Supplementary Material - Information Theoretic Intent Disambiguation via Contextual Nudges for Assistive Shared Control

Deepak Gopinath[1,3], Andrew Thompson[1,3], and Brenna D. Argall[1,2,3]

[1] Department of Mechanical Engineering, Northwestern University, Evanston, IL
[2] Department of Computer Science, Northwestern University, Evanston, IL
[3] Shirley Ryan AbilityLab, Chicago, IL

1 Illustration of the Intuitiveness of $D(s)$

In this section, we provide an illustration of how the proposed disambiguation metric matches our intuitions of what the maximally disambiguating states are for different prior beliefs over goals in a simple environment. The simulated environment under consideration is a 2D gridworld with three goal locations in which the robot state space $\mathcal{Q}$ is the (x, y) coordinate of each grid cell (Figure 1). The simulated teleoperation interface is a 1D interface with bi-directional mode-switching capabilities.[4]

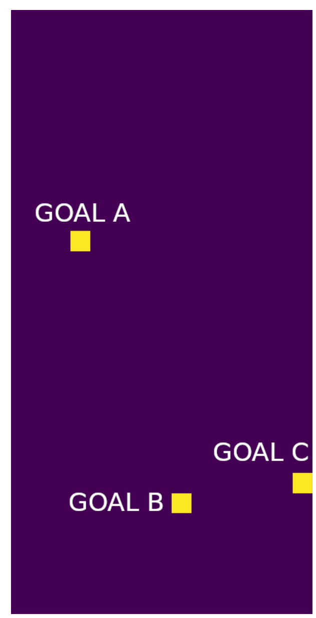

Fig. 1: Goal configuration used for simulation using a 15 $\times$ 30 2D grid.

Since the dimensionality of the interface is lower than the dimensionality of the robot, the control space is partitioned into two modes: specifically, $\mathcal{M} = \{$*Horizontal*, *Vertical*$\}$, allowing for motion along x and y dimensions respectively. The action space $\mathcal{A}$ consist of $\mathcal{A}_c = \{$*move-positive*, *move-negative*$\}$ that consists of actions that allow positive and negative motion along the active dimension at any given time and $\mathcal{A}_q = \{$*switch-right*, *switch-left*$\}$ which consists of mode-switch actions that result in mode switches. Note that for an ordered set $\mathcal{M}$, *switch right* results in a transition from mode m_i to m_{i+1} with wrap around to the first element of the set and *switch-left* results in a mode transition from m_i to m_{i-1} with wraparound to the last element of the ordered set. In this scenario (with just two control modes) a mode switch in one mode simply results in a transition to the other mode (Figure 2).

We model the simulated human as a Markov Decision Process described in Section 3.1 and use value iteration to obtain a goal-dependent policy, $p(a|s, g)$. The reward

[4] Note that we are exploring just one possible configuration of goals in this illustrative example. The proposed metric is applicable for *any* configuration as well as number of goals.

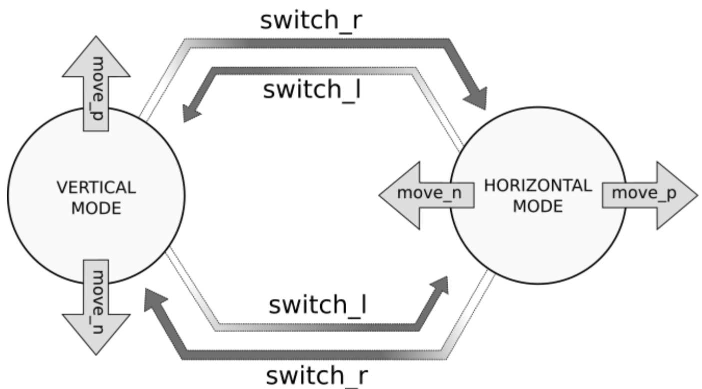

Fig. 2: Mode switching diagram for 2D grid world world.

function $\mathcal{R}$ is designed to minimize the distance travelled as well as the number of mode switches executed. The optimal policy shown in Figure 3 is obtained using value iteration. Upon inspection, we can see that following the policy from any states results in straight line paths and minimal number of mode switches.

The top row of Figure 4 shows the disambiguation metric computation for all states (visualized for each mode separately) for a uniform prior over goals and $\lambda = 0$. Within the bounds of the goal region, we see that the maximally disambiguating states for the *Horizontal* mode correspond to the states that are aligned with goal B along the y dimension; this meets our sanity check as any movement left in these states would suggest goal A, any movement right would suggest goal C and a mode switch to vertical motion would suggest goal C. Similarly, the maximally disambiguating states for the *Vertical* mode are states that are aligned with goal C along the x dimension. In this case we once again see that any movement up suggests goal A, any movement down suggests goal B and a mode switch to horizontal motion would suggest goal C. Since no two goals have the same action mapped to them, they allow for maximal goal disambiguation.

We also see that in Figure 4 (top and bottom) there are states in which the disambiguation metric is identically equal to zero (dark blue color). Under the assumption that the human approximately behaves like the MDP policy shown in Figure 3, we can see that for any states outside of the horizontal and the vertical limits of the goal region, there is a high likelihood that the human would choose the *same* action in those states regardless of the intended goal. As the actions are *indistinguishable*, an observer will not be able to determine (from executed actions alone) the true intended goal with certainty. In the bottom row of Figure 4, the priors are non-uniform with the probability associated with goal C set to be zero. This simulates a scenario in which the history of states and actions have already caused the Bayesian update of belief to assign zero probability to goal C. Effectively, it becomes a question of disambiguating between goals A and B. The disambiguation metric computed for this scenario correctly ignores goal C completely.

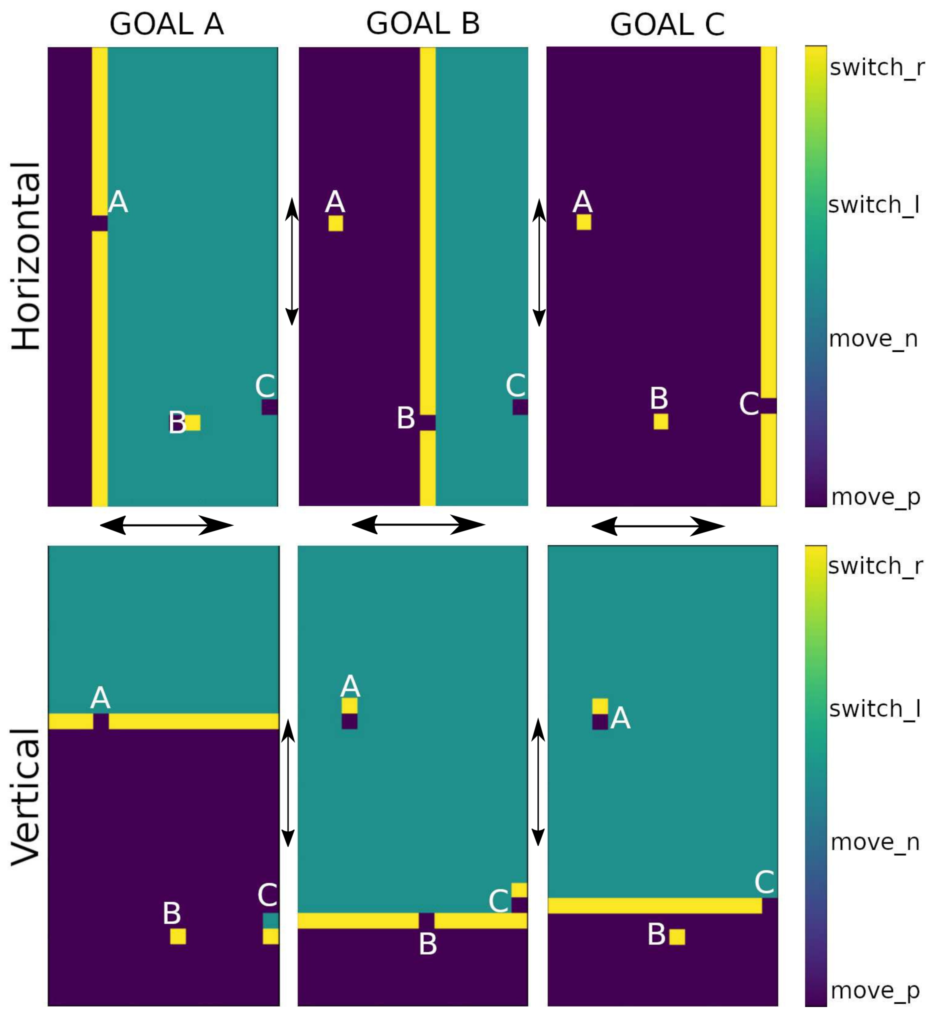

Fig. 3: Learned policy using value iteration for the each of the goals in the goal configuration shown in Figure 1.

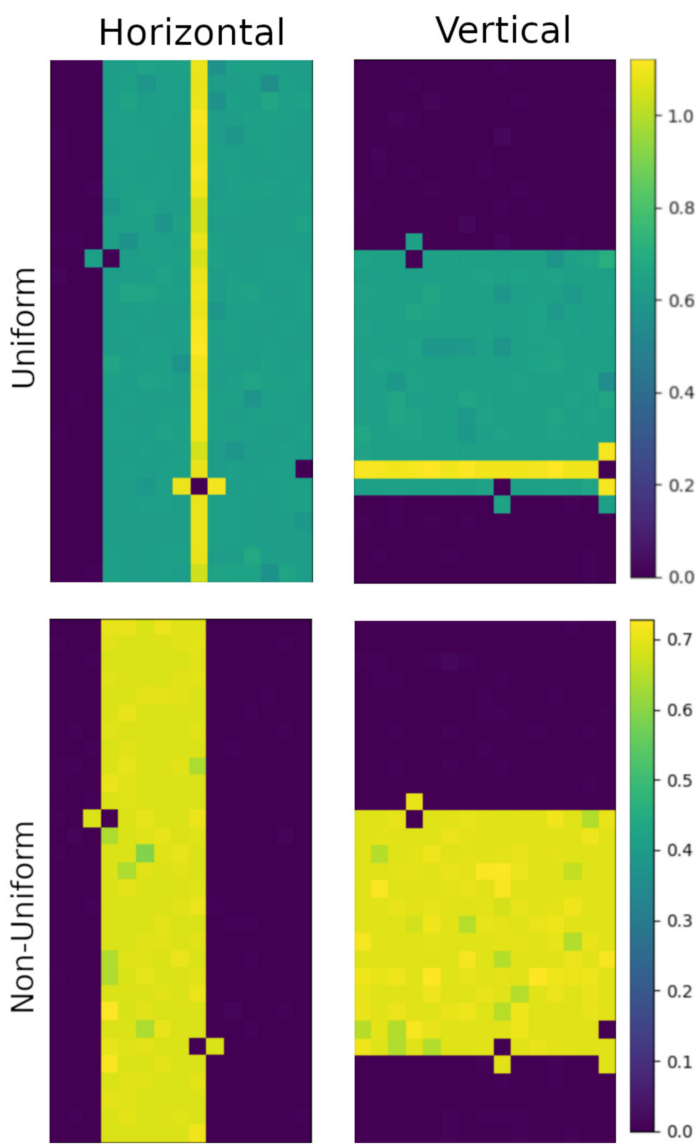

Fig. 4: Top Row: $D(s)$ computed for all states in the grid world over a uniform prior for each mode. Bottom Row: $D(s)$ computed for all states in the grid world for nonuniform prior for each mode. Note that λ was set to be 0.0 to highlight the contribution from the mutual information term.