Optimal control of reaching includes kinematic constraints

Michael Mistry, Evangelos Theodorou, Stefan Schaal and Mitsuo Kawato

J Neurophysiol 110:1-11, 2013. First published 3 April 2013; doi:10.1152/jn.00794.2011

J Neurophysiol 110: 1–11, 2013.
First published April 3, 2013; doi:10.1152/jn.00794.2011.

Optimal control of reaching includes kinematic constraints

Michael Mistry,[1,2] Evangelos Theodorou,[2] Stefan Schaal,[1,2] and Mitsuo Kawato[1]
[1]*ATR Computational Neuroscience Laboratories, Kyoto, Japan; and* [2]*Department of Computer Science, University of Southern California, Los Angeles, California*

Submitted 30 August 2011; accepted in final form 1 April 2013

Mistry M, Theodorou E, Schaal S, Kawato M. Optimal control of reaching includes kinematic constraints. *J Neurophysiol* 110: 1–11, 2013. First published April 3, 2013; doi:10.1152/jn.00794.2011.—We investigate adaptation under a reaching task with an acceleration-based force field perturbation designed to alter the nominal straight hand trajectory in a potentially benign manner: pushing the hand off course in one direction before subsequently restoring towards the target. In this particular task, an explicit strategy to reduce motor effort requires a distinct deviation from the nominal rectilinear hand trajectory. Rather, our results display a clear directional preference during learning, as subjects adapted perturbed curved trajectories towards their initial baselines. We model this behavior using the framework of stochastic optimal control theory and an objective function that trades off the discordant requirements of *1*) target accuracy, *2*) motor effort, and *3*) kinematic invariance. Our work addresses the underlying objective of a reaching movement, and we suggest that robustness, particularly against internal model uncertainly, is as essential to the reaching task as terminal accuracy and energy efficiency.

motor control; optimal control; reaching; force fields; motor adaptation

OPTIMIZATION HAS LONG BEEN suggested as a guiding principle behind sensorimotor control, planning, and execution. Associated with optimization is always a cost function: the minimization of which defines the "best" possible movement or control strategy. Various cost functions have been proposed for human reaching: the minimization of jerk (Flash and Hogan 1985), torque change (Uno et al. 1989), variance (Harris and Wolpert 1998), interaction torques (Goble et al. 2007), or combinations of such (Ohta et al. 2004). Recent models using the framework of stochastic optimal control (SOC) have demonstrated how human reaching behavior can result from a control strategy that minimizes the sum of motor commands during a movement and positional error at the end of the movement (Todorov and Jordan 2002). Since the cost functions typically employed in these models do not include kinematic constraints, such as a desired trajectory for the hand to the target, they suggest the possibly that the central nervous system (CNS) need not be concerned about hand trajectory to plan and execute a successful reaching movement.

However, an important question to consider is what happens during the adaptation process? When learning to reach in a novel dynamic environment [such as a rotating room (Lackner and DiZio 1994) or perturbing robot manipulandum (Shadmehr and Mussa-Ivaldi 1994)], what are the factors that drive the adaptation strategy of the CNS? If we are to believe that optimization guides this process, then the reduction of cost, on a trial-by-trial basis, should be the preeminent motivation. Investigating the concept of "adaptation as reoptimization," Izawa et al. (2008) proposed a framework whereby learning occurs via progressively improved estimations of internal models and subsequent reoptimizations based on those improved models. Additionally, because this procedure may result in a postadaptation trajectory that differs significantly from baseline, the authors of this work claim that the CNS cannot be optimizing to restore an invariant trajectory. Rather, the CNS adapts only for the sake of improving end point accuracy and reducing motor effort.

While accuracy and effort are certainly relevant criteria for the sensorimotor control system, there has been evidence that trajectory does play a significant role in adaptation. For example, in Wolpert et al. (1995) subjects received artificially curved visual feedback of their hand position during a reaching task that did not affect accuracy at the target. Subjects adapted by curving their hand trajectory in the opposite direction to maintain a visually straight path, effectively increasing motor effort. Similarly, Kistemaker et al. (2010) created a novel force field where the minimal energy trajectory differed substantially from the nominally straight hand movement. Even after practicing on the minimum energy path, subjects still chose to restore to the suboptimal straight trajectory.

In this work, we test how subjects react to a force field perturbation where the maintenance of a rectilinear hand trajectory actively conflicts with the desire to reduce motor effort. We ask how the CNS trades off the discordant requirements of *1*) target accuracy, *2*) motor effort, and *3*) kinematic invariance. We suggest how the simple inclusion of kinematic constraints in the stochastic optimal control framework unifies this powerful theory with previous kinematically motivated models. We also suggest that such constraints are a means to robustify the neural controller against environmental uncertainty to aid in task achievement while bootstrapping the process of exploration and learning.

MATERIALS AND METHODS

A total of 16 healthy, right-handed subjects (ages 20–32, 14 male, 2 female), participated in this study. The institutional ethics committee of ATR Institute International approved this experiment, and subjects gave informed consent.

Inspired by the visuomotor adaptation experiment of Wolpert et al. (1995), we attempt to create a dynamic environment that affects motor cost but not accuracy: while reaching to a goal, if a force perturbation first pushes the hand off course in one direction and then subsequently back in the opposite direction, the goal may still be achieved with minimal correction. Fighting this perturbation to maintain a rectilinear hand trajectory may unnecessarily increase motor effort. Although there are potentially many such force fields that can accomplish this directive (e.g., Kistemaker et al. 2010), we also wished to model such a disturbance within existing linear optimal control modeling tech-

Address for reprint requests and other correspondence: M. Mistry, School of Computer Science, Univ. of Birmingham, Edgbaston, Birmingham, UK (e-mail: m.n.mistry@bham.ac.uk).

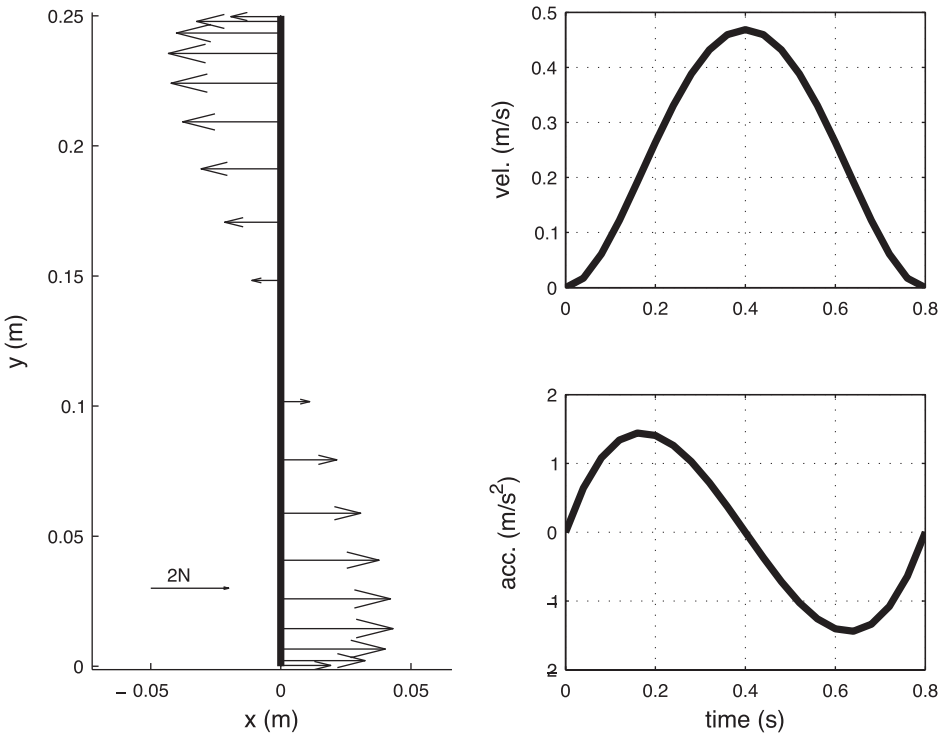

Fig. 1. Our acceleration force field shown as arrows for a minimum jerk reaching movement (vel., velocity). The force field, which only acts in the x direction, depends on y acceleration (acc.).

niques. Therefore, we chose a perturbation that is a linear function of hand kinematics:

$$F_x = 2.0\ddot{y} - 5.0\dot{x} \quad (1)$$

Thus, if the subject plans to execute a straight, minimum jerk-like movement, a disturbing force will first push the hand off course to the right during acceleration, and then subsequently left, back towards the target during deceleration (Fig. 1). The viscous term is required to dampen inertia sufficiently such that the deceleration disturbance actually reverses hand direction and additionally contributes to the overall stability of the device. The gains of the force field were tuned in pilot studies (results not shown) with the general goal of returning the hand to the target after a substantial initial off-course deviation (at least after the first "catch-trial" of force field onset, before adaptation occurs).

Additionally, because Wolpert et al. (1995) already addressed the issue of visually perceived error, we focus solely on the case of proprioceptive feedback. Thus we give no visual feedback to our subjects, except for the terminal position (hand position at the end of a trial). Also, because we wanted to focus our study to the adaptation of the acceleration term, we left the viscous damping term on for all "null" field trials.

Apparatus

We use the ATR parallel-link, 2 degree-of-freedom, direct drive, air-magnet floating manipulandum (PFM) shown in Fig. 2. Subjects were seated in a chair and strapped in with shoulder harness to keep their back against the chair. A wrist brace was attached to their right hand to restrict wrist motion and connect the subject to the handle of the device. The combination of wrist brace and shoulder straps only allows for planar shoulder and elbow motion. The subjects cannot see their arm, as the tabletop covers the entire workspace. Start and target locations are projected onto the table, with a ceiling mounted projector, as well as the positional feedback at the end of the movement. Kinematic data are recorded at a 500-Hz rate. The force applied by subjects at the handle of the PFM is measured with a force sensor (Nitta no. 328) with resolution of 0.06 N. The manipulandum and setup were described in detail previously (Gomi and Kawato 1996).

Experimental Procedure

Subjects are asked to execute 25-cm point-to-point reaching movements, within a time window of 800 ± 150 ms, to a specified target in a single direction, perpendicular to the frontal plane, away from the body. The target is indicated by a 2.5-cm diameter circle. Before the start of each trial, the subjects move their hands to the starting location, indicated by a 2-cm diameter circle. During this phase, visual feedback of the hand position is given (indicated by a dot) but only

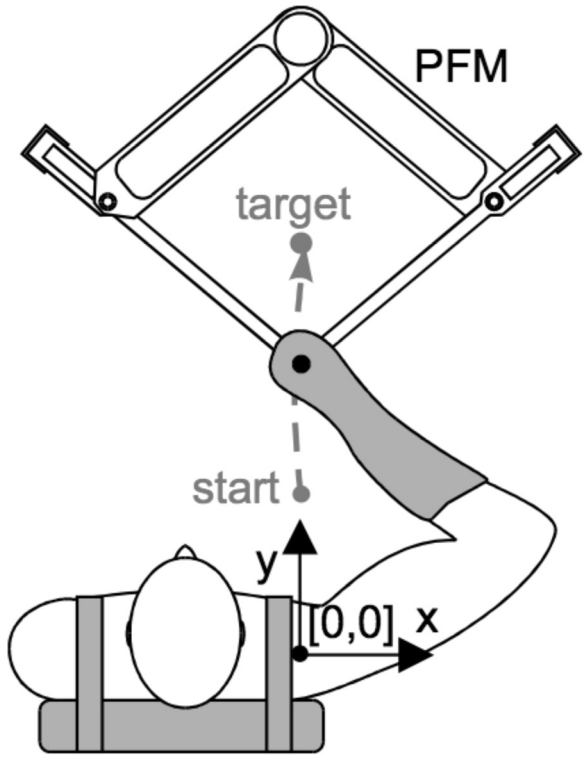

Fig. 2. ATR planar force manipulandum (PFM).

within 4 cm of the starting location, and only so the subject can properly position their hand to start. Once inside the starting circle, visual feedback is removed, and three consecutive beeps sound to indicate the start of the trial. Subjects are asked to initiate movement after conclusion of the three beeps and stop at the target within the specified time window. No other instructions regarding how to execute the movement or which path to take to the target were given. No visual feedback is given during the trial, except a single dot indicating the final hand position at the end of the trial. The subject is also told after each trial whether the trial was a success or failure, where success is defined by the hand arriving and stopping within the target circle during the specified time window. If the trial was a failure, the subject is told if the hand arrived too late, too early, or was off target.

Experimental Conditions

Each subject was tested under one of two conditions. Under the first condition (*experiment A*), after initial training in the null field, the subjects experienced the full strength of the force field within a series of random catch trials. For this group of subjects, we were able to estimate the Before Effects: reaching behavior in the force field before adaptation onset. As the Before Effects were strongly curved trajectories, significantly different from baseline, we were concerned that subjects would be aware of the abrupt change to motor behavior and subsequently make a conscious effort to adapt their motor strategy. To minimize the effect of conscious adaptation, we also tested a separate group of subjects under a condition where force field strength gradually increased over a series of trials (*experiment B*). In both conditions, after sufficient training in the force field we were able to measure After Effects (by switching the force field off in random catch trials).

Experiment A had six subjects (5 male, 1 female). Each was tested in five blocks of trials, with a 5-min break in between each block. In *block 1*, subjects trained on the null field condition until 70 successful trials were completed. In *block 2*, subjects executed 30 successful null field trials, followed by 80 trials containing 10 random catch trials (where the field is switched on unknowingly to the subject). *Blocks 3* and *4* required 60 successful trials each in the force field. Finally, in *block 5* subjects executed 30 successful force field trials followed by 80 trials containing 10 random catch trials (with the field off during these 10 trials) to test for After Effects. Over the course of the session, each subject trains for a total of 100 successful null field trials and 150 successful force field trials. For *experiment B*, we tested a total of 10 subjects (9 male, 1 female) with 4 blocks (again with 5-min breaks in between each block). *Block 1* was identical as in *experiment A*. In *block 2*, subjects executed 30 successful null field trials, followed by 18 trials with the force field gain linearly rising from trial to trial until reaching full strength and then followed by an additional 40 successful trials in the full strength force field. *Block 3* required 70 successful trials in the full strength force field. Finally, in *block 4* subjects executed 40 successful trials in the force field followed by 80 trials containing 10 random catch trials with the field off. Similarly to *experiment A*, each subject trains for a total of 100 successful null field trials and 150 successful force field trials.

Data Analysis

To quantify the curvature of a trajectory, we compute the average perpendicular distance among all points in the trajectory from a straight line that connects the start of the trial to the target. We assign this metric a positive value if the greater portion of the trajectory is on the right side of the line; otherwise, it receives a negative value. For the purpose of this data analysis, the start of a trial is considered to be the last time the velocity is 0.05 m/s before reaching peak velocity. The end of the trial is the first time the velocity reaches 0.05 m/s after peak velocity.

Modeling and Simulation

SOC has been a powerful tool for modeling the estimation and control of biological sensorimotor systems, particularly because such systems are inherently noisy and partially observable (Loeb et al. 1990). If one assumes a linear model of dynamics with additive Gaussian noise, and a globally quadratic cost function, a cost-minimizing optimal controller can be derived analytically via the Linear-Quadratic-Gaussian (LQG) framework. Additionally, Todorov (2005) extends the LQG framework to handle the case of signal-dependent noise, a property believed to exist in biological muscles and affect sensorimotor planning (Harris and Wolpert 1998).

Stochastic optimal control. The SOC model, in its general form, assumes the following linear dynamics equation:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \xi_t + \sum_{i=1}^{c} \varepsilon_t^i C_i \mathbf{u}_t \tag{2}$$

where $\mathbf{x}_t$ is the state vector; $\mathbf{u}_t$ is the control input vector; ξ_t and ε_t^i are the additive and multiplicative zero-mean Gaussian noise variables, respectively; A is the state transition matrix, B is the control input matrix, and C_i are scaling matrixes for control-dependent noise. The model also assumes sensory feedback is partial observable, and given by the equation:

$$\mathbf{y}_t = H\mathbf{x}_t + \omega_t \tag{3}$$

where $\mathbf{y}_t$ are the observations, H is the observation matrix, and ω_t is zero-mean Gaussian noise. Additionally, the following quadratic cost is accrued at each time step:

$$0 \leq \mathbf{x}_t^T Q_t \mathbf{x}_t + \mathbf{u}_t^T R \mathbf{u}_t \tag{4}$$

where the matrix R is a positive definite and the matrix Q_t is positive semidefinite. Under these assumptions, Todorov (2005) derives the optimal controller and estimator in the following form:

$$\mathbf{u}_t = -L_t \hat{\mathbf{x}}_t \tag{5}$$

$$\hat{\mathbf{x}}_{t+1} = (A - BL_t)\hat{\mathbf{x}}_t + K_t(\mathbf{y}_t - H\hat{\mathbf{x}}_t) + \eta_t \tag{6}$$

where $\hat{\mathbf{x}}_t$ is the state estimate at time t and η_t is zero mean Gaussian noise. The optimal control gain, L_t, and optimal estimator gain (or Kalman gain), K_t, are computed via an iterative procedure derived from dynamic programming principles. Please see Todorov (2005) for full details.

Model of sensorimotor control of human reaching. The assumptions of linear dynamics and a globally quadratic cost may seem to be too restrictive, particularly for the highly nonlinear problem of sensorimotor control of human reaching. However, as demonstrated in Todorov (2005), SOC still has a predictive power, particularly for systems operating under uncertainty. Following the model of the motor system derived in Todorov (2005), and also employed in Izawa et al. (2008), we represent the human arm with a 10-dimensional state vector as follows:

$$\mathbf{x}_t = \left(\mathbf{p}_t^T \mathbf{v}_t^T \mathbf{f}_t^T \mathbf{g}_t^T \mathbf{p}^{*T}\right)^T \tag{7}$$

where $\mathbf{p}_t$ and $\mathbf{v}_t$ are the two-dimensional position and velocity of a point mass, respectively. The vector $\mathbf{p}*$ represents the x and y target location. Muscles are modeled as a second order linear system, where their output force, $\mathbf{f}_t$, is a function of muscle activity, $\mathbf{g}_t$, and muscle activity is driven by control input with signal-dependent noise. The complete linear dynamics equations of the arm model are defined as follows:

$$\mathbf{p}_{t+1} = \mathbf{p}_t + \mathbf{v}_t \Delta_t \tag{8}$$

$$\mathbf{v}_{t+1} = \mathbf{v}_t + \left(\frac{\mathrm{f}_t + D\mathbf{x}_t}{m}\right)\Delta_t \tag{9}$$

$$\mathbf{f}_{t+1} = \mathbf{f}_t\left(1 - \frac{\Delta_t}{\tau_2}\right) + \mathbf{g}_t\frac{\Delta_t}{\tau_2} \quad (10)$$

$$\mathbf{g}_{t+1} = \mathbf{g}_t\left(1 - \frac{\Delta_t}{\tau_1}\right) + \mathbf{u}_t\left(1 + \sigma_\varepsilon\varepsilon_t\right)\frac{\Delta_t}{\tau_2} \quad (11)$$

where t is the time step and $D\mathbf{x}_t$ (used in *Eq. 9*) is some linear, state dependent, external force, such as a force field. For example, in the case of our acceleration-based force field (1), we write:

$$D\mathbf{x}_t = \beta\begin{bmatrix} 0 & 2.0/\mathrm{m} \\ 0 & 0 \end{bmatrix}\mathbf{f}_t - \begin{bmatrix} 5.0 & 0 \\ 0 & 0 \end{bmatrix}\mathbf{v}_t. \quad (12)$$

The parameter β is used in the modeling of *experiment B* to gradually scale field strength from 0 to 1. Otherwise, $\beta = 1$ when the field is on, and $\beta = 0$ in the null field case.

The total quadratic cost, accumulated for the entire movement, is defined as follows:

$$w_p\left(\mathbf{p}_{t_f} - \mathbf{p}^*\right)^T\left(\mathbf{p}_{t_f} - \mathbf{p}^*\right) + w_v\mathbf{v}_{t_f}^T\mathbf{v}_{t_f} + w_f\mathbf{f}_{t_f}^T\mathbf{f}_{t_f} + \sum_{t=0}^{t_f}\mathbf{u}_t^T R\mathbf{u}_t, \quad (13)$$

or in a more compact notation:

$$\mathbf{x}_{t_f}^T Q_{t_f}\mathbf{x}_{t_f} + \sum_{t=0}^{t_f}\mathbf{u}_t^T R\mathbf{u}_t, \quad (14)$$

where the parameters w_p, w_v, w_f, and R weight the importance of target accuracy, terminal velocity, terminal force, and control input, respectively. This cost function effectively defines the objective of the reaching movement: be at the target by a specified time (t_f) and stop there (with zero velocity and zero force), all while using the minimal amount of control effort as possible. It is important to note that the cost function does not specify any direction or path to the target. Assuming weight parameters are appropriately tuned, various human-like motions (e.g., minimum jerk trajectories) are shown to emerge from the minimization of this trade-off between target accuracy with motor effort (Todorov 2005). For reasons that will become apparent, we henceforth refer to this cost function as nondirectional.

Reoptimization as a model of learning. The work of Izawa et al. (2008) extends the SOC framework to incorporate learning and uncertainty of force field dynamics. Here, we will borrow their technique of modeling learning behavior. When learning a new force field, the subject does not have full knowledge of the external force matrix D. Rather, the subject can only estimate D as:

$$\hat{D} = \alpha D, \quad (15)$$

with $0 \leq \alpha \leq 1$. Thus, learning behavior can be approximated by gradually increasing α, effectively improving the subject's estimation of the force field. When implementing this framework, the optimal control and estimator gains are computed using the estimate, $\hat{D}$, while the resulting simulation uses the actual D. In this work, we also use reoptimization to model behavior during catch trials. Before Effects are modeled simply as $\beta = 1.0$ and $\alpha = 0.0$. To model After Effects, we use $\beta = 1.0$ and $\alpha = 1.0$ in the model estimate, however, when running the simulation, we set $\beta = 0.0$.

Modeling directional constraint. In this work, we also extend the SOC cost function (14) to include a kinematic constraint, or preferred direction of motion.[1] Calling $\mathbf{d} = [\, d_x\ d_y\,]^T$ a unit vector representing desired direction, we define the matrix Q_d as:

$$Q_d = \begin{bmatrix} d_y^2 & -d_x d_y \\ -d_x d_y & d_x^2 \end{bmatrix} \quad (16)$$

and add new terms to the original cost function, with the intent to penalize any position or velocity perpendicular to the desired direction:

$$\mathbf{x}_{t_f}^T Q_{t_f}\mathbf{x}_{t_f} + \sum_{t=0}^{t_f}\left[\mathbf{u}_t^T R\mathbf{u}_t + e^{-t\Delta_t/\tau}\left(k_p\mathbf{p}_t^T Q_d\mathbf{p}_t + k_v\mathbf{v}_t^T Q_d\mathbf{v}_t\right)\right], \quad (17)$$

where k_p and k_v weight the importance of error in position and velocity, respectively. An exponential decay term is also included to emphasize that the kinematic constraint need not exist for the entire motion. Note that the new cost function remains quadratic with respect to the state vector and thus can be used in LQG methods. For the purpose of our experiments and simulations, we set $\mathbf{d} = [\, 0\ 1\,]^T$ (the direction towards the target at the start of the motion), and the cost function simplifies to:

$$\mathbf{x}_{t_f}^T Q_{t_f}\mathbf{x}_{t_f} + \sum_{t=0}^{t_f}\left[\mathbf{u}_t^T R\mathbf{u}_t + e^{-t\Delta_t/\tau}\left(k_p p_x^2 + k_v v_x^2\right)\right]. \quad (18)$$

Simulation examples with a viscous curl field. For verification purposes, we test the SOC with reoptimization framework, under a traditional rightward viscous curl field. We use the same target distance and time constraints as our intended experiment and tune the parameters of the nondirectional cost function to achieve qualitatively the same performance as in (Izawa et al. 2008). Figure 3, *left*, shows the trajectory predictions made by the SOC framework with increasingly improved estimates of the force field. This demonstration repeats and confirms the results reported in Izawa et al. (2008) and additionally shows our predictions of Before and After Effects. Note that, as emphasized in Izawa et al. (2008), the fully adapted optimal controller does not move straight to the target but rather moves with a significant leftward curvature. Less total effort can be expended by first moving against the field, while velocity and force are low, and subsequently allowing the field to push the hand back towards the target. We add directional constraint to the cost function, with $\tau = 0.07$ and tune only the directional constraint terms to achieve a nearly straight trajectory in the fully adapted case. All other parameters are identical to the previous case. Not surprisingly, using a directional constraint can result in an optimal trajectory close to baseline (Fig. 3, *right*).

RESULTS

Figure 4, *right column*, shows the trajectory plots averaged across all 6 subjects of *experiment A* (A.DATA), as well as the average across the 10 subjects of *experiment B* (B.DATA). The subjects' nearly straight baseline trajectory is the result after at least 100 successful trails in the null field. The significant rightward curvature seen in the Before Effect (A.DATA) illustrates the preadaptation effect of the acceleration force field perturbation on the reaching movement: pushing the hand to the right before restoring towards the target. Final Force trajectories show subsequent adaptation after at least 150 successful training trials in the force field. Subjects either restored (in the case of *experiment A*) or maintained (in the case of *experiment B*) their hand trajectories to approximately baseline. Strong leftward curved After Effects (in both conditions) indicate the formation of a predictive internal model of the force field dynamics. The thin lines in A.DATA and B.DATA show some incremental stages of learning (the statistics of which are reported in Tables 1 and 2). The learning curve plots in Fig. 5 additionally show how average curvature progresses during force field training. In Fig. 6, DATA are the force measurements in the direction perpendicular to the target (same di-

[1] Here we treat kinematic invariance as a soft constraint, i.e., as an optimization objective, as opposed to a hard constraint, which must necessarily be satisfied by the optimization routine.

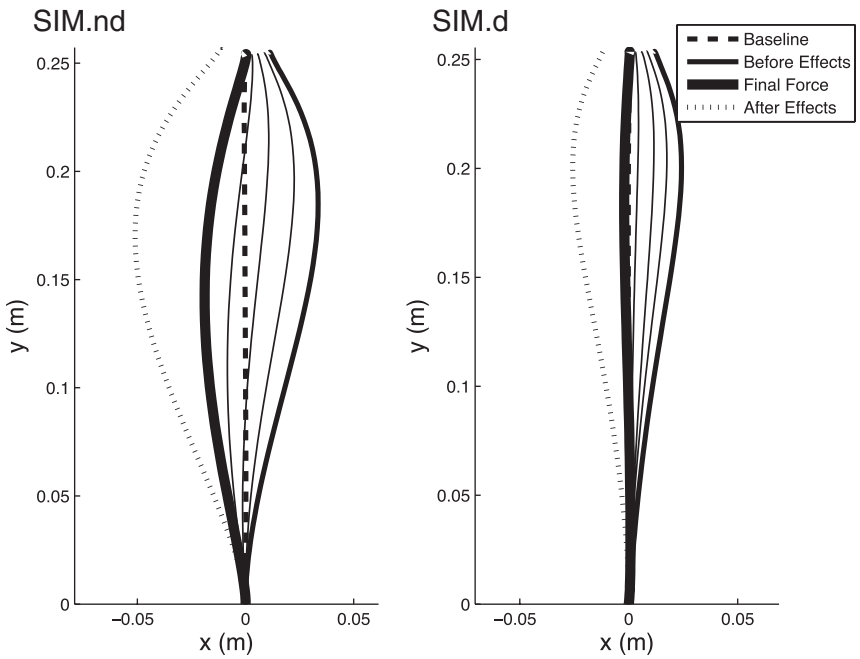

Fig. 3. Prediction made by stochastic optimal control under a rightward viscous curl field. Each line plots the average of 50 simulations. With the use of a nondirectional cost function (SIM.nd), the optimal action (shown by the thickest line) is to move with a significant leftward curvature, first working against the field while velocity is low and then subsequently allowing the field to push the hand back towards the target. While under the directional cost function (SIM.d), the optimal action is to restore the trajectory towards baseline. Before Effects (medium thickness line) show the action before learning occurs ($\alpha = 0.00$), while thin lines indicate incremental stages of learning ($\alpha = 0.25$, 0.50, and 0.75).

rection as the force field perturbation), recorded by the force sensor at the handle of the manipulandum. Both the averaged force trajectories (averaged across all 6 subjects of *experiment A*) as well as the individual trajectories from a single subject, have substantial leftward (negative) force, early in the movement, providing resistance against the force field perturbation. Evidence of resistive forces in preadaptation trajectories (Before Effects) indicate that the opposing force may be reactionary, correcting for error observed within a single trial, and not necessarily preplanned before execution.

Modeling Using Stochastic Optimal Control

Using the linear dynamics model described previously and the nondirectional cost function defined by (14), we use the stochastic optimal control framework to compute the optimal control policy under our force field. Cost function parameters were tuned to achieve bell-shaped velocity profiles and goal achievement in the baseline and fully adapted cases. Not surprisingly, the SOC model predicts that the optimal behavior in the force field is to move with a significant rightward curvature, allowing the disturbance to push the hand off course before restoring towards the target (Fig. 4, A.SIM.nd: Final Force). The preadaptation trajectory is curved to the right, but misses the target towards the left (Fig. 4, A.SIM.nd: Before Effects). The reoptimization model predicts increasingly outwards curvature as better estimates of the force field are obtained (Fig. 4, A.SIM.nd: thin lines between Before Effects and Final Force). Increasingly outwards curvature is also predicted if field strength is gradually increased from zero (Fig. 4, B.SIM.nd: thin lines between Baseline and Final Force). Results of these simulations imply that the stochastic optimal control and reoptimization framework fail to predict the correct adaptation behavior of our subjects.

We add the directional constraint term to the cost function as in (18) and repeat the simulations. We set the time constant of the decay term to $\tau = 0.07$ to achieve 90% decay at approximately the time of peak acceleration (see DISCUSSION). We additionally tune only the directional constraint terms of the cost function to change the adaptation behavior and better represent results from our subjects. All other parameters remained the same as in the nondirectional case. With the use of a directional cost function, reoptimization now predicts an inward adaptation trend (Fig. 4, A.SIM.d: thin lines between Before Effects and Final Force), until the final trajectory is close to Baseline (Fig. 4, A.SIM.d: Final Force). Gradually increasing field strength from zero, results in little change to curvature (Fig. 4, B.SIM.d). Output forces from the directional cost model (Fig. 6, SIM) resist the force field, matching the behavior observed in subjects. On the contrary, the nondirectional model applies force in the same direction as the force field at the initial stage of the movement. Overall, the SOC model with a directional cost function provides an improved resemblance to our behavioral results.

DISCUSSION

We investigated adaptation within a reaching task where perturbations altered nominal hand trajectory in a potentially benign manner: pushing the hand off course in one direction before subsequently restoring towards the target. The task was designed to elicit the underlying adaptation strategy employed by the CNS, since in this particular task, a strategy to reduce motor effort requires a distinct deviation from a rectilinear hand trajectory. Rather, our results display a clear directional preference during adaptation. Hypotheses proposing that kinematic error does not play a role in sensorimotor adaptation, e.g., as suggested by Izawa et al. (2008), cannot hold.

Although our results concur with a long-standing viewpoint that kinematics influences motor behavior, it remains clear that strict invariants such as minimum jerk or rectilinearity do not dominate motor control, e.g., as proposed in older works such as Flash and Hogan (1985) and Shadmehr and Mussa-Ivaldi (1994). Extended trials in viscous force fields resulted in an adaptation to energy optimal curved trajectories (Izawa et al.

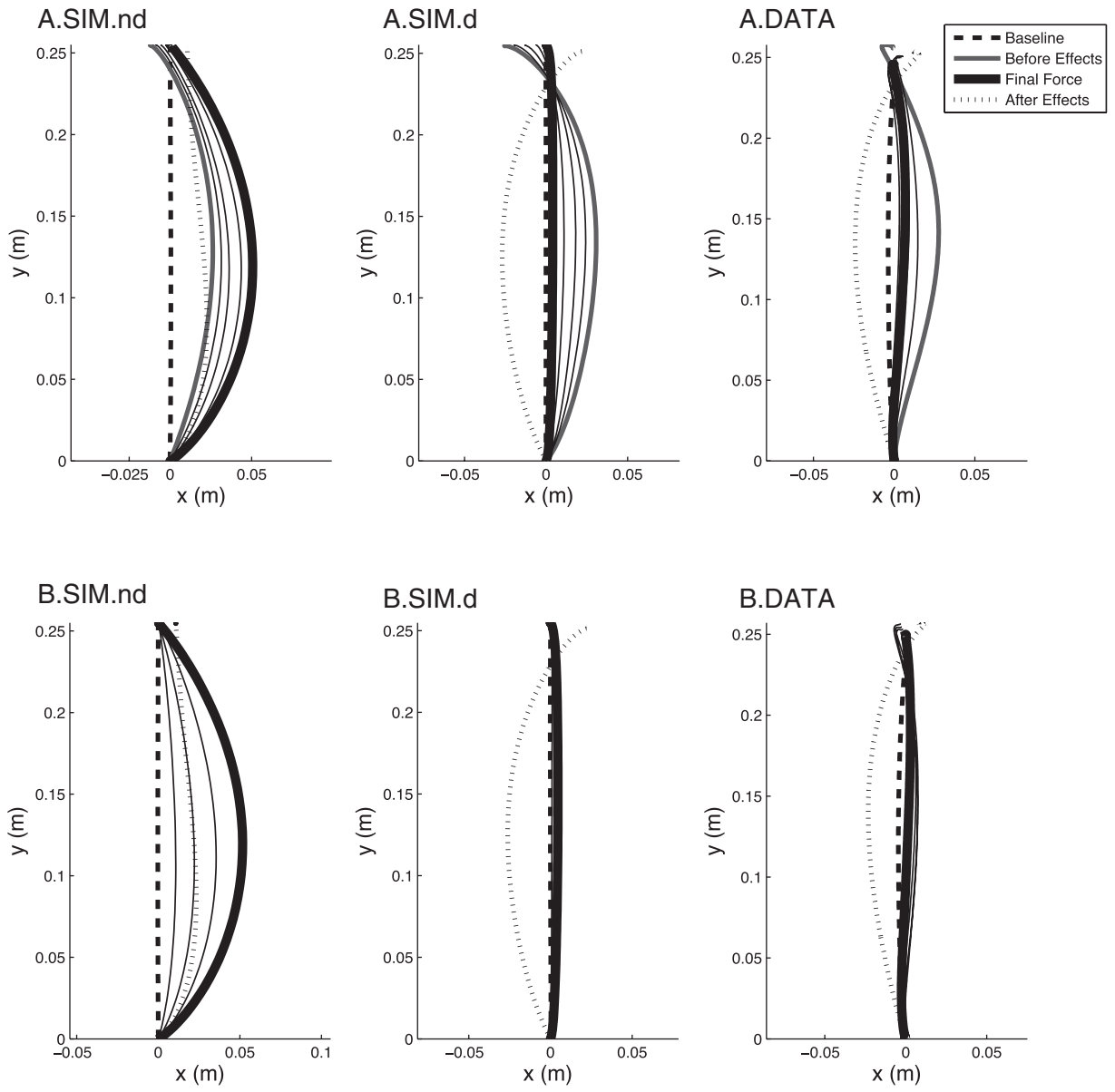

Fig. 4. *Top row*: trajectories and simulations from *experiment A*. *Bottom row*: from *experiment B*. In DATA plots (*right column*), each line is the average trajectory of all subjects (A.DATA: 6 subjects; B.DATA: 10 subjects). Baseline trajectories (long dashed lines) are the average of the last 15 trials in null field training. Final Force (thickest solid line) are the average of the last 15 trials in force field training. Before Effects (grey solid line, A.DATA only) averages the 10 catch trials following null field training. After Effects (short dashed lines) averages the 10 catch trials following force field training. Thin solid lines show incremental stages of learning, the statistics of which are reported in Table 1 (for A.DATA) and Table 2 (for B.DATA). *Left column*: simulations of the stochastic optimal controller using the nondirectional cost function (SIM.nd). *Middle column*: use of directional cost (SIM.d). Each SIM line is the average of 50 simulations. Under either cost function, preadaptation behavior in the force field (Before Effects, $\alpha = 0.0$, $\beta = 1.0$) is a rightward curved trajectory that misses the target to the left. Incremental learning is represented by thin lines (with $\alpha = 0.25$, 0.50, and 0.75) until complete learning (Final Force, $\alpha = 1.0$). In the nondirectional case (A.SIM.nd), the trajectory adapts outward, with increasing curvature. However, in the directional case (A.SIM.d), adaptation occurs inward with decreasing curvature. Similarly, a gradual increase in field strength predicts gradually increasing curvature in the nondirectional case (B.SIM.nd) and little change in the directional case (B.SIM.nd). Thin lines in B.SIM are $\beta = 0.25$, 0.50, and 0.75, with thickest line $\beta = 1.0$ (and perfect estimation is assumed, $\alpha = 1.0$, as field strength is increased).

2008), especially without visual feedback (Arce et al. 2009). We also speculate that with sufficient training in our acceleration force field, subjects may also learn to conduct curved trajectories. Indeed, the reduction of motor effort is a powerful motivation for the CNS. In this work, however, we address the short-term adaptation strategy, particularly during periods of significant environmental uncertainty. We argue that during adaptation, a robust strategy can be implemented via a control mechanism that trades off the desire for a rectilinear trajectory with the aspirations of terminal accuracy and low motor effort. In our model, we propose an optimal control cost function that represents this trade-off. Adding a kinematic constraint term to the SOC cost function introduces the additional requirement of rectilinearity while still able to exploit the predictive power of SOC for modeling biological systems with signal-dependent noise and partial observability. Thus we are able to unify the conflicting notions of desired trajectory and energy efficiency within an existing computational framework.

Table 1. *Statistics for experiment A*

Trajectory	Mean, m	SD, m	*P*	Significance	Success Rate
Baseline	−0.0020	0.0018			58.9%
Before Effects	0.0128	0.0028	3.0×10^{-6}	*	20.0%
Force Trials [*1–10*]	0.0086	0.0030	6.6×10^{-5}	*	25.0%
Force Trials [*11–20*]	0.0050	0.0036	3.6×10^{-3}	*	25.0%
Force Trials [*21–30*]	0.0038	0.0033	6.5×10^{-3}	*	36.7%
Force Trials [*31–40*]	7.7×10^{-4}	0.0050	0.2500		61.7%
Final Force	0.0032	0.0054	0.0694		65.6%
After Effects	−0.0131	0.0043	7.5×10^{-4}	*	16.7%

Trajectories Baseline, Before Effects, Final Force, and After Effects are as defined and plotted in Fig. 4, A.DATA). Force Trials represent bins of 10 of trials each from *block 3* (the 1st block of force field training) and are plotted as thin lines in Fig. 4, A.DATA). Mean and SD report the mean trajectory deviation and SD (respectively) of each trajectory for the 6 subjects of *experiment A*. *P* and significance report the significance from baseline (paired *t*-test, degrees of freedom = 5, $*P < 0.01$). Success rates for each trajectory are also shown.

There has been a long-standing discourse in motor control literature regarding the desired trajectory hypothesis (DTH), i.e., the notion that a movement trajectory is preplanned before execution. DTH has largely been supported by observations that subjects restored baseline trajectories after adaptation (Shadmehr and Mussa-Ivaldi 1994; Wolpert et al. 1995). DTH is also appealing from a computational perspective, as trajectory tracking is a simple, well-studied control methodology, commonly used in engineered systems. Recently, however, significant works have used stochastic optimal control theory to explain the trajectory as well as variability profiles of human reaching by means of only parsimonious task criteria: maximization of target accuracy and minimization of control effort. The main rationale behind these works is dictated by the minimum intervention principle (Todorov and Jordan 2002): deviations from baseline need only be corrected if they interfere with task performance. If the task is only to minimize energy and maximize end-point accuracy, prescribing a trajectory before execution makes little sense, as particular paths to the target are often inconsequential to goal achievement. Moreover, the process of enforcing such a trajectory may add extraneous motor cost and, in the case of systems with signal-dependent noise, introduce additional execution noise. Rather, the sensorimotor system is better left unrestrained by kinematic invariants and able to reserve decisions of how to best reach the target at execution time, when relevant contextual information is available.

This rational makes sense for the nominal case of learned reaching movements, where underlying dynamics are fully known and can be exploited to achieve the task accurately and efficiently. However, during a learning scenario, e.g., when exposed to a novel force field, the use of an inaccurate model becomes detrimental and potentially unsafe. In this case, to successfully reach the target, the brain must employ a robust strategy that is less sensitive to modeling inaccuracies. Any unknown force field dynamics can be treated as disturbances to be rejected, at least until an improved model is formed. Humans often apply such a strategy during learning by increasing arm stiffness, thereby reducing the effect of unknown disturbances (Takahashi et al. 2001) and then subsequently lowering stiffness levels as a new model is formed (Franklin et al. 2008). In our model, we see that adding kinematic constraints to the objective function increases the position and velocity gains of the controller, perpendicular to target direction, during the early part of the movement before the constraint has decayed (Fig. 7). For our linear system, these gain increases can effectively be viewed as stiffness and damping increases. The amplification of gains is likely to result in more overall energy expenditure, however, but is also likely to reduce the deviations from target direction. We see this effect when we compute the controller using the nominal mass value but test on varying mass values. A controller with directional constraint will be less sensitive to erroneous internal models (Fig. 8). Thus we argue that kinematic invariants, such as a directional preference, do more for the CNS than merely simplifying the control problem or improving its computational tractability. Kinematic constraints serve to robustify the neural controller against the uncertain environment to aid the arm in reaching the target, while bootstrapping the process of exploration and learning.

To be effective, these kinematic constraints need not be required for the full length or duration of the movement, dic-

Table 2. *Statistics for experiment B*

Trajectory	Mean, m	SD, m	*P*	Significance	Success Rate
Baseline	−0.0021	0.0036			58.0%
Ramp Trials [*1–10*]	0.0011	0.0047	0.1125		59.0%
Ramp Trials [*11–20*]	0.0025	0.0058	0.0511		34.0%
Force Trials [*1–10*]	0.0019	0.0062	0.1015		23.0%
Force Trials [*11–20*]	4.5×10^{-4}	0.0056	0.2498		34.0%
Final Force	0.0015	0.0048	0.0768		56.0%
After Effects	−0.0117	0.0017	4.0×10^{-6}	*	24.0%

The trajectories Baseline, Before Effects, Final Force, and After Effects are as defined and plotted in Fig. 4, B.DATA. Ramp Trials [*1–10*] and Ramp Trials [*11–20*] (plotted as thin lines in Fig. 4, B.DATA) are the first and last 10 trails, respectively, during the ramp stage where the force field gain gradually increases from trial to trial. Force Trials [*1–10*] and Force Trials [*11–20*] (also thin lines in Fig. 4, B.DATA) are the first 10 and next 10 trials immediately following the ramp stage (with force field at full strength). Mean and SD report the mean trajectory error and SD (respectively) for the 10 subjects of *experiment B*. *P* and significance report the significance from baseline (paired *t*-test, degrees of freedom = 9, $*P < 0.01$). Success rates for each trajectory are also shown.

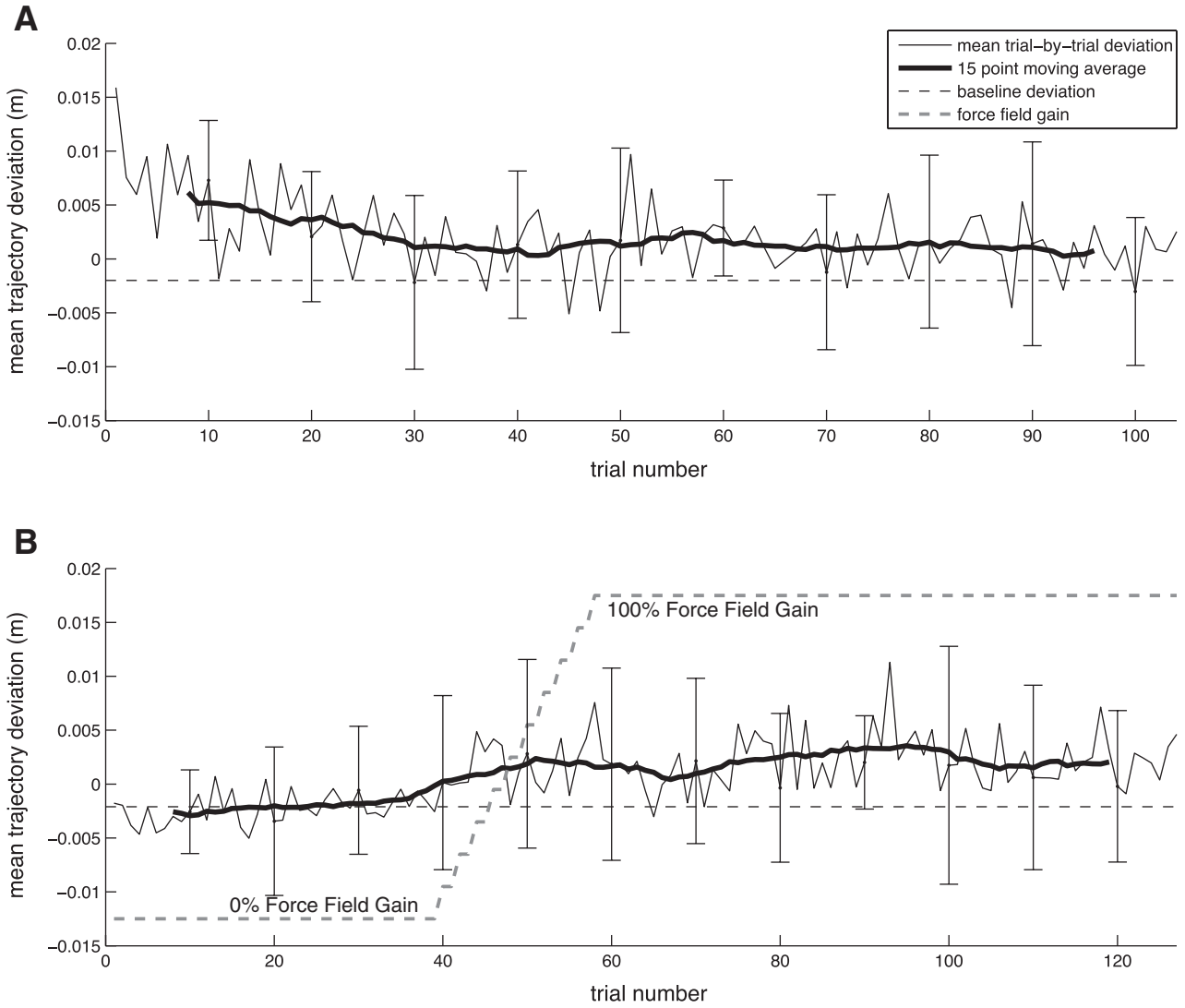

Fig. 5. *A*: averaged trial-by-trial learning curve during *block 3* of *experiment A* (the first block of force field training). *B*: averaged trial-by-trial learning curve during *block 2* of *experiment B*, which includes the ramp stage of gradually increasing the force field gain from zero (indicated by rising gray dashed line). In this case, the trial sequences of each subject are aligned by the start of the ramp stage. The *y*-axis in the plots is the average distance from the line connecting the start of the trial to the target. Baseline value is also shown with a dashed line. For selected trials, error bars indicate plus and minus one standard deviation. Thick line is a 15-point moving average.

tating a complete trajectory from start to goal. A vectorial position and velocity constraint, affecting only the initial part of the movement, is consistent with the idea that reaches are vectorially planned (Bullock and Grossberg 1989; Bock and Arnold 1992; Gordon et al. 1994; Ghilardi et al. 1995; Krakauer et al. 2000), i.e., planned in a hand centered coordinate system, with a desired initial direction of movement. For example, (Gordon et al. 1994) observed that directional variability at the start of a movement was constant over several directions and independent of distance, speed, and inertial properties of the limb, concluding that the directional variability is mostly due to errors in the planning process. The work of (Ghilardi et al. 1995) also showed that hand trajectories had systematic directional biases, in the absence of vision, even when hands where repositioned to a new workspace, and (Ghez et al. 2007) showed directional biases in slicing task (out and back in one continuous motion) when initial hand positions were covertly altered by a robotic manipulandum.

Our model of directional preference also includes a time-dependent decay. Although a constant, nondecaying, directional constraint term would achieve qualitatively the same outcome (results not shown), including the decay stresses that the directional constraint has significant influence at the start of the movement. Reduction of the constraint during motion is consistent with the notion that initial trajectory and end point accuracy are controlled sequentially (Scheidt and Ghez 2007). In that work, subjects who trained on visuomotor rotations with slicing movements did not transfer learning to a reaching task (and visa versa). They suggested that separate control mechanisms may exist for initial trajectory and final posture, rather than a complete spatial plan from start to goal. In their nonlinear model using inverse dynamics, the two controllers act sequentially with onset of the posture controller occurring at the time of peak acceleration. Motivated by this idea, we set the time constant of the decay term to achieve 90% decay at approximately the time of peak acceleration. Our results demonstrate how the concept of "sequential" control can

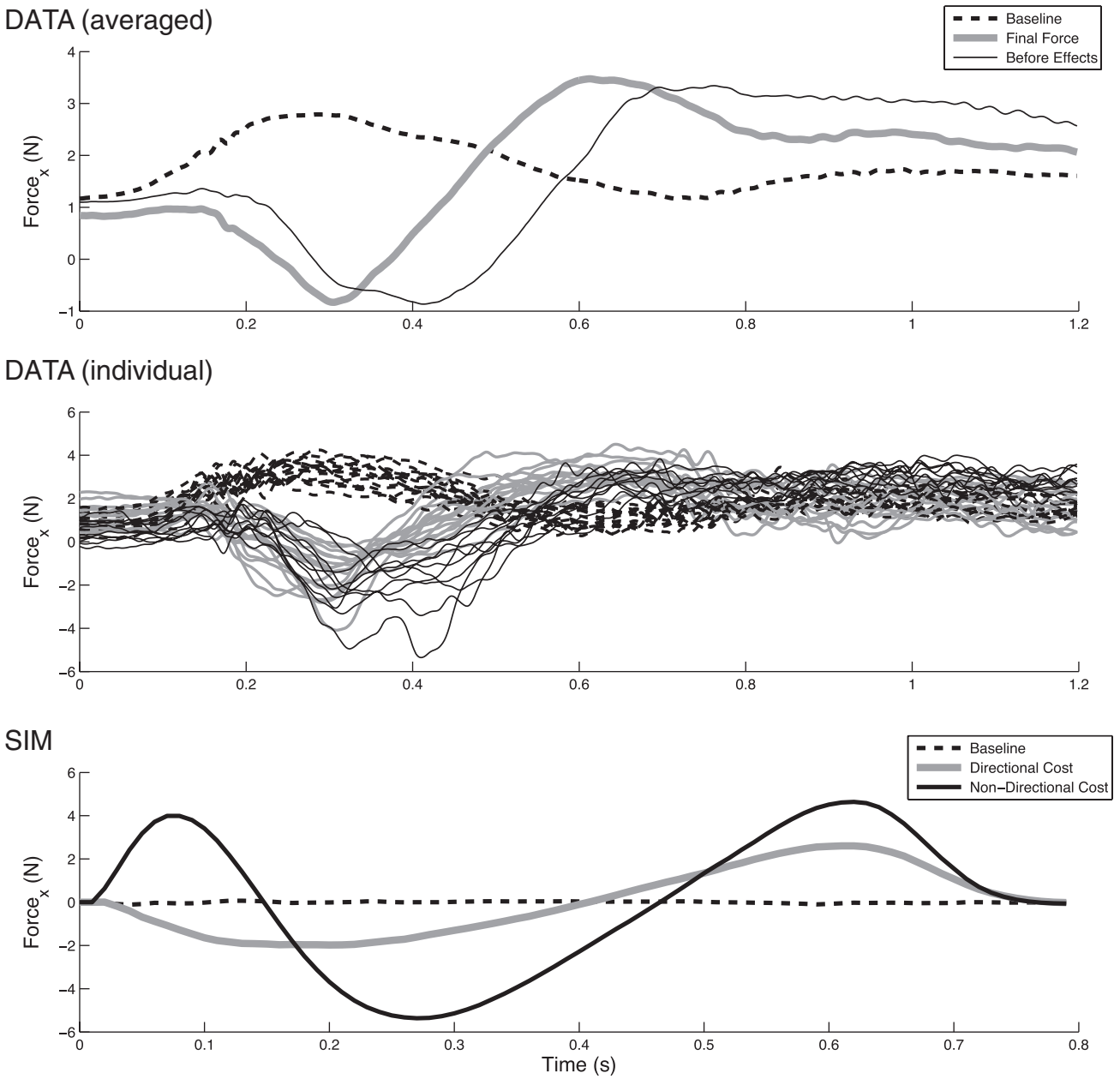

Fig. 6. DATA: forces in the *x* direction (the direction of the force field perturbation) as measured by the force sensor at the handle of the manipulandum. DATA (averaged) shows the average among all 6 subjects of *experiment A*. DATA (individual) shows individual trials from a single subject of *experiment A*. Baseline trajectories (dashed lines) are the last 15 trials in null field training. Final Force (thick solid lines) are the last 15 trials in force field training. Before Effects (thin solid lines) are the 10 catch trials following null field training. While the baseline force profile has a rightward (positive) force (likely due to subjects compensating for manipulandum dynamics), the force field disturbance elicits a strong corrective force in the leftward (negative) direction in the early part of the movement. SIM: output force, in the *x* direction, of the simulated optimal controller under the full strength force field (*experiment A*). The model with nondirectional cost function initially pushes in the same direction as the disturbance, while the directional model applies similar resistive forces as demonstrated by the subjects.

be encapsulated within optimal control theory, by trading off requirements in a cost function, instead of relying on a switching or blending of separate controllers.

Finally, there are certain physiological limitations that may inhibit the process of learning a novel dynamics environment. Evidence suggests that internal model formation occurs locally and does not generalize to untrained areas of the workspace for both dynamic perturbations (Gandolfo et al. 1996) and visuomotor transformations (Krakauer et al. 2000). Thus, for complete learning, the workspace must be exhaustively explored, which can be time consuming, energetically expensive, and risky. We also wish to learn to make energy-efficient movements; however, our sensory system has generally poor feedback of instantaneous motor cost. Feedback of energy consumption can be substantially delayed, perhaps on the time course of hours or days, in terms of muscle fatigue or exhaustion, e.g., pain in response to lactic acid (Allen et al. 2008). In the short term, however, a perceived deviation from a kinematic invariant can be relied on as an indication that something has gone awry with the current internal model of the dynamical

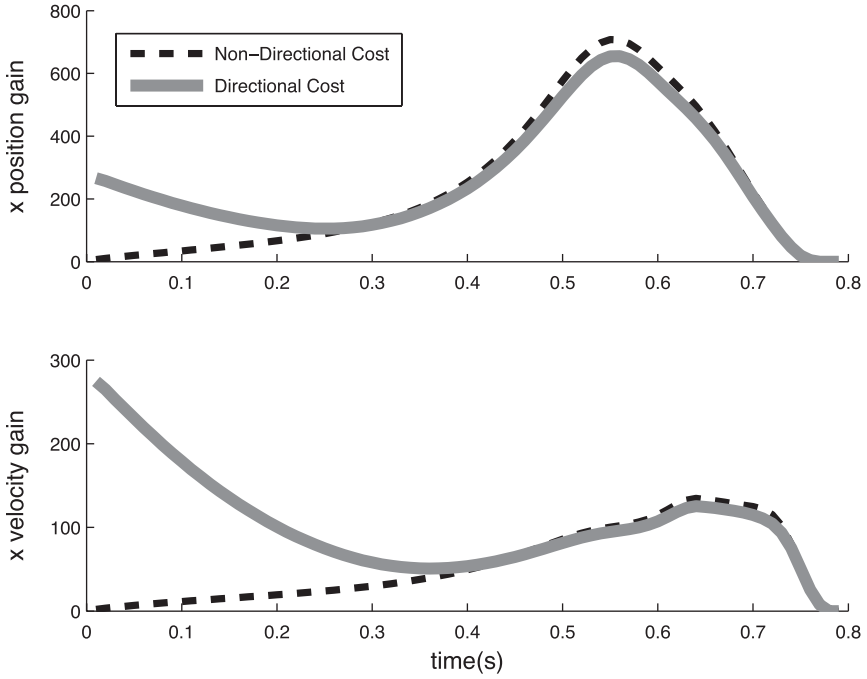

Fig. 7. Position and velocity gains of the optimal controller, in the x direction (same direction as the force perturbation) before and after addition of a directional constraint to the cost function. Adding the directional constraint effectively increases stiffness and damping during the early part of the movement.

system. Then, a strategy that systematically attempts to return the hand to a previously observed solution [perhaps utilizing *motor memory* (Ganesh et al. 2010)] allows for sufficient error information to be collected for an internal model update (Kawato and Gomi 1992). Once a reliable, local internal model is formed, and the ability to reach the target is restored, the CNS can then turn its efforts towards energy optimization.

In conclusion, our work addresses the intrinsic objective of a point-to-point reaching movement. While there are notable arguments made for the parsimonious case of only terminal accuracy and minimal motor effort, we suggest that robustness, especially against model uncertainly, should additionally be considered essential for reaching. A kinematic invariant, independent of the underlying dynamics, helps the CNS mitigate

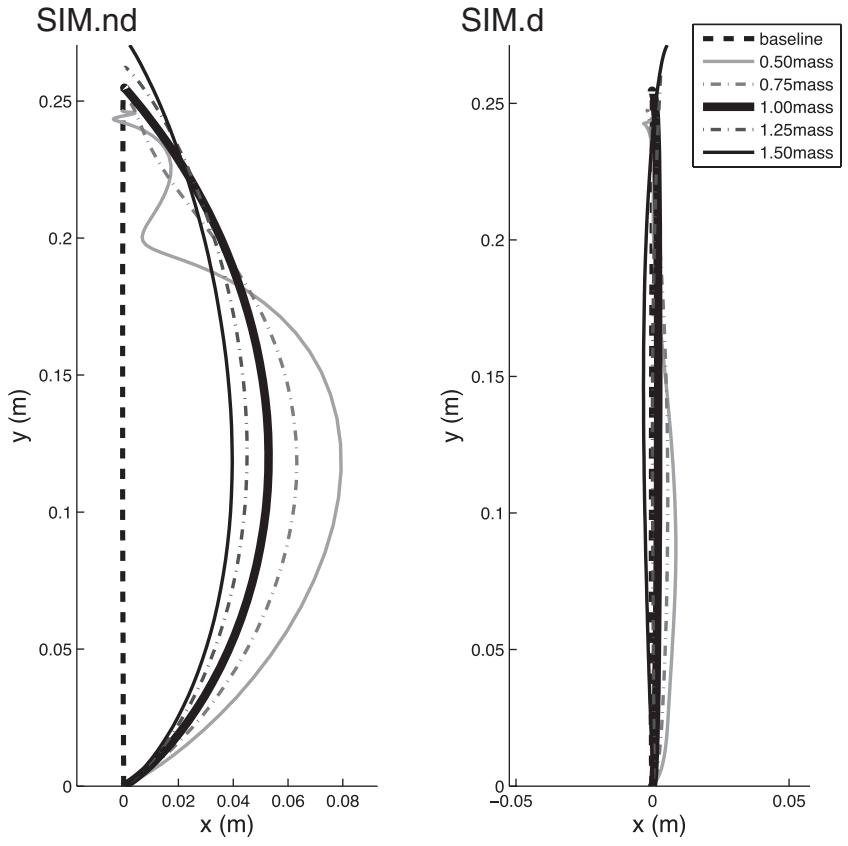

Fig. 8. Sensitivity of the controller with respect to unknown changes in arm mass. In each case, the optimal controller is computed assuming the nominal mass value, but tested on 50, 75, 100, 125, and 150% of the mass value. SIM.nd uses the nondirectional cost function, and SIM.d uses directional cost.

the risks of uncertainty while enabling the eventual learning of a more efficient control solution.

ACKNOWLEDGMENTS

We thank Gary Liaw for contributions during initial pilot studies of this experiment and Toshinori Yoshioka for implementation and operation of the experimental equipment.

GRANTS

This research was supported in part by National Science Foundation Grants ECS-0326095, IIS-0535282, IIS-1017134, CNS-0619937, IIS-0917318, CBET-0922784, EECS-0926052, CNS-0960061, DARPA program on Advanced Robotic Manipulation, Army Research Office, Okawa Foundation, and Strategic Research Program for Brain Sciences "Brain Machine Interface Development" by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: M. M, E. T., S. S., and M. K. conception and design of research; M. M. performed experiments; M. M. analyzed data; M. M, E. T., S. S., and M. K. interpreted results of experiments; M. M, E. T., S. S., and M. K. prepared figures; M. M. and E.T. drafted manuscript; M. M, E. T., S. S., and M. K. edited and revised manuscript; M. M, E. T., S. S., and M. K. approved final version of manuscript.

REFERENCES

Allen DG, Lamb GD, Westerblad H. Skeletal muscle fatigue: Cellular mechanisms. *Physiol Rev* 88: 287–332, 2008.

Arce F, Novick I, Shahar M, Link Y, Ghez C, Vaadia E. Differences in context and feedback result in different trajectories and adaptation strategies in reaching. *PLoS One* 4: e4214, 2009.

Bock O, Arnold K. Motor control prior to movement onset: preparatory mechanisms for pointing at visual targets. *Exp Brain Res* 90: 209–216, 1992.

Bullock D, Grossberg S. Vite and flete: neural modules for trajectory formation and postural control. *Volitional Action*, edited by Hershberger WA. Amsterdam, The Netherlands: North Holland, 1989, p. 253–297, 1989.

Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. *J Neurosci* 5: 1688–1703, 1985.

Franklin DW, Burdet E, Tee KP, Osu R, Chew CM, Milner TE, Kawato M. CNS learns stable, accurate, and efficient movements using a simple algorithm. *J Neurosci* 28: 11165–11173, 2008.

Gandolfo F, Mussa-Ivaldi FA, Bizzi E. Motor learning by field approximation. *Proc Natl Acad Sci USA* 93: 3843–3846, 1996.

Ganesh G, Haruno M, Kawato M, Burdet E. Motor memory and local minimization of error and effort, not global optimization, determine motor behavior. *J Neurophysiol* 104: 382–390, 2010.

Ghez C, Scheidt RA, Heijink H. Different learned coordinate frames for planning trajectories and final positions in reaching. *J Neurophysiol* 98: 3614–3626, 2007.

Ghilardi MF, Gordon J, Ghez C. Learning a visuomotor transformation in a local area of work space produces directional biases in other areas. *J Neurophysiol* 73: 2535–2539, 1995.

Goble JA, Yanxin Z, Shimansky Y, Sharma, S, Dounskaia, NV. Directional biases reveal utilization of arm's biomechanical properties for optimization of motor behavior. *J Neurophysiol* 98: 1240–1252, 2007.

Gomi H, Kawato M. Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science* 272: 117–120, 1996.

Gordon J, Ghilardi MF, Ghez C. Accuracy of planar reaching movements. I. Independence of direction and extent variability. *Exp Brain Res* 99: 97–111, 1994.

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

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

Kawato M, Gomi H. A computational model of four regions of the cerebellum based on feedback-error learning. *Biol Cybern* 68: 95–103, 1992.

Kistemaker DA, Wong JD, Gribble PL. The central nervous system does not minimize energy cost in arm movements. *J Neurophysiol* 104: 2985–2994, 2010.

Krakauer JW, Pine ZM, Ghilardi MF, Ghez C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci* 20: 8916–8924, 2000.

Lackner JR. DiZio P. Rapid adaptation to coriolis force perturbations of arm trajectory. *J Neurophysiol* 72: 299–313, 1994.

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

Ohta K, Svinin MM, Luo Z, Hosoe S, Laboissière R. Optimal trajectory formation of constrained human arm reaching movements. *Biol Cybern* 91: 23–36, 2004.

Scheidt RA, Ghez C Separate adaptive mechanisms for controlling trajectory and final position in reaching. *J Neurophysiol* 98: 3600–3613, 2007.

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

Takahashi CD, Robert Scheidt A, Reinkensmeyer DJ. Impedance control and internal model formation when reaching in a randomly varying dynamical environment. *J Neurophysiol* 86: 1047–1051, 2011.

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

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

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

Wolpert DM, Ghahramani Z, Jordan MI. Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study. *Exp Brain Res* 103: 460–470, 1995.