Assessing Structure Learning in Motor Tasks

Jonathan S. Berliner, Matthew M. Botvinick, & Jordan A. Taylor
Department of Psychology
Princeton University
jsb4@princeton.edu, matthewb@princeton.edu, jordanat@princeton.edu

Abstract

There is mounting evidence that humans utilize “structure learning,” the identification and utilization of the latent parameters driving action outcomes in a given environment, in motor learning tasks. There is also accumulating evidence suggesting that people engage in “active learning,” selectively sampling their environment in order to quickest reduce their “hypothesis space” of sets of variables that may underlie the environment. We sought to directly assess whether people actively sample their environment in order to best learn its latent structure. Participants made non-rewarded “training reaches,” which they used to inform their movements on rewarded “test reaches,” in a rapidly changing environment. We assessed whether participants would learn to prefer to make training reaches towards “information-bearing” areas that most reduced the hypothesis space of candidate environment structures. Participants learned to selectively sample the more information-bearing areas of the task environment. Further, given equal information-yield across the training space, participants preferred to train in areas near those in which they expected to later be tested, a trait not predicted by certain implementations of structure learning in the motor domain. We provide evidence suggesting that, when engaging in motor tasks, people may employ heuristic-based movement strategies that are more agnostic to the environment than strategies utilizing hypothesized latent structure would predict.

Keywords: Active Learning, Structure Learning, Motor Learning, Heuristics
Introduction

Learning curves during the learning of a novel motor task exhibit two stereotypical phases: an early rapid phase, accounting for the majority of performance gains, and a latter slower phase, which proceeds very gradually with only modest improvements in proficiency (Redding & Wallace 1996; Redding et al 2005; Smith et al, 2006). Traditional models of motor learning tend to capture the latter, but not former, portions of these learning curves (Snoddy, 1926; Crossman, 1959; Newell & Rosenbloom, 1981). Recent experiments have suggested that these two phases may be an instantiation of recent successful decision making models that also discretize policy learning into the learning of task structure and the setting of control parameters (Gershman & Niv, 2010; Braun et al 2009; Braun et al 2010; Turnham et al 2011). The former process has been called “structure learning,” which is the identification of latent variables driving action-outcomes in environments, while the later process has been called “parametric learning,” which is the incremental tuning of the parameters of the latent variables.

Most motor learning research and models tend to focus on parametric learning, but evidence is mounting arguing structure learning to be essential to capturing the initial rapid phase of learning. It has been argued that structure learning in motor tasks reduces parameter space to “metaparameters” defining the covariance of effector location along orthogonal axes in cartesian space (Braun et al, 2009; Turnham et al, 2011), thus collapsing a high dimensional parameter search space into one with only \( n^2 \) dimensions, \( n \) being the amount spatial dimensions in a task. Once the assumed covariance, or structure, is determined, then learning can shift focus to parameterization where control parameters are adjusted gradually. This would also hasten learning in any new motor task that shares a similar covariance structure, since the parameter space has already been identified and reduced. This account has had success matching humans’ ability to learn to overcome novel perturbations surprisingly fast after learning to overcome previously presented similar perturbations (e.g. Braun et al, 2009).

An additional process, called “active learning” may also hasten structure learning by selectively sampling the parameter space in a systematic manner as opposed to randomly sampling. Active learning affords faster performance gains by choosing to sample the most “informative” areas of an environment, typically taken to be areas that greatest reduce the set of candidate hypotheses (i.e. candidate structures) that could underlie the environment (Dasgupta, 2012). What constitutes an informative sample in motor tasks is currently ill-defined. If the metaparameter account of structure learning in motor-control is correct, then we should see sampling strategies in humans that aim to reduce the set of plausible covariance matrices of effector location.

Experiment 1

We designed a task to test whether participants systematically make reaching movements that best reduce the set of candidate structures underlying an environment. Participants made reaching movements in the environment while holding onto a robotic manipulandum. Participants viewed the environment by looking down on a horizontally-projected visual display, which occluded vision of their hand. The visual display consisted of two perpendicular lines (axes), oriented at 45° and 135°, which defined the “training dimension.” There were two designated “training areas” located at seven and twelve cm out along each axis (see figure 1). Participants were instructed to reach from a central starting position to any of the four areas. The robotic manipulandum created a virtual channel constraining subject reaches to the two axes. Feedback of the reach was withheld until a subject stopped their hand in a training area at which point visual feedback was provided in the form of a circular cursor. The cursor was displaced from the participants’ actual hand location by an amount defined by an x-y covariance matrix (see figure 2).

Following each reach, the robotic manipulandum gently returned the hand to the starting position before starting the next trial. After three such trials, a circular test target appeared in the center of the display. Participants were instructed to reach and stop their hand where they thought the cursor would land as close as possible to the target. Participants were never shown the actual location of their cursor given their terminated hand position, and instead given points proportional to the distance between the unseen cursor and test target. This sequence of three “training” trials and one “test” trial constituted a round in the experiment. Participants were told they would not score any points during training trials, and that these were to be used to best inform them for that round’s test reach. This design allows us to directly assess purely “information bearing” actions without any possible confounds that could arise from explore/exploit tradeoffs (Daw et al, 2006; Reverdy et al, 2012).

The covariance matrix defined a precise structure of the environment, a visuomotor mapping
between hand location and visual cursor feedback. Twenty participants were assigned to either a “single shear” (SS) or “double shear” (DS) conditions. For the SS group, all covariance matrices were shear-mappings with shear-factor $\lambda$ relative to a single axis (see Figure 2; Equations 1 and 2). This makes one axis, the “major” axis, information bearing in that it provides evidence of the value of $\lambda$, whereas the “minor” access shows veridical feedback for any value of $\lambda$ and so cannot reduce the hypothesis space. For the DS group, all covariance matrices were shear-mappings relative to both axes, and so both axes provided the same information about the value of $\lambda$. Eight shear factors uniformly distributed around 0 were each used for eight rounds for a total of sixty-four rounds (256 total reaching trials).

**Results and Discussion**

If participants were using active structure learning, then they should preferentially choose the major axis in the SS condition as it is more information bearing. In contrast, in the DS condition, both axes provide the same information about possible values of $\lambda$. Hence, an active structure learning account would predict participants to not acquire an axis preference. Indeed, participants in the SS group learned to prefer the major axis, while DS participants never grew to prefer a particular axis (Figure 3A). While DS participants never gained an overall axis preference as expected, the DS group systematically chose the axis which would put the cursor “inside” the axes (Figure 3B). This behavior is not predicted by discrete-phase structure learning, as the cursor falling inside or outside the axes does nothing to further reduce structure hypothesis space. An alternative account is that participants are simply using their training trials to land their cursor as close as possible to what they believe will be the test target location, rather than trying to best learn the structure of the environment. This is especially surprising considering that the participants know they cannot gain any reward from these actions. This behavior suggests that participants may eschew latent variable learning, and instead heavily focus their active learning on training in spots they expect to be near test trial target locations, rather than attempting to learn about the environment per se.

**Experiment 2**

We reasoned that, if participants were using their training reaches to try to land the cursor near expected target locations rather than to learn environment structure, we should see people learning through time to land their cursor reaches closer to possible target locations. To specifically distinguish whether participants were trying to use their practice reaches to land near expected target locations rather than to learn about the latent structure of the environment, twelve participants were again split into SS and DS conditions and recruited to play a modified version of the game used in experiment 1. Participants were free to reach to any point along the two axes during training trials. For test trials, the test target could now appear at one of two locations either 11cm or 22cm straight ahead. Target distance was increased from that of Experiment 1 to ensure that cursor termination the cursor near target locations was actually due to trying to land near the target and not due to a propensity to only reach a minimal distance.

**Results and Discussion**

Similar to experiment 1, major and minor axis preferences were again shown to be learned in SS but not DS conditions (Figure 4A). Inside and outside axis preferences were again shown to be learned in DS but not SS conditions (Figure 4B). Participants in the DS condition learned to stop their hand in training reaches so that the cursor would land closer to expected near-target and far-target locations (Figure 4C). This is especially apparent when analysis is limited to only the second training trial after they have learned from the first reach which axis is the inside axis (Figure 4D). These results, when taken together with experiment 1, suggest that participants may not be engaging in behaviors consistent with learning the latent structure of the environment.

Bias towards expected target location during training reaches suggests that participants considered these locations particularly information bearing. With so little opportunity to learn about the environment itself, we reasoned that subjects may be utilizing heuristics that focus more on expected target location and less so on latent environment structure. We created a simple “shift” model that predicts hand location for a test reach by taking a participants’ closest reach for said round and simply translating hand location by the vector that was the distance between cursor and target location (see Equations 3, 4, and 5). This model would predict training reaches attempting to minimize cursor distance to target, as its success relies on a sufficiently small translation vector if the subject expects the translation to perturb cursor location in a non-constant manner across the environment. We compared this model to the “ideal learner” that would attempt to perfectly counteract the displacement. To control for use-dependency, we also examined whether participants were just repeating previous actions that landed the cursor closest to the target (Figure 5A). The shift model predicted participant test reach location significantly better than that predicted by the model specifically attempting to counter the perturbation in the environment (5B), as well as the repeat model.
**General Discussion**

We sought out to see if participants actively sampled the task environment in an attempt to best learn latent structure. Participants’ behavior deviated from what would be predicted by an agent attempting to actively learn environment structure through training. During training trial reaches, participants appeared to employ an action policy of getting the cursor close to expected target locations rather than choosing locations to best reduce candidate latent structures. This is not to say that participants did not engage in a form of active structure learning; after all, participants in the SS condition preferred hypothesis-space reduction to target-proximity. However, these experiments at the very least do show that, given an area yields non-trivial information about itself, participants will preferentially explore areas where they presume there will later be reward. Our model results suggest that this is not simply a use-dependent effect; subjects appear to be employing action plans that specifically rely on having practiced sufficiently close to where they will be tested. Models like these may trade precision and generalization for faster local performance, and allow agents to make reasonably successful actions even when they have a poor understanding of their environment. They also shift the focus of utility functions in structure learning paradigms from best inferring latent structure to best inferring reward location.

**References:**


Figure 1: Task environment for experiment 1 included two training axes (gray lines), each with two training areas (pink). During training trials, the manipulandum constrained movements to be along the axes. Whenever a subject’s hand passed under a training area, the region brightened to let them know they could stop there. During test trials, the red test target appeared and the reaches were no longer constrained by the manipulandum. The task environment was similar for experiment 2, but there were no more training areas and the target could appear at two locations (11cm or 22cm from the starting position).

Figure 2: The mapping between hand to cursor locations was defined by a 2x2 shear transformation matrix. The Single Shear (SS) perturbations are defined in equation 1 and the Double Shear (DS) perturbations are defined by equation 2. A) vector field of hand to cursor mapping for $\lambda = .4$ (purple) and -.4 (red) in the SS condition. B) vector field of hand to cursor position for DS condition for $\lambda=2$. The “major axis” is treated as the x-axis and the “minor” treated as the y-axis.

<table>
<thead>
<tr>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $\begin{bmatrix} x_f \ y_f \end{bmatrix} = \begin{bmatrix} 1 &amp; \lambda \ 0 &amp; 1 \end{bmatrix} \cdot \begin{bmatrix} x_o \ y_o \end{bmatrix}$</td>
</tr>
<tr>
<td>(2) $\begin{bmatrix} x_f \ y_f \end{bmatrix} = \begin{bmatrix} \lambda &amp; 0 \ 0 &amp; \lambda \end{bmatrix} \cdot \begin{bmatrix} x_o \ y_o \end{bmatrix}$</td>
</tr>
<tr>
<td>(3) $\text{transl} = \begin{bmatrix} \text{shift x} \ b_{\text{shift}} \end{bmatrix}$</td>
</tr>
<tr>
<td>(4) $\text{transl} = \begin{bmatrix} \text{shift x} \ b_{\text{shift}} \end{bmatrix}$</td>
</tr>
<tr>
<td>(5) $\text{transl} = \begin{bmatrix} \text{shift x} \ b_{\text{shift}} \end{bmatrix}$</td>
</tr>
</tbody>
</table>

Figure 3: Axis preferences for experiment 1. A) Major vs. minor axis preferences during training for participants in the SS (red) and DS condition (green). B) Axis preference for the cursor to land inside or outside the two axes. An epicycle contained one round containing every shear factor (8 movements rounds per epicycle).

Figure 4: A) Major or minor axis preference for the cursor to land inside or outside the two axes. An epicycle contained one round containing every shear factor (8 movements rounds per epicycle). B) Inside or Outside Axis Preference.

Figure 5: Model comparison for test trial reaches. A) Participants’ hand position (black), ideal hand position to offset shear displacement (green), “shift” model predicting simple translation of the training reach that landed the cursor closest to the target (blue), and “repeat” repeating closest training reach for this round (magenta). All three predictions scale expected location by individuals’ propensities to undershoot where they expect to reach to. The shift model is defined by equations 3, 4, and 5. Repeat is the special case of B) Average distance from model-predicted-to-actual hand location for each model. Superimposed circles are average error for individuals participants. The shift model on average has significantly less error than the gain-shifted ideal model ($p=0.03$).

Figure 4: Axis preferences for experiment 2. A) Major vs. minor axis preferences during training for participants in the SS (red) and DS condition (green). B) Axis preference for the cursor to land inside or outside the two axes. C) Average reach distance along axis relative to the close (red) and far test target location (blue), averaged over all training reaches during a round. D) Average reach distance along axis relative to close and far targets for only the second reach in the round.