Learning and optimization of novel sensorimotor feedback loops: Internal models meet classical conditioning

Miro Enev[1] and Emanuel Todorov[2]

Department of Computer Science and Engineering[1]

Departments of Applied Mathematics and Computer Science and Engineering[2]

University of Washington

Abstract

We investigated how novel sensorimotor feedback loops can be formed in the course of learning. More specifically, we examined motor adaptation in an experiment which systematically paired a lateral force pulse at movement onset with a delayed visual target perturbation. Learning in this context means associating the cue and the perturbation, such that sensory feedback about the cue triggers a corrective action suitable for the upcoming perturbation.

The data from the experiment reveals that human subjects gradually embraced the information content of the force impulse and used it to predict the forthcoming target displacement. Behaviorally adaptation manifested itself as (1) movements towards the anticipated target position (after the force pulse ended and before the target perturbation occurred), (2) reduction of the counter-productive stretch-reflex-like response to the force pulse, and (3) reduction in grip force (without change in arm impedance).

To model the main effects of our study, we developed an extension to optimal control which uses a hedging approach to mix target-specific optimal feedback controllers weighted by an agent's belief in the plausibility of future goal outcomes. Using this method we accurately modeled the movement trajectories in the different phases of learning and observed that subject's beliefs converged to the true task statistics. We believe that our extension to optimal control is applicable to other tasks where the central nervous system (CNS) needs to maintain multiple hypotheses about future goals under consideration and prune them in an online fashion as novel information becomes available.

1. Introduction

Feedback plays an important role in sensorimotor control, and is particularly critical in complex behaviors and in the presence of noise and uncertainty. While the sensory guidance of movement has been extensively studied both experimentally and theoretically, most studies have painted a static picture where the underlying feedback loops remain unchanged. One exception in this regard are context-dependent modulation in reflex gains (Loeb et al 1999). Our goal here is to investigate a different type of change, namely how novel sensorimotor feedback loops can be formed in the course of learning.

The work presented here lies at the intersection of two largely disconnected literatures: internal models, and classical conditioning. In our experiments, human subjects learn to anticipate perturbations (target displacements orthogonal to the direction of a reaching movement) and to take corrective actions before these perturbations occur. Thus our paradigm is technically similar to a number of studies on learning internal models (Shadmehr and Mussa-Ivaldi 1995, Kawato 1999, Donchin et al 2003). However such studies have focused on learning open-loop corrections, i.e. corrections for perturbations that can be predicted before movement onset. In contrast, here the perturbation cannot be predicted based on any information that is available before movement onset. Instead prediction becomes possible only after movement onset, when we deliver a cue telling the subject which way the target will be displaced. Learning in this context means associating the cue and the perturbation, such that sensory feedback about the cue triggers a corrective action suitable for the upcoming perturbation. In other words, the learning we study requires the formation of a novel sensorimotor feedback loop.

Our work is related to classical conditioning - where repeated presentation of a conditioned stimulus (CS) followed by an unconditioned stimulus (US) causes the CS to elicit anticipative behaviors appropriate for the US. In some sense, classical conditioning can be thought of as the formation of a novel feedback loop mapping the CS to the US-appropriate behavior. Of course classical conditioning is rarely discussed in these terms, because the time intervals in that literature are longer than what is normally considered to be the domain of feedback control, and also because the response elicited by the CS is typically a full-blown behavior (e.g. eye blink or salivation) rather than a correction to an ongoing movement. Nevertheless one can conceptualize our experiments as a classical conditioning protocol applied within the duration of a single reaching movement.

We found that when the cue is a brief force pulse (delivered to the hand by a haptic robot) learning is successful. However when the cue is either visual or auditory, there is no evidence of learning in the movement trajectories, even though the subjects are aware of the fact that the cue is predictive of the perturbation (they are told so before the experiment begins). This dissociation can be explained mechanistically with the presence of appropriate neural pathways, or computationally with a Bayesian prior which says that lights and sounds are unlikely to correlate with object displacements.

We present a computational model of these results, combining the ideas of optimal feedback control (Todorov 2002, 2004) and Bayesian inference (Kording and Wolpert 2006) in a way that resembles hedging. Models of reaching usually assume that the target is known, and rely on feedback to compensate for any perturbations (Flash and Henis 1991, Hoff 1992, Liu and Todorov 2007). However the feedback mechanisms in such models are not adapted to the statistics of the perturbations, thus they do not address the form of learning we study here. In our new model we mix optimal feedback controllers for different targets. The mixing is somewhat elaborate (see below) and depends on the Bayesian posterior probability over the final target position after the perturbation. The posterior is updated continuously during the movement as new sensory information becomes available. Thus, at movement onset, the hand is controlled by a uniform mixture of feedback controllers. Later, when the cue is detected, the mixture becomes narrower and matches the learning state of the subject. Finally, when the actual target displacement is detected, the system switches to a single controller suitable for this target. It can be shown mathematically that such a strategy of progressive narrowing corresponds to the optimal way to act in this setting. The model's behavior is very similar to the experimental data. Preliminary results were presented in conference format (Enev and Todorov, Neural Control of Movement 2009).

2. Materials and Methods

2.1. Experiments

2.1.1. Experimental setup and data collection

Subjects performed reaching movements in a virtual environment in which we could create controlled perturbations in the form of visual target jumps and haptic force impulses. The setup is shown in Figure 1A. It consisted of a 20" CRT monitor mounted facing down, a horizontal mirror which reflected images shown on the monitor and made them appear in the workspace below, and a 3D robot (Delta Haptic, from Force Dimension Inc.) used to record movement trajectories and deliver force pulses. The handle was instrumented with a 6-axis ATI force sensor which recorded the interaction force between the robot and the subject, as well as a single-axis force sensor used to measure grip force. Subjects could not see their hand; instead they saw a cursor which tracked the horizontal hand position with unnoticeable latency. Subjects grasped the robot handle with their thumb and index finger and could move in a volume measuring about 25x25x25 cm. The lightweight robot arm was easy to manipulate and we provided gravity compensation to minimize the required effort.

The robot's end-effector position (coinciding with the subject's hand position) was recorded at 500 Hz using optical encoders built into the motors. Velocity and acceleration were computed offline by numerical differentiations and filtering (2nd-order Butterworth filter). Force sensor data were recorded at 2000Hz. A custom multi-threaded C++ program (running in Windows XP) was written to enable real-time rendering of the experimental stimuli as well as data collection and control of the robot.

2.1.2. Experimental procedures and subjects

Ten healthy volunteers were recruited for the main experiment. The subject pool was a mixture of university undergraduate and graduate students, 9 males and 1 female, average age 23, 8 right handed and 2 left handed. Every subject used their (self-reported) dominant hand. Another 12 subjects participated in the pilot experiments described later. Prior to data collection subjects we given a chance to get used to the virtual environment, but in general the system was very intuitive and participants quickly moved on to the actual experiment.

Although we allowed movements in 3D, the cursor position rendered in the virtual environment only reflected hand position in the horizontal plane. At the beginning of each trial there was a re-centering phase which required subjects to initiate their movements from an origin defined in 3D, so as to avoid drift over trials in the vertical dimension. Subjects accomplished this using additional visual feedback (showing the vertical deviation) which was removed once re-centering was completed. Subjects were allowed to take breaks between blocks; they usually opted to rest for less than 30 seconds.

Following successful re-centering, the target was shown, always at the same initial location 18cm in front of the hand. In some conditions (see below) the target could be displaced 9 cm left or right, 200 msec after movement onset. Movement onset was detected using a positional threshold (thus the onset of muscle force was earlier). Reaches were self-initiated, however once initiated the movement had to be completed within a condition-specific time limit. Successful reaches were those in which the cursor hit the target circle before the time limit, and the maximum allowed speed (1 m/s) was never exceeded. The speed had to be limited to prevent subjects from hitting the target before it was displaced. At the moment of contact, an animated target explosion was triggered and the number of particles and the intensity of an accompanying sound were proportional to the speed with which the target was hit. This was used to motivate subjects to perform the task well. On unsuccessful reaches subjects were presented with two high pitched beeps and a text message indicating the reason for the error ("time expired" or "maximum speed exceeded").

2.1.3. Experimental design: Main experiment

Each subject performed 8 blocks of reaching movements, totaling 440 trials and lasting approximately 45 minutes. The first four blocks were the experimental condition; the last four blocks were baselines. The numbers of trials and allowed movement durations per block are given in Table 1. Reducing the allowed time limit in the experimental blocks made the task progressively harder.

In the experimental condition (blocks 1-4), the robot applied a force pulse which was triggered at movement onset. The force had a predefined profile: a truncated Gaussian (in time) with mean 90 msec, standard deviation 40 msec, and peak force 12 N (see Figure 1). The force was directed either left or right, orthogonal to the reach direction. At 200 msec the target was displaced by 9 cm left or right, always in the same direction as the force. The direction of force/displacement was randomized over trials, with both directions having 50% probability. To provide consistency, the same sequence of randomized perturbations was presented to every subject. A constraint was enforced to prevent more than three consecutive trials with identical directions - so as to avoid traditional open-loop adaptation. This resulted in a "psychologically" random sequence, even though statistically it was not completely random.

Subjects were instructed to "move as quickly and accurately as possible", and were explicitly informed of the contingency between the force and jump direction with the following verbal script "the target will always jump either left or right during your movement, prior to the jump the robot will always produce a force in the direction of the jump."

The baseline conditions (blocks 5-8) were used to measure the response to the force pulse alone and the target perturbation alone. For half of the subjects, we applied force pulses without target perturbations in blocks 5 and 7, and target perturbations without force pulses in blocks 6 and 8. For the other half of the subjects the protocol was reversed. Before every block, subjects were informed about the forces or target perturbations they were about to experience.

2.1.4. Experimental design: Pilot experiment

Prior to the main experiment, we ran two pilot experiments with 6 subjects each. These experiments differed from the main experiment as follows. First, only the experimental condition (blocks 1-4) was used. Second, instead of a force pulse, we used either a sound (pilot 1) or a visual cue (pilot 2). The pitch of the sound in pilot 1, and the color of the cue in pilot 2, was predictive of the perturbation direction. Subjects were again told about the contingency between the cue and the target perturbation.

2.2. Modeling

The present model is based on the stochastic optimal control framework, which we have previously used to model reaching and other motor behaviors (Todorov 2002, 2004). The novel element here has to do with adapting the control scheme to the statistics of the perturbations. This is done via a decision-tree like approach (see Figure 2) in which the branches correspond to hypotheses about the final target position. Mathematically the branches are finite-horizon discrete-time linear quadratic regulators, constructed as in (Todorov 2002) and connected into a tree as follows. The initial state of each branch equals the final state of its ancestor. The final cost of each branch is a weighted mixture of the costs-to-go of its descendents, evaluated at the descendent initial states and weighed by the descendent probabilities. The costs-to-go are computed recursively using standard Riccati equations. The branch points correspond to the points in time when new information about the final target position becomes available. We now describe the model in more detail.

2.2.1. Dynamics and costs

We model the hand as an $m = 1$ kg point mass moving in a horizontal plane, with viscosity $b = 10$Ns/m approximating intrinsic muscle damping. The controller affects the point mass through two force actuators which can induce positive or negative forces along two orthogonal dimensions; this scheme is intended to resemble two sets of agonist-antagonist muscles. We impose further constraints on the force actuators by making them behave as first-order low-pass filters of the control signals, with time constant $\tau = 0.05$ s.

Let $\boldsymbol{p}$(t), $\boldsymbol{v}$(t), $\boldsymbol{a}$(t), $\boldsymbol{u}$(t) be the two-dimensional hand position, velocity, actuator state, and control signal, respectively. The corresponding units are m, m/s, N, N. The time index varies depending on which branch is being computed: branches in stage 1 have an initial time of 0ms, branches in stage 2 start at 157ms, and those in stage 3 are initialized at t = 357ms. The state is augmented with the final target position $\boldsymbol{p}^*$ which is constant within a branch but varies between branches. The plant dynamics in continuous time are modeled as follows:

$$\dot{\boldsymbol{p}}(\mathrm{t}) = \mathbf{v}(\mathrm{t})$$

$$m\dot{\boldsymbol{v}}(\mathrm{t}) = \mathbf{a}(\mathrm{t}) - b\mathbf{v}(\mathrm{t})$$

$$\dot{\boldsymbol{a}}(\mathrm{t}) = \frac{\mathbf{u}(\mathrm{t}) - \boldsymbol{a}(\mathrm{t})}{\tau} + \boldsymbol{g}(t)\, d$$

Here $\boldsymbol{g}$(t) represents the impulse force, and only acts in the lateral directions:

$$\boldsymbol{g}(t) = [force(t); 0]$$

The function *force*(t) is the predefined force profile, and *d* is the force direction (+1 or -1). We can assemble all variables into an eight-dimensional state vector

$$\boldsymbol{x}(t) = [\boldsymbol{p}(t); \boldsymbol{v}(t); \boldsymbol{a}(t); \boldsymbol{p}^*],$$

and write its dynamics in general first-order form as follows:

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t) + C\boldsymbol{g}(t)$$

with A, B, and C obtained from the above equations.

The objective function being minimized for a terminal branch *i* is

$$J_i = \left\|\boldsymbol{p}^* - \boldsymbol{p}(t_f)\right\|^2 + w_{stop}\left(\left\|\boldsymbol{v}(t_f)\right\|^2 + \left\|\boldsymbol{a}(t_f)\right\|^2\right) + w_{energy}\int_0^{t_f} \|\boldsymbol{u}(t)\|^2 dt$$

These three cost terms encourage endpoint positional accuracy, stopping at the target, and energetic efficiency respectively. We adjusted the relative weights of the these parameters so that they would fit the baseline behavior ($w_{stop} = .05, w_{energy} = .0000005$) Each non-terminal branch had the same energy cost, and final cost obtained by mixing the costs-to-go at the initial states of its descendent branches. Recall that the cost-to-go function is the cost accumulated starting at a given state; in this setting the cost-to-go is always a quadratic function of the state.

Given the above continuous-time formulation, we discretized the time axis at 15 msec time steps, and computed the optimal feedback control law

$$\boldsymbol{u}^*(t) = K(t)\,\boldsymbol{u}$$

where $K(t)$ is the time-varying sequence of optimal feedback gains.

2.2.2. Hypothesis mixing

To account for the uncertainty in the task, we created a branching structure for mixing control hypotheses intended to span the space of plausible future target states. Each control hypothesis was assigned a likelihood (reflecting the perceived probability) which was used as a mixture weight at the branch points. Branch points lie at the boundaries of the three experimental stages and represent moments during the reaching movement at which novel information about the final target position becomes available. The structure of the model is illustrated in Figure 2.

To better understand the hypothesis mixing scheme consider the information available to a subject through the different stages of a trial in the main experimental condition. At the start of stage 1, the subject has high uncertainty because a target jump is forthcoming but there is no information about its direction. Hence the subject constructs a policy which equally weighs the likelihood of a left and right target jump. As a result both weights in Stage 1 are 0.5. At the start of stage 2, the subject has detected the direction of the force pulse. If subjects were fully rational, at this point they would assign probability 1 to the correct outcome, and furthermore there would be no learning in this experiment because subjects are told in advance that the force direction predicts the target displacement. However subjects are clearly not rational (see Results). Instead they basically ignore the verbal instructions, and gradually adapt their behavior in the course of the experiment. We model this as a change in the probabilities used at the start of stage 2: beginning from a uniform distribution and gradually transitioning towards a delta function centered at the correct outcome (this limit is unlikely to be reached within the duration of our experiments). Lastly, at the start of stage 3, the target has jumped to its final location and so there is no uncertainty left.

As an example, consider a rightward trial. The subject has started in the root node (V1) and experienced a push to the right, which at the start of stage 1 has placed him/her in node V3. The target has not yet jumped. Node V6 is the control strategy for reaching the left target while node V7 is the control strategy for the right target. If the subject had learned the task, he/she would place high probability on the rightward hypothesis (p) and low probability on the leftward one (1-p). Mathematically this can be expressed as:

$$V_3(t_{s2f}) = (1-p) * V_6(t_{s2f}) + p * V_7(t_{s2f})$$

where the time index t_{s2f} is used to indicate that the mixing occurs at the final time in stage 2 (although its effects influence all of stage 2). At the end of stage 2 the target jump will be realized and the subject will complete the reaching movement without further consideration of multiple control strategies. The probabilities used in the model are given in Table 2.

3. Results

3.1. Main experiment

The results from the main experiment are shown in Figures 3 and 4. Figure 3 shows movement trajectories and force data averaged over subjects. Figure 4 shows measures of learning for each subject and trial. Since left and right perturbations were symmetric, for analysis purposes we mirrored the left-perturbed trials and pooled them with the right-perturbed trials.

3.1.1. Timing

Although we know the exact timings of all events in the experiment, subjects react to these events with delays, which we inferred from the data. We analyzed the lateral acceleration data using a multi-way ANOVA (Matlab 'anovan') to find times at which the blocks are statistically distinguishable with a threshold of $p < .05$. In our setup the factors were the block numbers while the measures were the average acceleration of each subject (within a block) for a particular time point (1 msec resolution).

The first significant difference occurred at 157 msec after movement onset. This is the time (on average) when subjects began to use the information provided by the force pulse. Thus 157 msec marks the end of **stage 1** and the start of **stage 2**. A similar analysis found the 376 msec mark as the next point of divergence (end of **stage 2** start of **stage 3**) between the blocks, when subjects realized that the target had jumped and altered their movements accordingly. These times are marked as the blue vertical lines in Figure 3.

3.1.2. Learning

The main learning effect is shown in Figure 3A. In block 1, the force pulse perturbed the hand and triggered a simple stretch-reflex-like corrective response (albeit with a long latency). A very similar response (in terms of kinematics as well as grip and interaction forces) was observed in the force-only baseline, indicating that on average there was little learning in block 1. When the subject detected the target jump later in the movement, a corresponding visually-guided correction was triggered, resulting in an S-shaped trajectory. Note that the force pulse here is assistive in the sense that it pushes the hand towards the location where the target will be at the end of the movement. Resisting this force is clearly a suboptimal strategy, yet subjects used it early in the experiment even though they were told in advance that the force will be assistive.

With practice the shape of the trajectory changed substantially and became more straight, and the task-inappropriate response to the force pulse was reduced - as can be seen in the kinematic data. The red horizontal lines in Figure 3A are standard errors, allowing for a visual test of statistical significance. The change from block 1 to block 4 is highly significant. The grip force data also showed changes over blocks: subjects gradually relaxed and held the robot with smaller grip force. We did not observe slips, thus the grip force was sufficient even at the end of the experiment. Note that the subjects grasped a flat metal piece with the thumb on top and the index finger underneath, thus the interaction forces acting in the horizontal plane had to remain within the friction cone created by the grip force in order to prevent slip.

Previous studies have shown that both adaptation (Thoroughman and Shadmehr 1999) and reduction in grip force (Tsuji et al 1995) tend to reduce arm impedance, by reducing co-contraction of arm muscles. However this was not the case in our experiments. While we did not measure impedance directly, it can be inferred from the early effects of the force pulse. If arm impedance had decreased over blocks, the same force pulse would cause a larger deviation (in movement stage 1) towards the end of the experiment. There was no such trend in the data; indeed we defined stage 1 as the time interval when no significant changes in acceleration were found, and this time interval turned out to be about as long as the force pulse itself. Thus the adaptive suppression of the inappropriate stretch-reflex-like response is not associated with reduction in muscle co-contraction, but rather a change in the underlying sensorimotor loop.

Figure 4 shows the trial-by-trial lateral deviation of the hand at 80% into the movement, which is where the largest difference between blocks 1 and 4 was observed. Also shown is a learning index for each subject. This index is the number of successful trials in blocks 3 and 4 (which were the most challenging because they had the shortest allowed movement duration), multiplied by average speed at impact and then normalized so that the maximum over subjects is 100. It is interesting to note that different subjects learned at rather different rates. Two of the 10 subjects were very fast learners and produced nearly straight movements almost from the beginning of the experiment (subjects 3 and 10). One subjects showed very little learning even at the end (subject

1), while the remaining seven subjects showed more gradual adaptation. Of the subjects that learned (9 out of 10) everyone eventually settled to a very similar strategy.

3.2. Model

The trajectories generated by the model are illustrated in Figure 5. Note the similarity to the corresponding subplots in Figure 3. The learning effect is easily captured. Recall that the amount of learning in the model corresponds to the probability that a force pulse in a given direction will be followed by a target jump in the same direction. While this probability is 1.0 in the experimental design (and subjects are told so in advance), subjects behave as if this probability is initially set to 0.5 and then gradually increases, as shown in Table 2. Given the structure of the model, it is obvious that increasing this parameter will result in a learning effect in the same direction as the experimental data. However the quantitative resemblance to the data is remarkable, especially since the model is rather simple in terms of the assumed dynamics.

3.3. Pilot experiments

Earlier (pilot) experiments did not show the expected learning effect, despite our prolonged and frustrating efforts to design an experiment that does show learning. Recall that these experiments had identical design except that the force pulse was replaced by a visual cue or an auditory cue predicting the direction of the target jump. Figure 6 shows the average trajectories for blocks 1, 2 and 3 in the auditory cue task (the data for the visual cue task were essentially the same). We omit plotting the trajectories from block 4 because there were very few successful trials. Furthermore, we often had to end data collection early because subjects got tired.

In these experiments the behavior did change consistently over blocks, but in a very different way from the force-pulse experiment. Here subjects gradually adopted the strategy of moving as quickly as possible to the central position of the target (where it was at the beginning of the trial before jumping), and then making a sharp turn once they detected the target jump. The cue was effectively ignored, and the opportunity to predict the target jump and move towards the expected target position was not taken advantage of. Given that the cue was ignored, such a speeding strategy was necessary because otherwise the movement could not be completed on time in the later blocks. Another systematic change was observed in the grip force data: the grip force now increased over blocks, instead of decreasing as in the main experiment. This increase was probably needed in order to be able to generate larger horizontal accelerations while preventing slip.

4. Discussion

We examined a new form of learning which combines elements of classical conditioning, internal model learning, and reflex gain modulation. Systematic pairing of a force pulse at movement onset and a later target perturbation in the same direction caused gradual changes in behavior. We observed reduction of the counter-productive stretch-reflex-like response to the force pulse, accompanied with reduction in grip force but no change in arm impedance. The movement trajectories in the different phases of learning were accurately modeled by a mixture of target-specific optimal feedback controllers weighted by the probabilities of their corresponding targets. Such a mixture is the optimal strategy in the present setting. The proposed extension to optimal control modeling should also be applicable to other situations where the CNS needs to keep multiple options under consideration and prune them when sufficient information becomes available (Cisek and Kalaska 2010).

While the above summary fits nicely in the theoretical framework of Bayesian inference and stochastic optimal control - which postulates that the CNS makes the best possible use of information available to it so as to maximize movement performance - we were surprised to discover that the learning studied here is by no means general. On the contrary, it took several months and several failed experiments to find a combination of stimuli that elicit the learning we set out to find. Results from two of these failed experiments (pilot 1 and pilot 2) were presented above. Even though the experimental design was conceptually identical to the main experiment, we found that replacing the force pulse with either a visual or an auditory cue abolished learning. Instead subjects developed a speeding-up strategy, which was sufficient to reach the target within the specified time limit, but was nevertheless quite fatiguing and suboptimal. Thus there is a big difference between using a force pulse as a cue vs. using visual or auditory cues. We now consider several possible explanations. Two of them were already mentioned in the Introduction: Bayesian priors which make the pairing of visual/auditory cues and target displacements unlikely (and thus require large amounts of evidence before learning is manifested), or neural pathways which make it easier to use tactile and proprioceptive feedback in the formation of the sensorimotor loops.

Another possible explanation is that the learning studied here is limited to modulation of existing feedback loops as opposed to formation of new feedback loops. Indeed the main learning effect had to do with suppression of a pre-existing response. However we do not think that such suppression is sufficient to explain our results. In block 4 subjects are clearly moving towards the predicted target position after the force pulse has ended. This is best seen in the velocity plot (Figure 3E) which shows that, in block 4, the lateral velocity starts to increase long after the force pulse has ended but before the target jump is detected. Nevertheless it is interesting that the only case of learning we found could be at least partly explained with modulation of existing feedback gains. Perhaps this form of learning has to start with gain modulation before more elaborate changes in the feedback control structure can be made.

Yet another complicating factor is the target perturbation itself. Previous work on standard (open-loop) adaptation has shown that, while subjects readily learn force fields and visuo-motor rotations, learning to anticipate predictable target jumps is much harder (Diedrichsen et al 2005, Liu and Todorov 2007). In our recent work (Liu and Todorov, unpublished manuscript) we have found that subjects can learn to anticipate predictable target jumps but only if we arrange the experiment so that goal achievement becomes impossible without anticipation. Our working explanation is that a visual target acts as an attractor, and even if subjects know that the target will be displaced, they are compelled to keep moving towards the current position of the target. It is as if the perceived task is not "hit the target at the end of the movement", but instead "move towards the current target". Of course in the real world these two tasks are usually the same, and we can again invoke the concept of Bayesian priors to explain why learning to anticipate future goals is very difficult. Thus, in retrospect, the target jumps we decided to use as the "unconditioned stimulus" may have been a poor choice. A better choice may have resulted in learning for arbitrary cues ("conditioned stimuli") and not just force pulses. This is an interesting possibility worth exploring in future work. After all, most of the sensorimotor feedback loops underlying human movement have been learned, and we need to find ways to investigate such learning in laboratory conditions. The present work is a step in that direction.

References

P. Cisek and J. Kalaska, "Neural mechanisms for interacting with a world full of action choices," Annu. Rev. Neurosci., vol. 33, pp. 269-298, 2010

J. Diedrichsen, Y. Hashambhoy, T. Rane, and R. Shadmehr, "Neural correlates of reach errors," J Neurosci, vol. 25, pp. 9919-9931, 2005

O. Donchin, J. Francis, and R. Shadmehr, "Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control," J Neurosci, vol. 23, no. 27, pp. 9032-9045, 2003

T. Flash and E. Henis, "Arm trajectory modification during reaching towards visual targets," J Cogn Neurosci, vol. 3, pp 220 –230, 1991

B. Hoff, "A computational description of the organization of human reaching and prehension," PhD thesis, University of Southern California, 1992

M. Kawato, "Internal models for motor control and trajectory planning," Curr Opin Neurobiol, vol. 9, no. 6, pp. 718-27, 1999

K. Kording and D. Wolpert, "Bayesian decision theory in sensorimotor control," Trends in Cognitive Sciences, vol. 10, no. 7, pp. 319-326, 2006

D. Liu and E. Todorov, "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control," Journal of Neuroscience, vol. 27, pp. 9354-9368, 2007

R. Shadmehr and F. Mussa-Ivaldi, "Adaptive representation of dynamics during learning of a motor task," Journal of Neuroscience, vol. 14, no. 5, pp. 3208-3224, 1994

K. Thoroughman and R. Shadmehr, "Electromyographic correlates of learning an internal model of reaching movements," J Neurosci, vol. 19, no. 19, pp. 8573-8588, 1999

E. Todorov, and M. Jordan, "Optimal feedback control as a theory of motor coordination," Nature Neuroscience, vol. 5, no. 11, pp. 1226-1235, 2002

E. Todorov, "Optimality principles in sensorimotor control," Nature Neuroscience, vol. 7, no. 9, pp. 907-915, 2004

T. Tsuji, P. Morasso, K. Goto and K. Ito, "Hman hand impedance characteristics during maintained posture," Biological Cybernetics, vol. 72, pp. 475-485, 1995

G. Loeb, I. Brown and E. Cheng, "A hierarchical foundation for models of sensorimotor control," Exp. Brain Res., vol. 126, pp. 1–18, 1999

FIGURES AND TABLES

	Blocks	**Duration (ms)**	**Num. Trials**
Main	1-4	(800,650,550,450)	80
Baseline 1	5, 7	650	30
Baseline 2	6, 8	650	30

Table 1. Summary of the experiment's block structure including durations and trial counts.

	Cued Direction (p)	**Non-Cued Direction (1-p)**
Block 1	.5	.5
Block 2	.63	.37
Block 3	.81	.19
Block 4	.93	.07

Table 2: Belief that the target will jump in the direction of the force cue at the onset of Stage 2. Parameters settings are based on best fit to behavioral data.

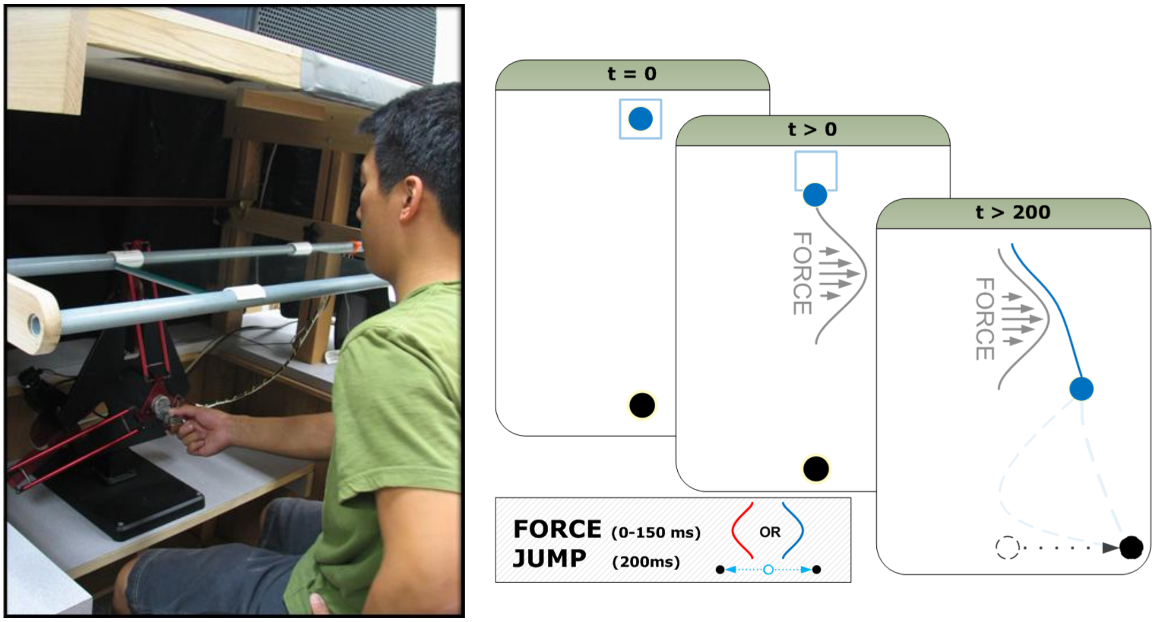

Figure1. (**Left)** Experimental Setup: Subject is shown performing a reach by manipulating the haptic robot while looking at the reflective surface onto which the visual stimuli are rendered. **(Right)** A graphical depiction of a trial with left perturbations, at t=0 the blue cursor is in the home base square, at t > 0 the subject has left the home base region and initiated a reach and the force impulse has also been activated, at t > 200 the target has jumped to its final location; Note that forces are not visible to the participant and are only shown here for illustrative purposes.

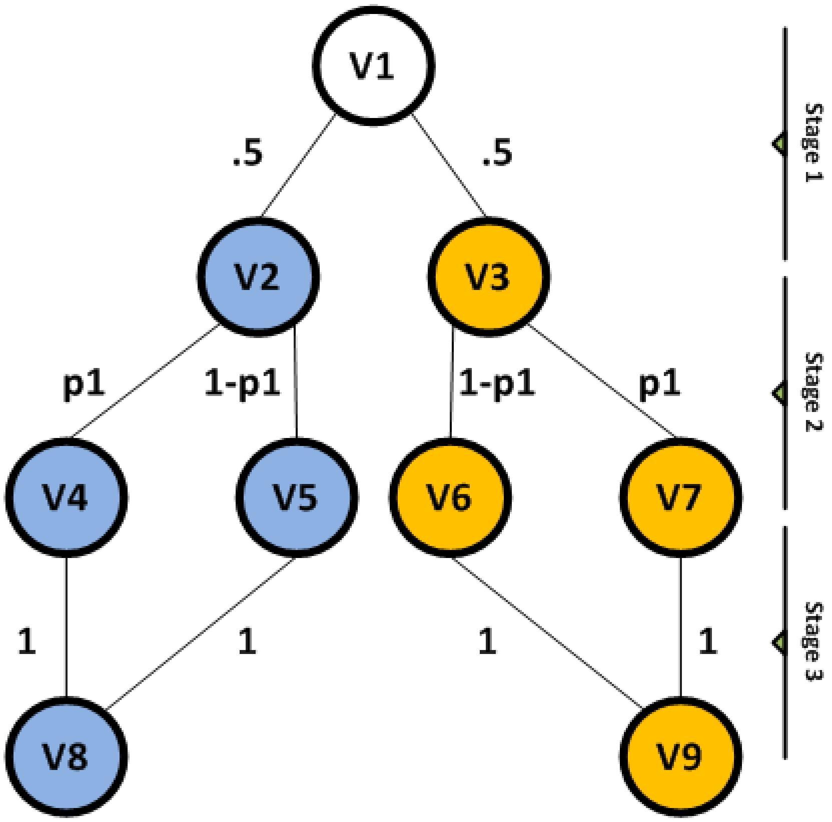

Figure 2: Branching structure of our experiment. A rightward trial causes you to traverse the tree through the orange nodes while a leftward trial places you in the blue nodes. Stage 1 ends at t = 157, Stage 2 ends at t = 376.

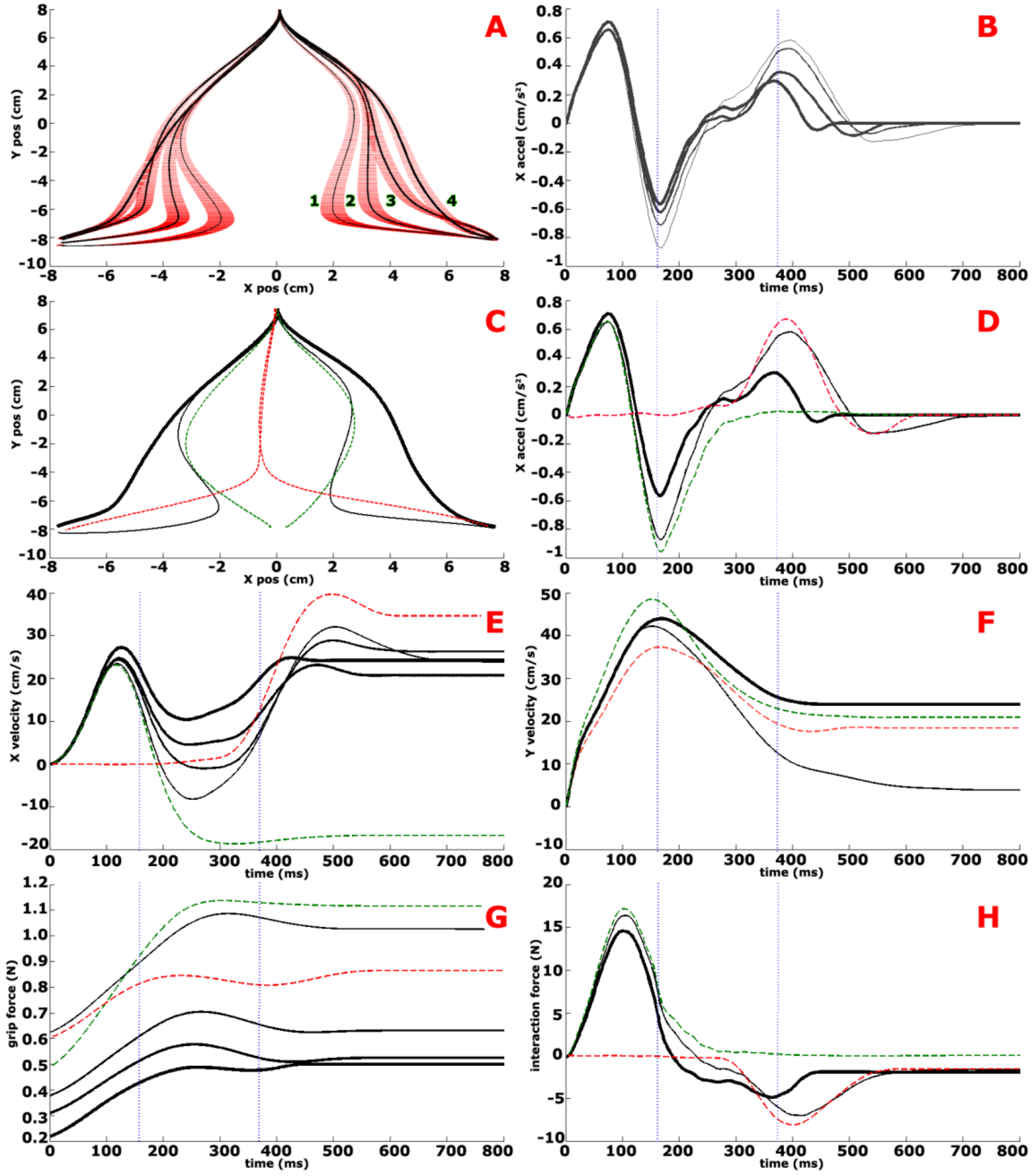

Figure 3. Experimental results shown as time series averaged amongst all subjects for individual blocks (line thickness increases with block number); in all but panels A and C right trials are reflected and combined with left trials; Baseline averages are shown as green (Baseline 1) and red (Baseline 2) dashed lines; the blue vertical dashed lines indicate the separators of the experimental stages (Section 4.1) **A)** Movement trajectories for the 4 main condition blocks (black lines, numbered for clarity), standard errors are shown in red **B)** Lateral velocities for the 4 main condition blocks. **C)** Movement trajectories for block 1 and block 4 with baseline trajectories superimposed. **D)** Lateral accelerations for block 1 and block 4 with baselines superimposed. **E)** Lateral velocities for blocks 1 through 4 with baselines superimposed. **F)** Vertical velocities for block 1 and block 4 with baselines superimposed. **G)** Grip force profiles for blocks 1 through 4 with baselines superimposed. **H)** Interaction forces between the hand and the robot arm for blocks 1 and 4 with baselines superimposed.

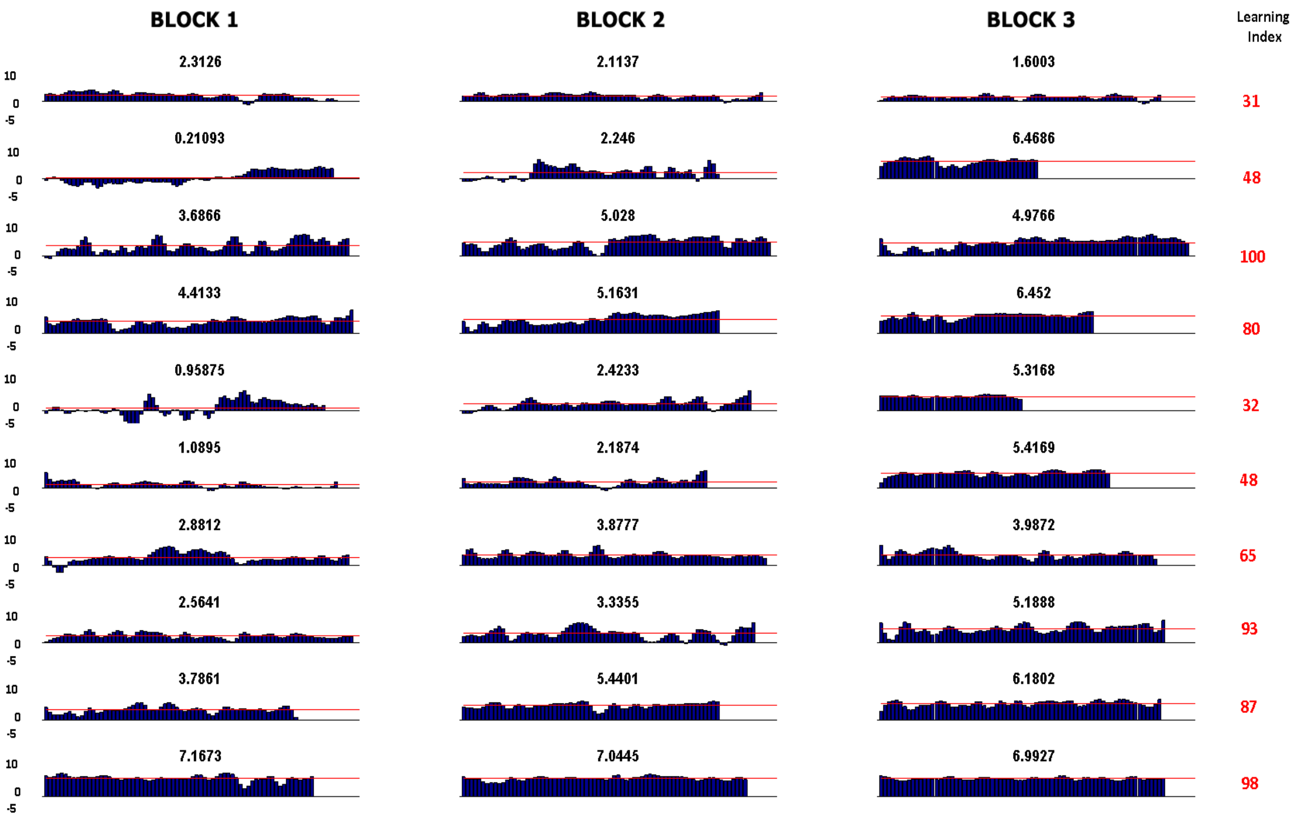

Figure 4: Bar graph of the lateral positional deviations of successful trials in the first 3 blocks of the main condition. The lateral deviation is measured in centimeters at the point in the movement in which participants had coved 80% of the vertical distance to the target. Red lines indicate the block average. Low pass filtered for clarity.

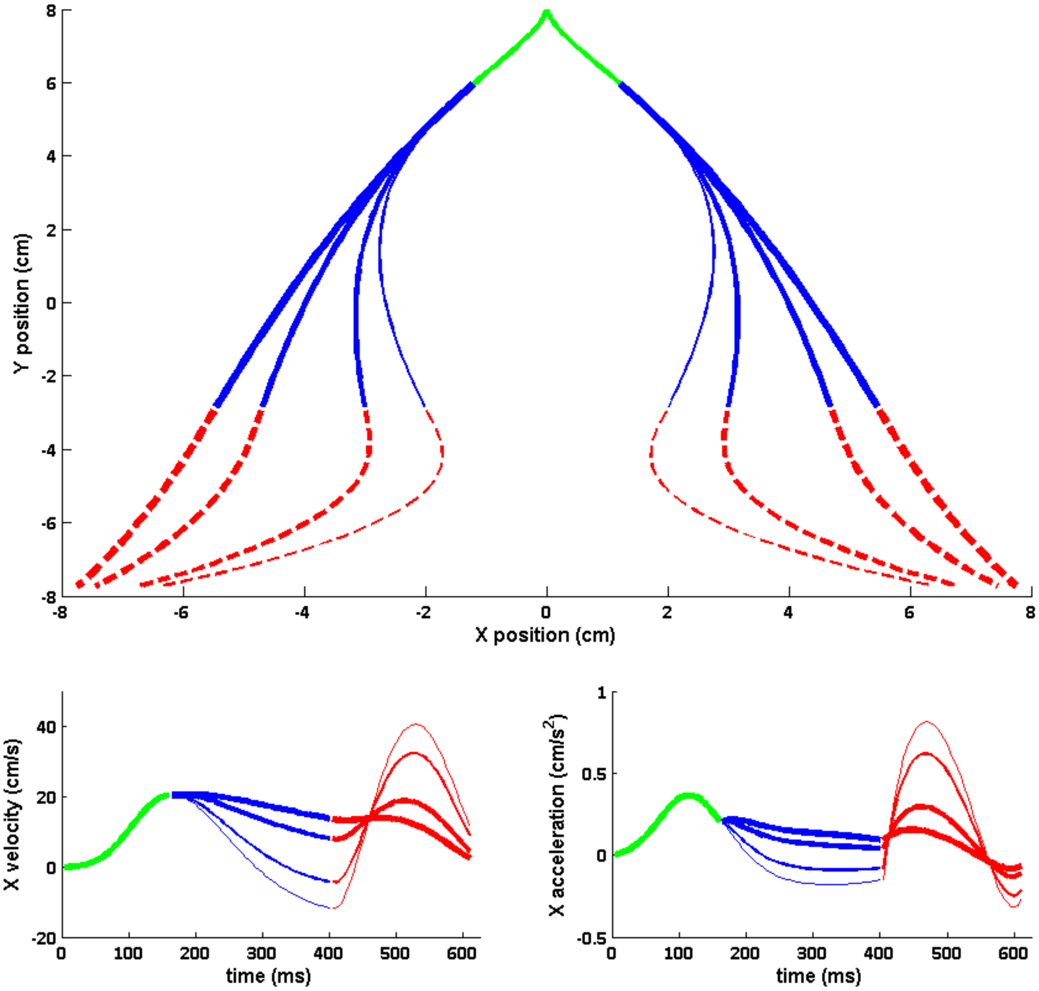

Figure 5: Model Results which capture the average subject behavior in the four experimental blocks. Green segments represent the model outputs during stage 1, blue during stage 2, and red during stage 3. Due to the symmetry of the experiment the velocity and acceleration panels show the behavior for only one perturbation direction.

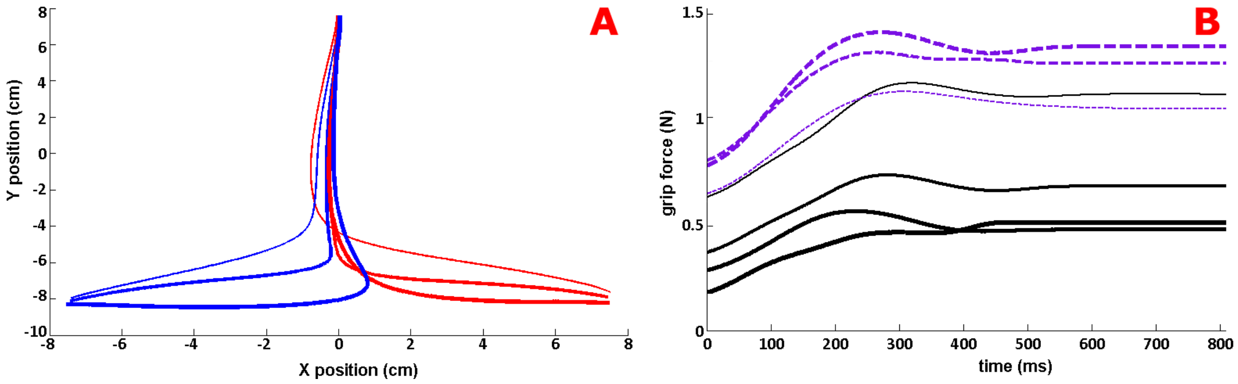

Figure 6: Failed experiments. Panel A shows the position trajectories (colors are used for easier identification in overlapped regions). Panel B shows the average grip force for each block in the auditory cue condition (dashed lines); note the high level of grip force in comparison to the force cue condition (solid lines); in both cases line thickness increases with block number.