Brief Communications

Near-Optimal Human Adaptive Control across Different Noise Environments

Manu Chhabra and Robert A. Jacobs
Departments of Computer Science and Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14627

A person learning to control a complex system needs to learn about both the dynamics and the noise of the system. We evaluated human subjects' abilities to learn to control a stochastic dynamic system under different noise conditions. These conditions were created by corrupting the forces applied to the system with noise whose magnitudes were either proportional or inversely proportional to the sizes of subjects' control signals. We also used dynamic programming to calculate the mathematically optimal control laws of an "ideal actor" for each noise condition. The results suggest that people learned control strategies tailored to the specific noise characteristics of their training conditions. In particular, as predicted by the ideal actors, they learned to use smaller control signals when forces were corrupted by proportional noise and to use larger signals when forces were corrupted by inversely proportional noise, thereby achieving levels of performance near the information-theoretic upper bounds. We conclude that subjects learned to behave in a near-optimal manner, meaning that they learned to efficiently use all available information to plan and execute control policies that maximized performances on their tasks.

Key words: **visuomotor control; motor learning; human; performance; movement; ideal actor analysis**

Introduction

A person learning to control a complex system needs to learn about both the dynamics and the noise of the system (Shadmehr and Mussa-Ivaldi, 1994; Wolpert et al., 1995; Baddeley et al., 2003). When learning about the dynamics, the person learns about the relationships between control signals and the expected responses to these signals (Miall and Wolpert, 1996; Flanagan and Wing, 1997; Johansson, 1998; Desmurget and Grafton, 2000). A person learning to hit a hockey puck toward a teammate, for example, learns about the mapping from the state of their body and the commands sent to their muscles to the expected trajectory of the puck. In contrast, when learning about the noise, the person learns about the relationships between control signals and the expected variances in the responses to these signals (Wolpert et al., 1995; Baddeley et al., 2003). For example, a hockey player may learn that large muscle commands lead to a distribution of puck trajectories with a large variance, whereas small commands lead to a distribution with a small variance.

An approach to the study of human adaptive control that has recently yielded novel insights is to compare human performances with the mathematically optimal performances of "ideal actors" (Busemeyer, 2002; Jagacinski and Flach, 2003; Todorov, 2004). A motivation for this approach can be found in the work of Marr (1982). Marr defined three levels of analysis of an information processing device. The top level, known as the computational theory, examines what the device does and why. A distinguishing feature of this level is that it provides an explanation for why a device does what it does by studying the goals of the device. An important property of research that uses ideal actors is that this research formalizes goals as mathematical constraints or criteria, searches for behaviors that optimize the criteria, and compares the optimal behaviors with human behaviors. If there is a close match, then it is hypothesized that people are behaving as they do because they are efficiently satisfying the same goals as were built into the ideal actor.

This article reports research using ideal actors to study people's abilities to learn to control a dynamic system in an optimal manner. Different groups of subjects learned to control a system under different noise conditions. The comparisons of subjects' performances with the performances of the ideal actors for each condition provided at least two advantages. First, because optimal performances depended on the dynamics of the system as well as its noise, people needed to learn about both the dynamics and the noise to behave optimally. If subjects did indeed behave optimally, then we can conclude that they learned to efficiently use all available information to plan and execute control policies that maximized performance on the experimental task. Second, the comparisons of subjects' performances with the ideal actors' performances allowed us to understand subjects' behaviors in a more quantitative and mathematically rigorous manner than would otherwise have been the case.

Materials and Methods

Subjects. Subjects were students at the University of Rochester with normal or corrected-to-normal vision and with normal motor abilities. Subjects were naive to the purposes of the study.

Stimuli and procedures. Our experiments required subjects to control a simulated device using a computer mouse. Experiments requiring subjects to control simulated dynamical systems via human–computer interfaces are increasingly widespread in the literature because they allow researchers to easily and precisely control the properties of the dynamical systems (Foulkes and Miall, 2000; Robles-de-la-Torre and Sekuler, 2004;

Received March 7, 2006; revised Aug. 17, 2006; accepted Sept. 13, 2006.

This work was supported by research grants from the National Institutes of Health and the Air Force Office of Scientific Research. We thank D. Knill for many thoughtful conversations and R. Shadmehr and an anonymous reviewer for their helpful comments on a previous version of this manuscript.

Correspondence should be addressed to Robert A. Jacobs at the above address. E-mail: robbie@bcs.rochester.edu.

DOI:10.1523/JNEUROSCI.2238-06.2006

 0270-6474/06/2610883-05$15.00/0

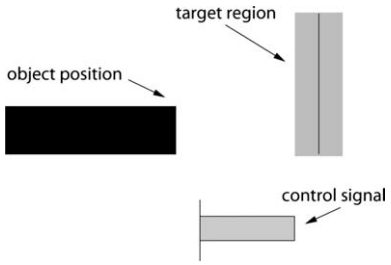

Figure 1. Schematic of a visual display illustrating the current position of the object, the current value of the control signal, and the position of the target region.

Krakauer et al., 2005). For our purposes, the use of simulated dynamic systems controlled through a human–computer interface provided two advantages: because of the flexibility of simulation, it allowed us to easily test subjects in a variety of noise conditions, and it allowed us to constrain the experimental environment so that we could construct mathematical models known as ideal actors that performed the same tasks as human subjects.

The experimental task was to use the computer mouse to apply forces to a simulated object so that it spends as much time as possible in a target region. The object was constrained to move along a horizontal line. The object can be characterized as a second-order dynamical system: $m\ddot{x} = f - b\dot{x}$, where x, $\dot{x}$, and $\ddot{x}$ are the position, velocity, and acceleration, respectively, of the object, m is its mass, b is its viscous resistance, and f is the force applied to the object. In our simulations, we set $m = 10$ units and $b = 5$ force units second/spatial unit (in which the spatial units of the workspace ranged from -50 to 50). Subjects selected control signals by setting the horizontal position of the computer mouse. This position was sampled at a rate of 30 Hz. The visual display (subtending ~40° of visual angle in the horizontal dimension) provided the current position of the object, the current value of the control signal, and the position of the target region (Fig. 1). The position of the object was given by the position of the rightmost edge of a black horizontal bar (the leftmost edge of the bar was always adjacent to the leftmost edge of the workspace). The value of the control signal was given by a red horizontal bar in which the bar extended to the left or right of the center of the workspace to indicate the sign of the signal (either negative or positive), and the length of the bar indicated the magnitude of the signal. The target region was denoted by a gray vertical bar (this region subtended ~1.6° of visual angle in the horizontal dimension). Subjects sat ~50 cm from the computer monitor and viewed the monitor at a comfortable angle.

Each trial lasted 10 s, and subjects performed 180 trials during a 1 h experimental session. At the start of each trial, the positions of the object and target regions were set to random values such that the target region was to the right of the object. If the horizontal dimension of the workspace had coordinates between -50 and 50, then the initial location of the object was sampled from a uniform distribution between -36 and -13, and the center of the reward region was sampled from a uniform distribution between 0 and 23. At the end of each trial, subjects received feedback in the form of a score equal to the percentage of time during the trial that the object was in the target region.

The experiment included three experimental conditions. In the no-noise (NN) condition, the force applied to the object was a deterministic linear function of the horizontal position of the computer mouse. If the horizontal dimension of the workspace had coordinates between -50 and 50, then the mouse position r was an integer in this range, and the subject's control signal $u(r)$ was set to $15 \times r$. In the NN condition, the force was set to the control signal: $f = u(r)$. In the proportional-noise (PN) condition, the force was a stochastic function of the control signal: $f = u(r) + \varepsilon$, where the noise ε was sampled from a normal distribution whose mean was 0 and whose SD was proportional to the control signal [$\mathrm{SD} = k \times u(r)$, where $k = 1.5$]. Last, the inversely proportional-noise (IPN) condition was identical to the PN condition, with the exception that the SD of the noise was inversely proportional to the control signal [$\mathrm{SD} = k/15^{0.75}$ if $u(r) < 15$, and $\mathrm{SD} = k/u(r)^{0.75}$ if $u(r) \geq 15$, where $k = 1500$].

The experimental task was relatively difficult for at least two reasons. First, it required subjects to learn the dynamics governing the movement of the object. This aspect of the task was challenging, particularly because we set the viscous resistance to a small value. For example, because of the small value of this resistance, the application of a positive force to the object caused the object to move rightward even after this force had been set to 0. Second, it required subjects to learn about the stochastic relationships between subjects' control signals and the forces applied to the object.

Design of the ideal actors. As mentioned above, it is often useful to compare human performances on perceptual or motor tasks with the performances of "ideal observers" or ideal actors (Barlow, 1957; Geisler, 1989; Baddeley et al., 2003; Todorov, 2004; Berthier et al., 2005). Because these computational devices optimally use all available information when performing a task, their performances serve as "gold standards" against which human performances can be benchmarked.

We calculated the performances of ideal actors on our adaptive control task. For every trial performed by every subject, we determined the optimal feedback control law of an ideal actor. The ideal actor was provided with knowledge of the dynamics of the object, the noise characteristics of the environment, and the task requirements. It was also provided with the same initial position of the object and location of the target region as seen by a subject on a given trial. Using this information, the ideal actor performed the same task as a subject, i.e., it searched for an optimal control law. This law is useful because it provides an action for each state of a system that maximizes the expected amount of time that the object spends in the target region (Sutton and Barto, 1998; Engelbrecht et al., 2003; Robles-de-la-Torre and Sekuler, 2004). We emphasize that the optimal control law depends on the noise characteristics of the environment, meaning that different ideal actors were created for different noise conditions.

Mathematically, the experimental task can be formalized as a Markov decision task whose optimal solutions were found via dynamic programming (Bellman, 1957; Bertsekas and Tsitsiklis, 1996). To solve the optimal control problem, we first discretized the continuous environment into a discrete state and action space. The position of the object x was discretized into 100 equally sized bins between -50 and 50. The velocity of the object $\dot{x}$ was discretized into 100 equally sized bins between -50 and 50. The control signal was already a discrete variable (bin size, 15). It was constrained to be between -750 and 750.

We then transformed the optimal control problem to a Markov decision problem (MDP) on this discretized control space. The state space S of the Markov decision problem is defined as $S = \{s : s = (x, \dot{x})\}$, where x and $\dot{x}$ are bin centers as defined above. The action space A is the set $\{-735, -720 \ldots 750\}$. The next-state distribution of the MDP [i.e., $P(\cdot|s, a)$, $s \in S$ and $a \in A$] depends on the dynamics given in the above equation and the nature of the noise. Because the system is linear and because the noise in all experimental conditions has a normal distribution, the next-state distribution is also a normal distribution. The reward function for the MDP was defined as $R(s_t) = 1$ if x was in the target region at time t, and $R(s_t) = 0$ otherwise. Given the next-state distribution and the reward function, the MDP can be formulated as a dynamic programming problem whose solution is a mapping from states to actions such that the action chosen at each state maximizes the expected sum of rewards, denoted $E[\Sigma_t R(s_t)]$, where $E[\cdot]$ is the expectation operator, t is an index over time steps, and the sum is taken over all time steps of a trial. Note that we used dynamic programming to find globally optimal solutions to the optimal control problem (as opposed to approximate solutions that are typically found by methods from the machine learning literature, such as reinforcement learning methods). In addition, a solution to the optimal control problem is not unique (i.e., there are many control policies that maximize performance on the experimental task), and dynamic programming can be used to find multiple solutions.

As discussed below, we found that subjects learned to perform nearly as well as ideal actors in all experimental conditions. There are many possible reasons why subjects did not perform exactly as well as the ideal actors and, thus, did not achieve statistical efficiencies of 1. Some reasons are relevant to subjects' motor learning processes. For example, subjects may have performed suboptimally because they learned imperfect internal models of the system they needed to control whereas an ideal actor was provided with a perfect model of the system, or because subjects inaccurately estimated the values of the state variables whereas an ideal actor was provided with perfect state observations. Other reasons are not

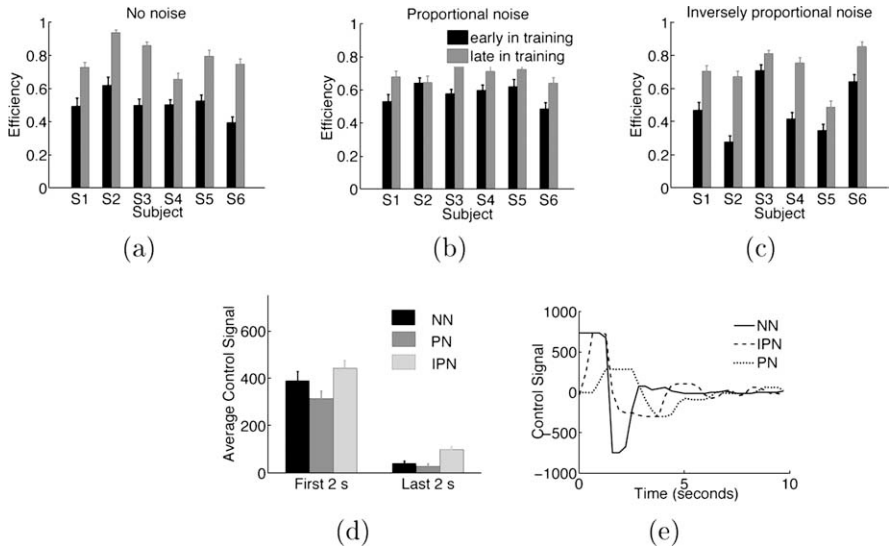

Figure 2. ***a–c***, Subjects' efficiencies in the NN, PN, and IPN conditions, respectively. Efficiency on an individual trial is defined as the ratio of a subject's performance (the amount of time that the object spent in the target region) to the average performance of the ideal actor. Black bars indicate a subject's efficiency during the first 30 trials of a session, whereas gray bars indicate the efficiency during the last 30 trials (error bars give the SEMs). In the NN condition, six of six subjects showed significant improvement in efficiency (two-tailed t test, $p < 0.01$), in the PN condition, five of six subjects showed significant improvement ($p < 0.03$), and, in the IPN condition, six of six subjects showed significant improvement ($p < 0.03$). ***d***, Subjects' average control signals (based on the magnitude of these signals, not their signs) during the first and last 2 s of the final 30 trials (error bars show SEM). ***e***, Typical control strategies during the last 30 trials of randomly selected subjects in the NN, PN, and IPN conditions.

relevant to subjects' learning processes. For example, an ideal actor could choose a novel control signal at each moment in time although subjects' control signals were necessarily correlated at neighboring time steps, or an ideal actor did not include temporal lags although subjects' motor and cognitive processing contained lags. For these reasons, our ideal actors provided information-theoretic upper bounds on performances. Designing ideal actors that include motor, perceptual, and cognitive limitations resembling those of people is an important area of future research.

Results

Figure 2*a–c* shows subjects' efficiencies in the NN, PN, and IPN conditions, respectively, in which efficiency is defined as the ratio of a subject's performance (the amount of time that the object spent in the target region) to the expected performance of the ideal actor (this expectation was taken with respect to the probability of the noise in the PN and IPN conditions and was computed via Monte Carlo simulation). Seventeen of 18 subjects showed significant learning during the course of the experimental session (two-tailed t test, $p < 0.03$). In addition, subjects achieved high statistical efficiencies in all conditions at the end of the session.

To evaluate whether subjects learned control strategies tailored to the specific noise characteristics of each experimental condition, we examined their average control signals (using the magnitudes of these signals, not their signs) during the first 2 s (mostly reflecting subjects' planning or feedforward control policies) and last 2 s (mostly reflecting subjects' feedback policies) of each of the last 30 trials. The left side of Figure 2*d* shows that, on average, subjects in the PN condition used significantly smaller control signals during the first 2 s than subjects in the NN or IPN conditions (based on a two- tailed t test, the difference between the PN and IPN groups is statistically significant at $p < 0.05$ level). This result indicates that subjects in the PN condition learned that large control signals lead to large amounts of noise corrupting the forces applied to the object. Consequently, they avoided large control signals. The right side of Figure 2*d* shows that subjects in the IPN condition used significantly larger control signals during the last 2 s than subjects in the NN or IPN conditions (based on a two-tailed t test, the difference between the PN and IPN groups is statistically significant at the $p < 0.01$ level). This result suggests that subjects in the IPN condition learned that small control signals lead to large amounts of noise and, thus, they avoided small control signals.

We claim that subjects in the PN condition learned to use smaller control signals and that subjects in the IPN condition used larger signals, because subjects learned control strategies that were tailored to the specific noise characteristics of their training conditions. An alternative explanation that is a logical, albeit unlikely, possibility is that subjects in the IPN condition used larger control signals because they had larger errors. That is, subjects in the IPN condition may have used large control signals because the object was often far from the target region. To evaluate this possibility, we measured subjects' average errors [the average difference between the position of the object and the position of the nearest edge of the target region (using magnitude only, not the sign of the difference); this difference was set to 0 if the object was inside the target region]. Contrary to the hypothesis being considered here, subjects in the IPN condition actually had smaller errors than subjects in the PN condition during the first 2 s of a trial (based on a two-tailed t test, the difference between the two groups is statistically significant at the $p < 0.05$ level). During the last 2 s of a trial, subjects in the IPN condition again had smaller errors, although the difference between the two groups was not statistically significant. We conclude that it is not the case that subjects in the IPN condition used larger control signals because they had larger errors.

Figure 2*e* shows typical control strategies during the last 30 trials of randomly selected subjects in the NN, PN, and IPN conditions. The subject in the NN condition learned to initially provide a large positive control signal, thereby accelerating the object to the right toward the target region, followed by a large negative control signal, thereby decelerating the object so that it came to rest in or near the target region. This subject's strategy can be characterized as an approximation to what is known in the engineering literature as a "bang–bang" control policy. The control strategy of the subject in the PN condition is an approximation to a bang–bang policy that has been smoothed to avoid control signals with large magnitudes. Because large control signals lead to large amounts of noise corrupting the forces applied to the object, it makes sense that the subject chose to avoid large signals. The strategy of the subject in the IPN condition can also be characterized as a bang–bang policy, but it differs from that of the subject in the NN condition in that the control signals toward the end of the trial differed significantly from 0 although the object is in or near the target region. It seems that the subject learned that small signals lead to large amounts of noise and, thus, avoided very small signals.

The data in Figure 2 suggest that subjects learned near-optimal control strategies. To more directly evaluate this hypothesis, we plotted the optimal control policies of the ideal actors, as well as the subjects' control policies. Figure 3, *a* and *b*, shows the average control policies of the ideal actors in the PN and IPN conditions, respec-

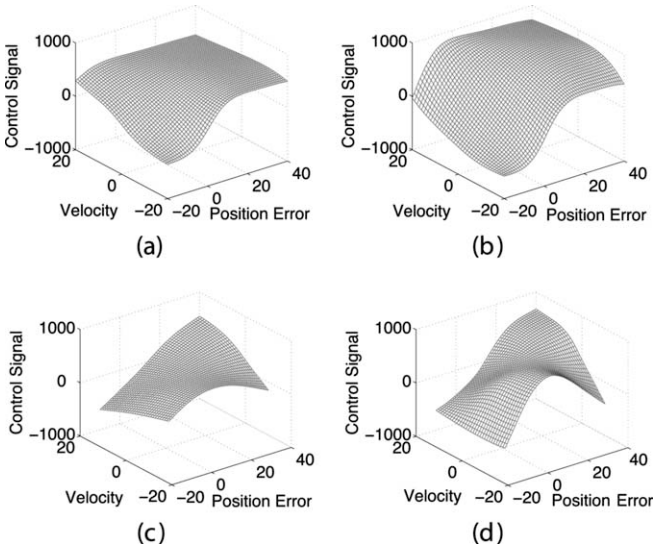

Figure 3. ***a***, ***b***, Average optimal control policies for the PN and IPN conditions, respectively. ***c***, ***d***, Subjects' average control policies in the PN and IPN conditions, respectively.

tively. [Recall that a control policy is a mapping from the position error ($x_T - x$: the position of the center of the target region minus the position of the object) and the velocity of an object to a control signal.] The data in these graphs were formed as follows. For each noise condition, we first found all optimal policies (i.e., all policies that maximize the expected amount of time that the object spends in the target region). We then smoothed each optimal policy through Denauley triangulation, followed by linear interpolation, and, finally, averaged these smoothed policies. These graphs illustrate the fact that optimal policies depend on the noise characteristics of the environment. For example, optimal policies in the PN condition tend to use smaller control signals than optimal policies in the IPN condition.

Figure 3, *c* and *d*, shows subjects' control policies in the PN and IPN conditions, respectively. For each subject, we collected this information from a subject's last 30 trials. Identical to Figure 3, *a* and *b*, a smooth surface was formed from the data through Denauley triangulation, followed by linear interpolation. This was done for each subject in an experimental condition, and then the surfaces for all subjects within a condition were averaged. Clearly, subjects in the PN and IPN conditions learned different policies: subjects in the PN condition tended to use smaller (in magnitude) control signals than subjects in the IPN condition, and their control policies tended to be smoother.

We also evaluated the complexity of subjects' acquired control strategies by examining the amount of variability in these strategies that could be accounted for by a linear controller known in the engineering literature as a proportional-derivative (PD) controller. A PD controller is defined as $u = K_1(x - x_T) + K_2\dot{x}$, where x and $\dot{x}$ are the position and velocity of the object, x_T is the center of the target region, K_1 and K_2 are gain parameters, and u is the control signal. Using a subject's last 10 trials, we fit a PD controller (i.e., we determined values for K_1 and K_2) by linear regression, where $x - x_T$ and $\dot{x}$ are the independent variables, and u is the dependent variable.

We first asked whether each subject's acquired policy could be accounted for by a fixed PD controller. On average, the PD controller accounted for 42% of the variability in a subject's control signal (SD of 8%). We next asked whether the fit could be improved by using different PD controllers for the initial portion (reflecting a subject's planning and feedforward control policy) and final portion

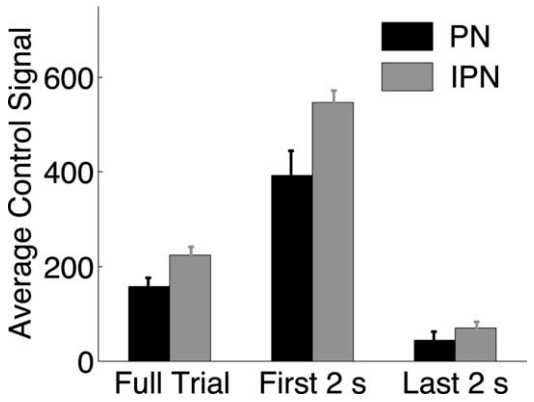

Figure 4. Subjects' average control signals (based on the magnitude of these signals, not their signs) from the PN and IPN groups during an entire catch trial or during the first or last 2 s of a catch trial (error bars show SEM).

(reflecting a subject's feedback policy) of each trial. When we fit a PD controller to the data from the first 2 s of a subject's trials, the controller accounted for 27% of the variability in a subject's control signal on average (SD of 12%). When we fit a PD controller to the data from the last 2 s, the controller accounted for 32% of the variability in a subject's control signal on average (SD of 18%). We conclude that PD controllers do not provide good accounts of subjects' acquired control strategies. Moreover, it seems unlikely that any linear controller will provide a good account of this data, suggesting that subjects acquired nonlinear control policies.

Finally, we further evaluated the claim discussed above that subjects in the PN condition learned to use smaller control signals and that subjects in the IPN condition used larger signals, by running an additional experiment. Five subjects were run in the PN condition and five subjects were run in the IPN condition, in which each subject performed four blocks of trials in which each block consisted of 60 trials. At random, 10 trials of the final block were selected to serve as "catch" trials. On a catch trial, the initial position of the object and the center position of the target region were always set to fixed values (-27.5 and 12.5, respectively, in which the horizontal dimension of the workspace had coordinates from -50 to 50), and the forces applied to the object were not corrupted by noise. Consequently, catch trials had the same properties for all subjects, and differences in performances on catch trials can be attributed to differences in subjects' training histories. Figure 4 shows subjects' average control signals (using the magnitudes of the signals, not signs) during an entire catch trial and during the first 2 s or last 2 s of a catch trial. Subjects trained in the PN condition consistently used smaller control signals than subjects trained in the IPN condition (based on two-tailed t tests, the differences between the PN and IPN groups are statistically significant at the $p < 0.05$ level when considering an entire catch trial and when considering the first 2 s of a catch trial; the difference is not significant when considering the last 2 s of a catch trial). Clearly, subjects learned control strategies tailored to the specific noise characteristics of their training conditions: subjects in the PN group used smaller control signals, indicating that they learned that large signals lead to large amounts of noise corrupting the forces, whereas subjects in the IPN group used larger control signals, indicating that they learned that small signals lead to large amounts of noise corrupting the forces.

Discussion

This article has reported research evaluating human subjects' abilities to learn to control a stochastic dynamic system when the

forces applied to the system were corrupted with noise whose magnitudes were either proportional or inversely proportional to the sizes of subjects' control signals. We also used dynamic programming to calculate the mathematically optimal control laws of an ideal actor for each noise condition. The results suggest that people learned control strategies tailored to the specific noise characteristics of their training conditions. In particular, they learned to use smaller control signals when forces were corrupted by proportional noise and to use larger signals when forces were corrupted by inversely proportional noise. These control strategies allowed subjects to achieve levels of performance near the information-theoretic upper bounds. We conclude that subjects learned to behave in a near-optimal manner, meaning that they learned to efficiently use all available information to plan and execute control policies that maximized performances on their tasks.

These results suggest several issues that will need to be addressed in future research. First, our results indicate that subjects were effective learners regardless of whether they were trained in the PN or IPN condition. This might be regarded as a surprising result. The PN condition closely resembles biological motor noise [recent data and theories suggest that noise in biological motor systems scales with the size of the control signal (Clamann, 1969; Matthews, 1996; Harris and Wolpert, 1998)], whereas the IPN condition does not. Therefore, in some sense, subjects entered our experiment with lots of experience with the type of noise present in the PN condition but with little experience with the type of noise present in the IPN condition. Consequently, one might have predicted that subjects would find it easy to learn to perform the task in the PN condition but difficult or impossible to learn to perform this task in the IPN condition. The results reported here suggest that this prediction is not correct. Future research will need to address this issue in a more detailed manner. That is, future work will need to investigate biases on human motor learning and adaptive control by studying constraints that determine the types of dynamics and noise people can easily learn versus the types they can learn only with significant difficulty, if at all.

Second, as mentioned above, there are many open issues concerning the construction of ideal actors. Ideal actors typically have unlimited perceptual, motor, and cognitive capacities. For example, ideal actors often have unlimited memory, attentional, and processing resources. According to one school of thought, such actors are appropriate because they allow researchers to define the information-theoretic upper bounds on performances that arise solely because of task constraints and limitations on the information provided to the actors by the environment. However, according to another school of thought, ideal actors will provide more useful benchmarks for human performance when they are restricted by many of the same mental limitations as possessed by people. Future research will need to study a broad range of ideal actors to determine when and how human performances are constrained by task and environmental factors versus internal, mental factors.

Third, the ideal actors reported in this article used dynamic programming to calculate optimal control policies. Given that subjects' performances were near-optimal, did they also use dynamic programming to calculate their control policies? It seems unlikely that people perform dynamic programming in their heads, mostly because of its high memory and processing requirements. For similar reasons, researchers in artificial intelligence have explored ways of solving complex tasks through techniques that approximate dynamic programming but have much smaller computational costs (Si et al., 2004). One such class of techniques is referred to as "reinforcement learning" techniques (Sutton and Barto, 1998). Interestingly, neuroscientists have found that the activities of some neurons are closely related to quantities that appear in reinforcement learning equations (Seymour et al., 2004; Knutson et al., 2005; Lee, 2006). For example, Schultz et al. (1997) identified dopaminergic neurons in the primate whose outputs signal errors in the predictions of future salient or rewarding events. Future research will need to directly address the neural substrate of reinforcement learning and other approximate dynamic programming methods.

References

Baddeley RJ, Ingram HA, Miall RC (2003) System identification applied to a visuomotor task: near-optimal human performance in a noisy changing task. J Neurosci 23:3066–3075.

Barlow HB (1957) Increment thresholds at low intensities considered as signal/noise discriminations. J Physiol (Lond) 136:469–488.

Bellman R (1957) Dynamic programming. Princeton, NJ: Princeton UP.

Berthier N, Rosenstein M, Barto A (2005) Approximate optimal control as a model for motor learning. Psychol Rev 112:329–346.

Bertsekas DP, Tsitsiklis JN (1996) Neuro-dynamic programming. Belmont, MA: Athena Scientific.

Busemeyer JR (2002) Dynamic decision making. In: International encyclopedia of the social and behavioral sciences (Smelser NJ, Baltes PB, eds). Oxford: Elsevier.

Clamann HP (1969) Statistical analysis of motor unit firing patterns in human skeletal muscle. Biophys J 9:1233–1251.

Desmurget M, Grafton S (2000) Forward modeling allows feedback control for fast reaching movements. Trends Cogn Sci 4:423–431.

Engelbrecht SE, Berthier NE, O'Sullivan LP (2003) The undershoot bias: learning to act optimally under uncertainty. Psychol Sci 14:257–261.

Flanagan JR, Wing AM (1997) The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads. J Neurosci 17:1519–1528.

Foulkes AJ, Miall RC (2000) Adaptation to visual feedback delays in a human manual tracking task. Exp Brain Res 131:101–110.

Geisler WS (1989) Sequential ideal-observer analysis of visual discrimination. Psychol Rev 96:267–314.

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

Jagacinski RJ, Flach JM (2003) Control theory for humans. Mahwah, NJ: Erlbaum.

Johansson RS (1998) Sensory input and control of grip. Novartis Found Symp 218:4559.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. J Neurosci 25:4806–4812.

Krakauer JW, Ghez C, Ghilardi MF (2005) Adaptation to visuomotor transformations: consolidation, interference, and forgetting. J Neurosci 25:473–478.

Lee D (2006) Neural basis of quasi-rational decision making. Curr Opin Neurobiol 16:191–198.

Marr D (1982) Vision. New York: Freeman.

Matthews PBC (1996) Relationship of firing intervals of human motor units to the trajectory of post-spike after-hyperpolarization and synaptic noise. J Physiol (Lond) 492:597–628.

Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural Netw 9:1265–1279.

Robles-de-la-Torre G, Sekuler R (2004) Numerically estimating internal models of dynamic virtual objects. ACM Trans Appl Percept 1:102–117.

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

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak R (2004) Temporal difference models describe higher order learning in humans. Nature 429:664–667.

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

Si J, Barto AG, Powell WB, Wunsch D (2004) Handbook of learning and approximate dynamic programming. Piscataway, NJ: Wiley-IEEE.

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.

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

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