Effect on Movement Selection of Evolving Sensory Representation

Ashvin Shah[1] and Andrew G. Barto[2]

[1]Neuroscience and Behavior Program and [2]Department of Computer Science, University of Massachusetts Amherst

ABSTRACT

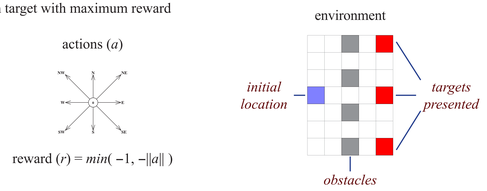
We describe a computational model that focuses on decision making based on *evolving sensory representation*. When selecting a goal-directed movement, we often consider our current state, possibly including a representation of target. Under most accounts of this process, a decision must be made based on the current representation of state; the state representation does not evolve without the learner interacting with the environment. However, sensory information takes time to process. Assuming it evolves from a crude estimate to a more precise one, it may be advantageous to make a quick decision – possibly suboptimal for the actual task – based on the crude estimate rather than wait for a more precise estimate.

Our model explicitly includes sensory information which evolves from a crude estimate to a more precise one over time (independent of any decision made). The model builds on a multiple controller scheme [1, *cf.* 2,3], based on biological studies, in which a Planner controller makes reasonable decisions based on fully resolved state information. Simpler controllers, which require training but less computational resources, learn to assume control under appropriate circumstances and can select movements based on crude state information.

In this poster, we apply the model to a simple decision-making task in which a learning agent must execute a series of actions (analogous to movement selection) to hit a designated spatial location (target). Targets are presented randomly from a small set. Target representation evolves from a probability distribution over all possible locations – built through experience – to a more precise one in which only the current target is represented. Through reinforcement learning and Hebbian learning, simpler controllers learn to select actions based on crude sensory information. We discuss model behavior and implications to motor control.

ENVIRONMENT and TASK

- environment is a "grid-world," learning agent is in location (l)
- agent must choose an action (a) to navigate towards a target ($targ$)
- each action taken incurs a context-dependent reward (r)
  - goal: reach target with maximum reward



STATE REPRESENTATION

State: location *and* target

- s: location l and target $targ$; $N_{states} = N_{locations}$ X $N_{targets}$
- state vector ($\mathbf{s}_l$): $N_{targets}$-element vector corresponding to location l
  - each element of $\mathbf{s}_l$ corresponds to a possible target

Location of agent is known, but *target location is uncertain*

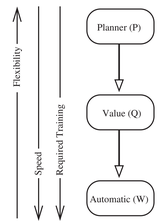
Target belief vector (**b**): $N_{targets}$-element vector:

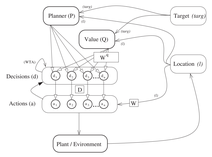
- element i: belief that target i is the actual target (sum(**b**) = 1)
- b_i increases with time if i is the actual target
  - all other b_i's decrease
- target belief evolution occurs *independent* of any decisions made

How does evolving target belief affect decision-making?

MULTIPLE CONTROLLER MODEL

- planning areas
  - takes goal into account when planning
  - requires attention, thought, and time
  - planning and cognitive areas of cortex
- with repetition, simpler controllers are engaged
  - learn how "valuable" each movement taken in each context is
  - requires less resources
  - less cognitive areas of cortex and BG
- repeat same decisions and movements enough times, use simplest scheme possible: *motor skill*
  - sensory information elicits movement (similar to SR mapping), goal not represented
  - requires least resources
  - thalamus to striatum?





Connectionist model

- action taken when *action* neuron is excited past threshold
- **P** and **Q** excite *decision* neurons, which excite *action* neurons
- WTA in *decision* neuron array
- **W** excites *action* neurons directly

Planner (P)

- use AI algorithm (A*) to calculate best actions for a target
- excites decision neurons strongly
- requires *fully resolved* target belief ($b_{targ} = 1$)

Value-based (Q)

- agent has an estimate, $Q(s,a)$, of how valuable each a is for each s [4]
  - thought to be mediated by dopamine in PFC and BG
- learns these values with experience (visiting locations and taking actions)
- $Q(\mathbf{s}_{l,t},a_t) = Q(\mathbf{s}_{l,t},a_t) + \alpha(r + \gamma Q(\mathbf{s}_{l,t+1},a_{t+1}) - Q(\mathbf{s}_{l,t},a_t))\mathbf{b}_t$
  - all $Q(\mathbf{s}_{l,t},a_t)$ updated, weighted by $\mathbf{b}_t$
- Q used to train W^q
  - W^q excites *decision* neuron array (noise allows for exploration)
  - W^q grows from weak connections (no *decision* neuron wins WTA) to stronger connections

Automatic (W)

- $W(l,a)$, weight from l to a, is strengthened for each (l,a) experienced
- $W(l,a)$ for all actions *not* taken is weakened

Arbitration scheme: **W** is faster than **Q**, which is faster than **P**

- **W** is engaged earliest
- if no *action* is selected, **Q** is engaged next
- if no *action* is selected, **P** is engaged

Training:

- at first trial, **b** is uniform distribution over all locations
- at each trial, one of the three target locations is selected randomly
  - multiple controller model is used to select actions to move agent from initial location to selected target location
  - as trial progresses, **b** evolves to represent actual target with belief of 1
- as training progresses,
  - initial **b** approaches uniform distribution over only the three targets presented to the model
  - simpler controllers are engaged

HYPOTHESES

Model Behavior

- During early trials, the model will wait until **b** is resolved enough to move. During later trials, the model will use simpler controllers to move right away towards the intermediate target.

Learning with Fully-Resolved Targets

- This will bias the agent toward taking the most direct route. However, if presented with an evolving target representation, it will have to wait a few steps before moving.

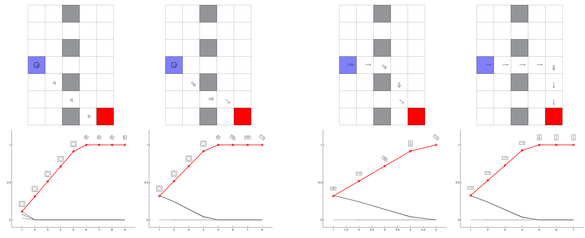
MODEL BEHAVIOR

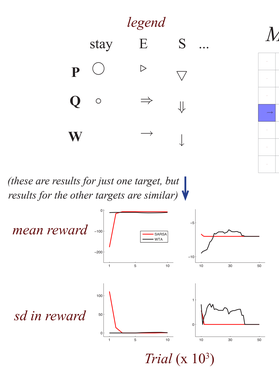
Early trials:

- **P** dominates control and it must wait until **b** is fully resolved.
- Eventually, **Q** and **W** are trained enough to make some decisions.

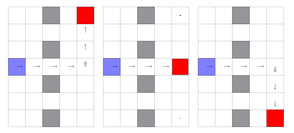
Late trials:

- **P** isn't used at all and the model makes movements right away.
- Eventually, it develops an *MS* to move straight toward the intermediate target.





MS's reflect target distribution



- Presence of Planner aids in early trials
  - compared performance with planner versus decision making based on picking highest Q-value (**W** still active)
  - the latter needs to explore a lot before it finds the target
  - at areas of (s,a)-space where agent has little experience, **P** ensures reasonable performance

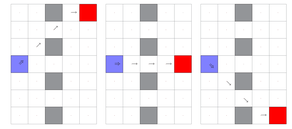
References and Acknowledgements

1. Shah, A. and Barto, A.G. (2007). Functional mechanisms of motor skill acquisition. Poster presented at Computational Neuroscience Meeting, July 7-12, Toronto, ON, CA.
2. Daw, N., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience, 8:1704--1711.
3. Shah, A., Barto, A.G., and Fagg, A.H. (2006). Biologically-based functional mechanisms of coarticulation. Poster presented at Neural Control of Movement Conference, May 2-7, Key Biscayne, FL.
4. Sutton, R., and Barto, A.G. (1998). Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
5. Battaglia, P.W. and Schrater, P.R. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast-reaching task. The Journal of Neuroscience, 27:6984-6994.
6. Hudson, T.E., Maloney, L.T., and Landy, M.S. (2007). Planning movements toward probabilistic mixtures of possible targets. Journal of Neurophysiology (in press).

This research was made possible by NIH grant # NS 044393-01A1

LEARNING WITH FULLY-RESOLVED TARGETS

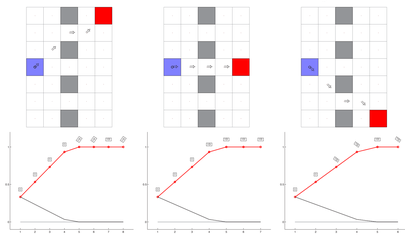
Model trained with a *fully resolved* **b** at the start of each trial



- agent takes most direct route and forms appropriate *MS*'s

How does this training and behavior affect Q-values?

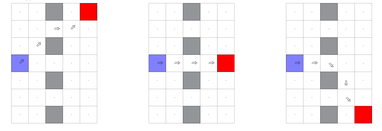
- to better assess Q-values, *MS*'s were turned "off" (**W** set to **0**)
- the trained model was presented with an *evolving* **b**
- **b** was initialized to uniform distribution over only the three targets



Performance suffered

- agent must wait several steps before making movement
- in many cases, makes bad decisions (*not shown*)

Model trained with an evolving **b** was able to act reasonably well when presented with a fully resolved **b** at the start of a trial



Discrepancy of behavior between the two training methods:

- due to use of *MS*'s (**W**)
- if models were trained without *MS*'s, Q-values (and thus behavior) were very similar (*not shown*)
  - suggests possible behavioral experiments to elucidate learning and use of motor skills

DISCUSSION

Much research on motor control and decision-making investigates how goal specification and environmental attributes (*e.g.*, dynamics, kinematics, perturbations) affect behavior. In this poster, we focused on how an evolving sensory representation affects decision-making and motor skill development. We show that, rather than wait for the sensory representation to resolve, the agent instead chooses to move immediately toward an intermediate target. The results of our modeling work agrees, on a qualitative level, with recent experimental work investigating movement selection under an evolving sensory representation:

Battaglia and Schrater [5] show that human subjects, presented with crude target information that becomes more precise during a trial, will reach for the perceived target based on crude information in order to allow them time to make an accurate movement. An explicit trade-off between perception quality and movement accuracy was observed. Hudson et al. [6] devised a reaching task in which human subjects were presented with a probability distribution over possible targets; only after the subject completed part of the movement did the target information become fully resolved. They showed that the subjects' initial direction of movement was based on the initial probability distribution and that direction deviated toward the target after the fully-resolved target information was provided.