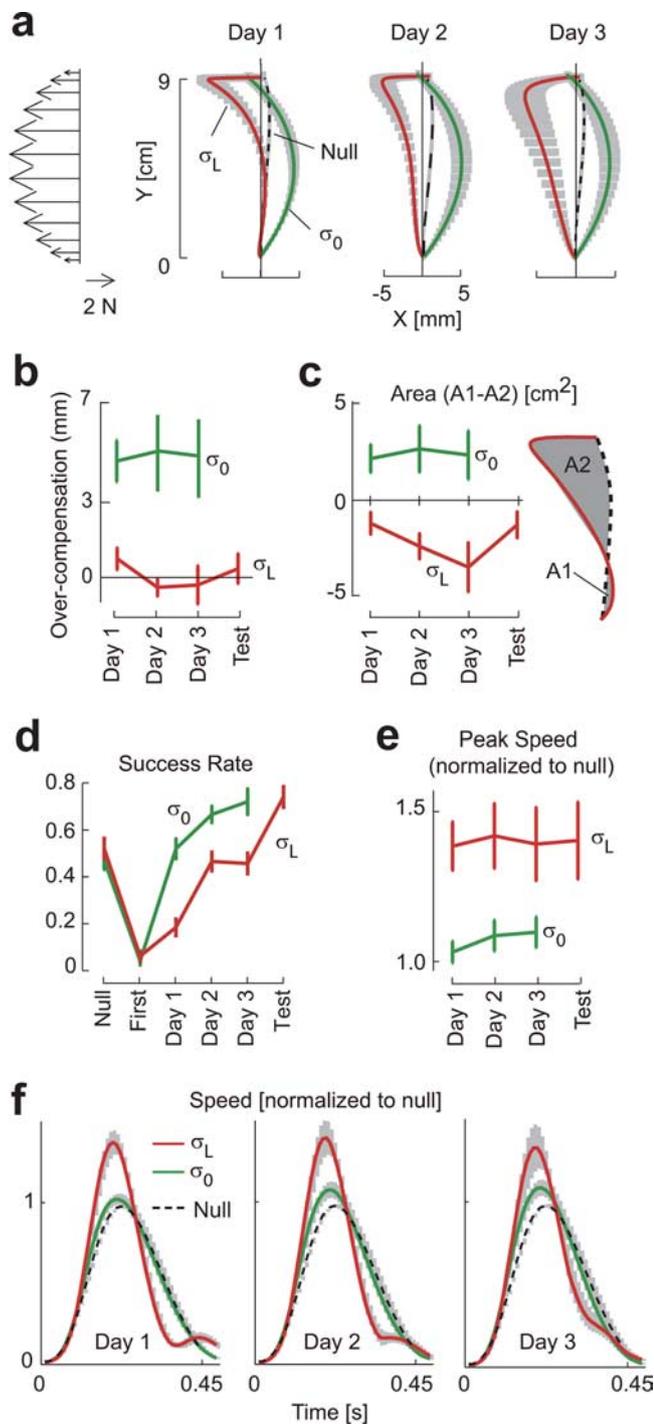
Figure 4. Same as Figure 3, except for a CCW stochastic curl field. The data for the deterministic and stochastic groups are labeled as σ_0 and σ_L. **a**, Mean hand path (averaged for the last 50 trials of training and across subjects) for the two groups. The dashed line represents the across-group average hand path in the null condition (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 right 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. Areas A1–A2 are show 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. All error bars are SEM.

ducing 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 σ_L 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 σ_S and a group that experienced a field with larger variability σ_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 σ_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 σ_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 σ_S and σ_L conditions, respectively (within-subject change across all subjects: main effect of condition, $p = 0.005$; for the σ_L 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
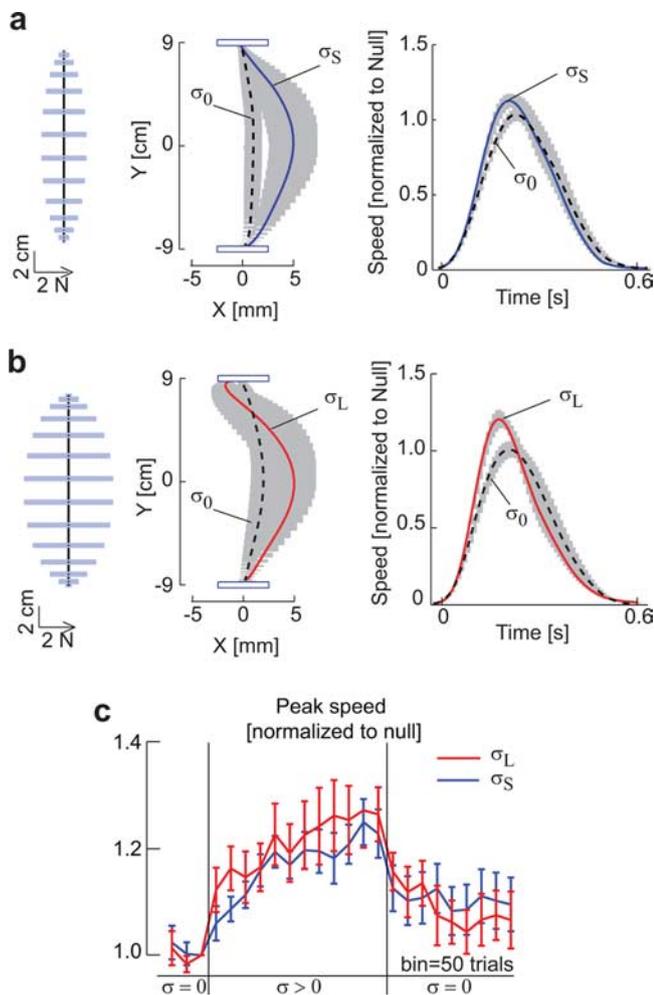
Figure 5. Reaching in a stochastic environment in which mean force is zero. Here, the environmental forces were $\mathbf{f} = D\dot{\mathbf{x}}$, where $D = [0, 13; -13, 0]$ Nm/s and δ are a normally distributed random variable with zero mean and variance σ^2. The task is to reach to the target in 0.6 s. **a**, Left, Schematic representation of a low-variance zero-mean curl field. Middle, Average hand paths in the deterministic (σ_0, dashed line) and small variance (σ_S, solid line) conditions. Right, The speed profiles. Data are for the last 50 trials in each condition. **b**, Same as in **a**, except for a high-variance condition. **c**, Peak speed for the two stochastic conditions during the experiment.

Figure 6. Reaching through a via-point in a stochastic zero-mean field. The task was to have the hand at the via-point target at 400 ms and at the final target at 1.0 s. **a**, The trajectory produced by the optimal controller when the environment had zero (σ_0), small (σ_S), or large (σ_L) variance. The hand path 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. **b**, Mean trajectories (last 50 trials of training) of subjects in the σ_0 and σ_S conditions. **c**, Trajectories in the σ_0 and σ_L conditions. Gray bars are SEM.

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 $\sigma_L = 0.6$ or $\sigma_L = 0.3$, followed by three sets of zero mean, zero variance force fields $\sigma_0 = 0$. 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 σ_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
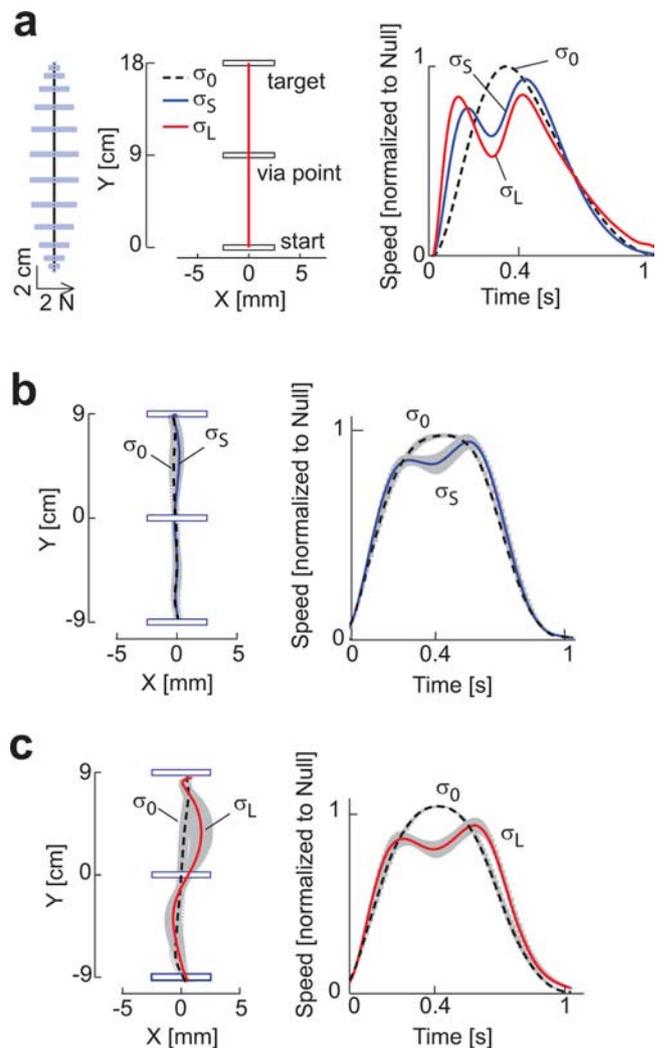
the σ_0 condition was significantly higher than the σ_L and σ_S conditions ($p < 0.01$ in each case).

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 Thoroughman et al. (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

Burdet E, Osu R, Franklin DW, Milner TE, Kawato M (2001) The central nervous system stabilizes unstable dynamics by learning optimal impedance. Nature 414:446–449.

Chen H, Hua SE, Smith MA, Lenz FA, Shadmehr R (2006) Effects of human cerebellar thalamus disruption on adaptive control of reaching. Cereb Cortex 16:1462–1473.

Chhabra M, Jacobs RA (2006) Near-optimal human adaptive control across different noise environments. J Neurosci 26:10883–10887.

Chib VS, Patton JL, Lynch KM, Mussa-Ivaldi FA (2006) Haptic identification of surfaces as fields of force. J Neurophysiol 95:1068–1077.

Conditt MA, Gandolfo F, Mussa-Ivaldi FA (1997) The motor system does not learn the dynamics of the arm by rote memorization of past experience. J Neurophysiol 78:554–560.

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

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.

Emken JL, Benitez R, Sideris A, Bobrow JE, Reinkensmeyer DJ (2007) Motor adaptation as a greedy optimization of error and effort. J Neurophysiol 97:3997–4006.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898–1902.

Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47:381–391.

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

Harris CM, Wolpert DM (2006) The main sequence of saccades optimizes speed accuracy trade off. Biol Cybern 95:21–29.

Hwang EJ, Shadmehr R (2005) Internal models of limb dynamics and encoding of limb state. J Neural Eng 2:S266–S278.

Hwang EJ, Donchin O, Smith MA, Shadmehr R (2003) A gain-field encoding of limb position and velocity in the internal model of arm dynamics. PLoS Biol 1:209–220.

Hwang EJ, Smith MA, Shadmehr R (2006) Adaptation and generalization in acceleration-dependent force fields. Exp Brain Res 169:496–506.

Jones KE, Hamilton AF, Wolpert DM (2002) Sources of signal-dependent noise during isometric force production. J Neurophysiol 88:1533–1544.

Krebs H, Hogan N, Hening W, Adamovich SV, Poizner H (2001) Procedural motor learning in Parkinson's disease. Exp Brain Res 141:425–437.

Li W, Todorov E (2007) Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic systems. Int J Control 80:1439–1453.

Martin TA, Keating JG, Goodkin HP, Bastian AJ, Thach WT (1996) Throwing while looking through prisms. I. Focal olivocerebellar lesions impair adaptation. Brain 119:1183–1198.

Maschke M, Gomez CM, Ebner TJ, Konczak J (2004) Hereditary cerebellar ataxia progressively impairs force adaptation during goal-directed arm movements. J Neurophysiol 91:230–238.

Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27:7105–7116.

Nowak DA, Timmann D, Hermsdorfer J (2007) Dexterity in cerebellar agenesis. Neuropsychologia 45:696–703.

Scheidt RA, Dingwell JB, Mussa-Ivaldi FA (2001) Learning to move amid uncertainty. J Neurophysiol 86:971–985.

Scheidt RA, Reinkensmeyer DJ, Conditt MA, Rymer WZ, Mussa-Ivaldi FA (2000) Persistence of motor adaptation during constrained, multi-joint, arm movements. J Neurophysiol 84:853–862.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

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

Smith MA, Shadmehr R (2005) Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. J Neurophysiol 93:2809–2821.

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

Thoroughman KA, Shadmehr R (2000) Learning of action through adaptive combination of motor primitives. Nature 407:742–747.

Thoroughman KA, Wang W, Tomov DN (2007) The influence of viscous loads on motor planning. J Neurophysiol 98:870–877.

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.

Uno Y, Kawato M, Suzuki R (1989) Formation and control of optimal trajectory in human multijoint arm movement. Minimum torque-change model. Biol Cybern 61:89–101.

van Beers RJ, Haggard P, Wolpert DM (2004) The role of execution noise in movement variability. J Neurophysiol 91:1050–1063.

Wainscott SK, Donchin O, Shadmehr R (2005) Internal models and contextual cues: encoding serial order and direction of movement. J Neurophysiol 93:786–800.

Wang T, Dordevic GS, Shadmehr R (2001) Learning the dynamics of reaching movements results in the modification of arm impedance and long-latency perturbation responses. Bio Cybern 85:437–448.