Expected Value of Communication for Planning in Ad Hoc Teamwork

William Macke, Reuth Mirsky, Peter Stone

@ AAAI
February 2021

Learning Agents Research Group
The University of Texas at Austin

Ad Hoc Teamwork

“To create an autonomous agent that is able to efficiently and robustly **collaborate** with **previously unknown** teammates on tasks to which they are all individually capable of contributing as team members”.

Stone, P., Kaminka, G. A., Kraus, S., & Rosenschein, J. S. (2010, July). **Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination**. In AAAI Conference on Artificial Intelligence (p. 6).

Related Work

Are you playing offense?




- Ad Hoc Teamwork
  - Plastic Policy [Barrett and Stone, 2015]
  - Task Learning [Melo and Sardinha, 2016]
  - AATEAM [Chen et al., 2020]
- Communication
  - Dec-POMDP [Goldman and Zilberstein, 2004]
  - Human Feedback with RL [Griffith et al., 2013]
  - Multiagent RL for giving advice [Lowe et al., 2017]
- So far, work in the intersection has largely been focused on non-sequential problems
  - Ad hoc multi-agent multi-armed bandit with shared observations [Barrett et al., 2014]
  - Ad hoc teamwork for exploration in multi-agent multi-armed bandits [Chakraborty et al., 2017]

Communication in Ad Hoc Teamwork (CAT)

Previously unknown teammates ≠ solo player

Use existing communication channels

Learn new communication channels

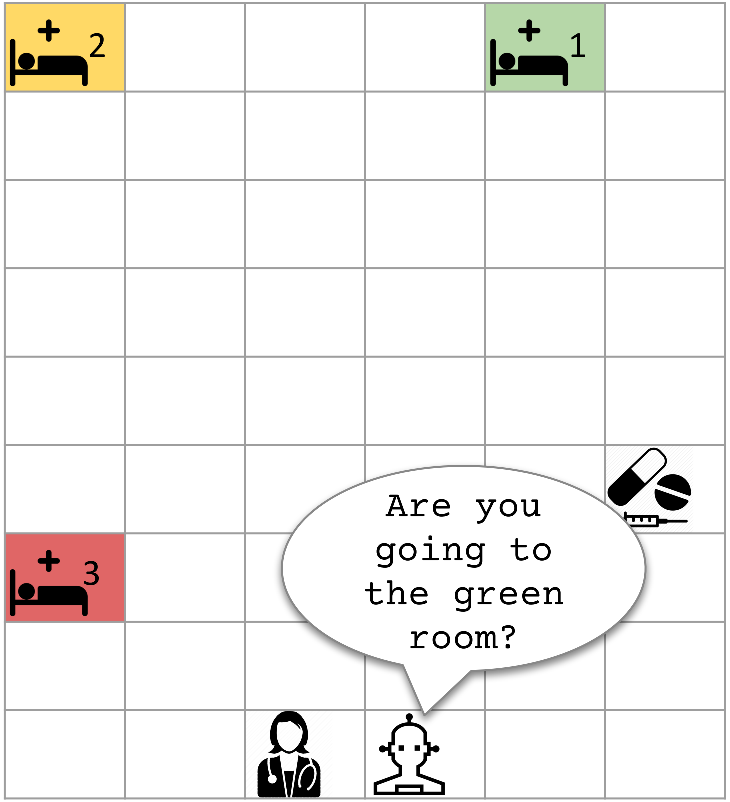
Teach other teachable agents

Ackerman, E. "**Moxi Prototype from Diligent Robotics Starts Helping Out in Hospitals**." *IEEE Spectrum. https://spectrum.ieee.org/automaton/robotics/industrial-robots/moxi-prototype-from-diligent-robotics-starts-helping-out-in-hospitals* (2018).

SOMALI CAT

Sequential
One-shot
Multi-
Agent
Limited
Inquiry

Communication
in
Ad hoc
Teamwork



When to Communicate

Zone of Information (Z_I)

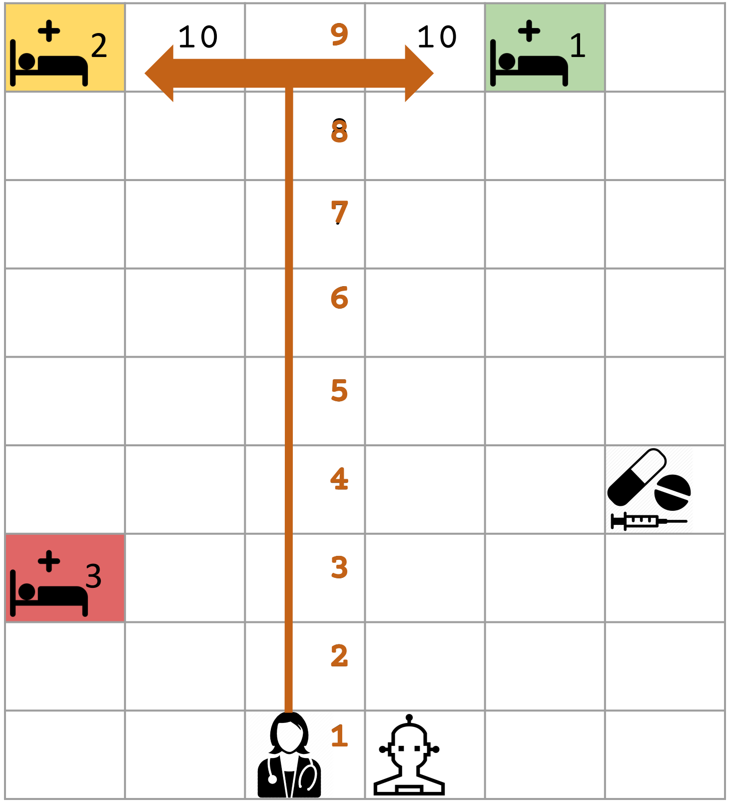
Ad hoc agent may have

Uncertainty about the physician's goal

worst cases distinctiveness (wcd)[1]

$Z_I = \{t \mid t \leq wcd_T(i,j)\}$

$Z_I(1,2) = 1\text{-}9$



1. Keren, S., Gal, A., and Karpas, E. (2014). **Goal recognition design**. In ICAPS.

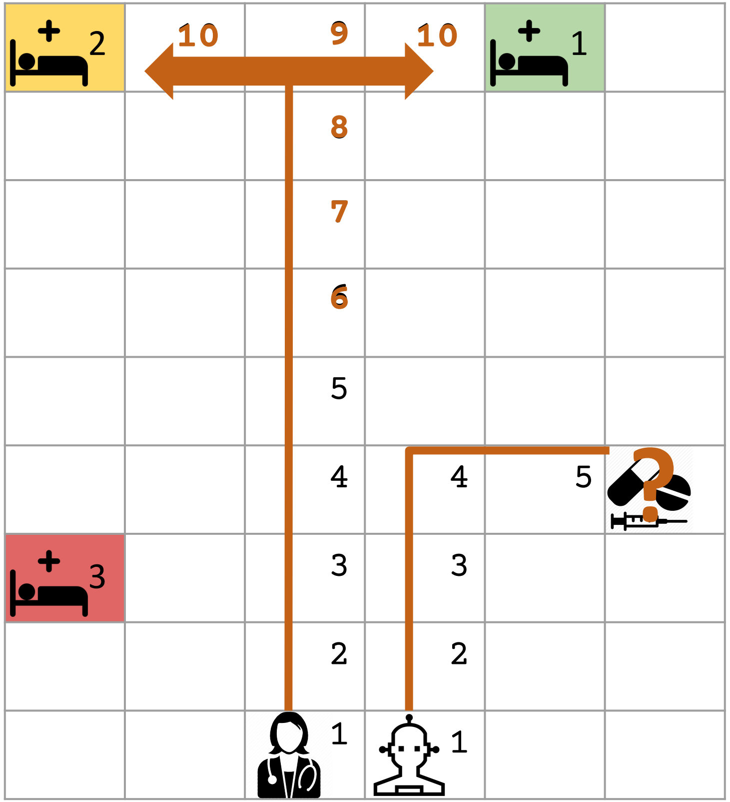
When to Communicate

Zone of Plan Branching (Z_B)

Ad hoc agent must commit to goal

$Z_B = \{t \mid t \geq wcd_A(i,j)\}$

$Z_B(1,2) = 6\text{-}10$



When to Communicate

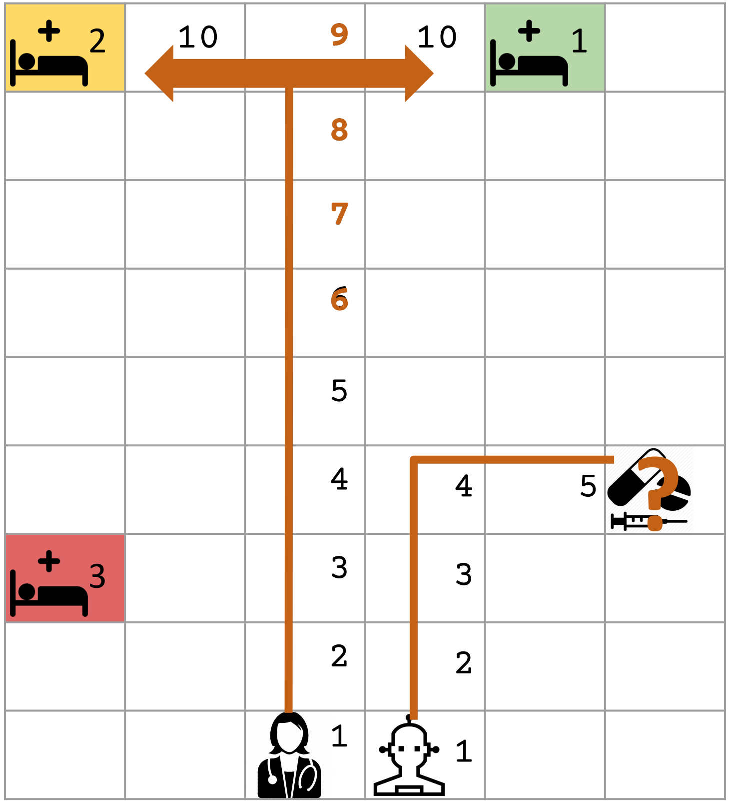
Query when:

Ad hoc agent both is uncertain and must commit

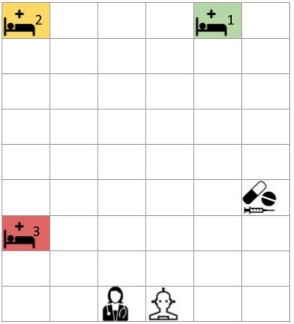
Zone of Querying

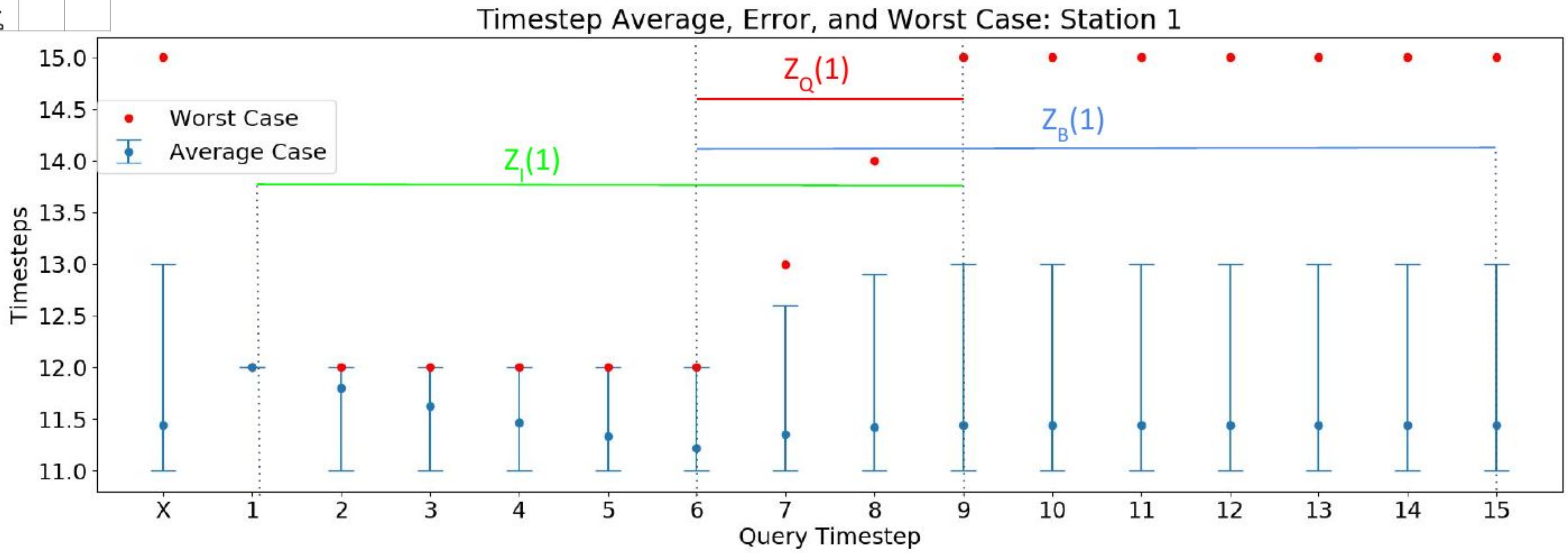
$$Z_Q = Z_I \cap Z_B$$
$$Z_Q = \{6, 7, 8, 9\}$$

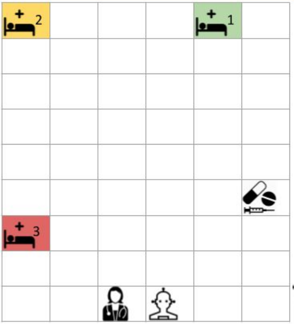
Critical Querying Point(CQP) $= 6$



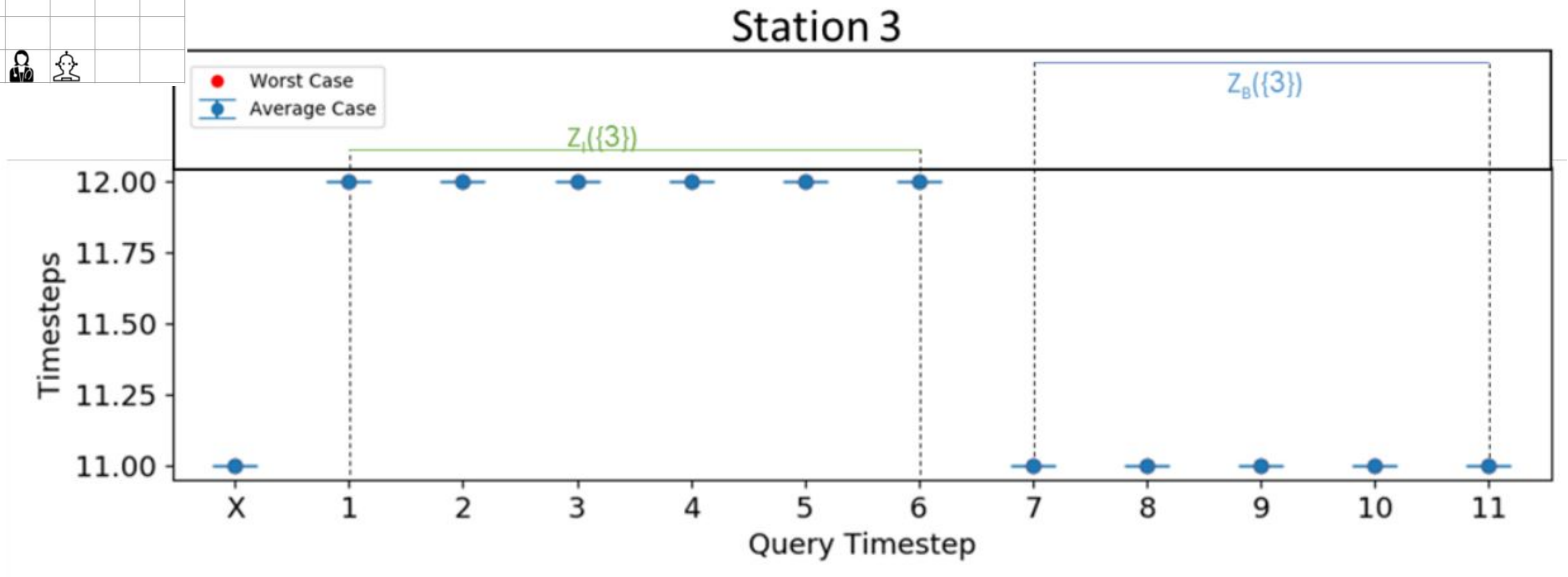
Methodical evaluation





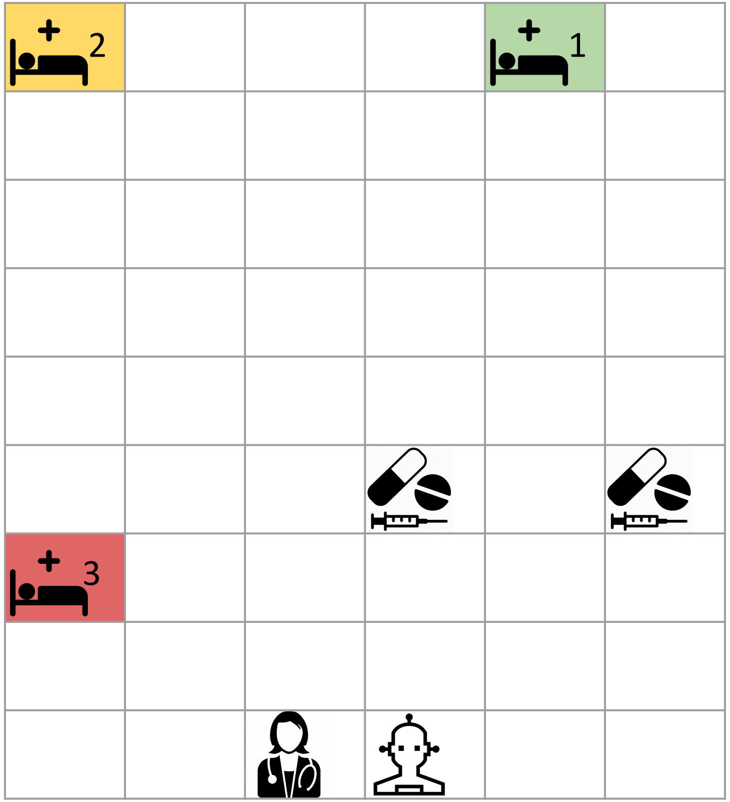


Methodical Evaluation



Generalizations to Domain

- Additional Medicine Cabinets
- Different Probability Distributions
- Different Query Cost Models
  - Cost = base_cost + |Q|*per_station_cost



Expected Divergence Point

Divergence Point (dp)

$dp(\pi \mid Tr) = \min\{t \mid \pi(a_t \mid s_t) = 0)\}$

Expected Divergence Point (EDP)

$EDP(s,\pi_1 \mid \pi_2) = E_{Tr\sim\pi 2}[dp(\pi_1 \mid Tr)]$

8	6 / 10	5.57 / 9	5.14 / 8	4.71 / 7	4.29 / 6	3.86 / 5	3.43 / 4	3 / 3
7	4.33 / 9	4 / 8	3.67 / 7	3.33 / 6	3 / 5	2.67 / 4	2.33 / 3	2 / 2
6	2.4 / 8	2.2 / 7	2 / 6	1.8 / 5	1.6 / 4	1.4 / 3	1.2 / 2	g_1
5	2.75 / 4.5	2.5 / 4	2.25 / 3.5	2 / 3	1.75 / 2.5	1.5 / 2	1.25 / 1.5	1 / 1
4	3.33 / 3.33	3 / 3	2.67 / 2.67	2.33 / 2.33	2 / 2	1.67 / 1.67	1.33 / 1.33	1 / 1
3	4.5 / 2.75	4 / 2.5	3.5 / 2.25	3 / 2	2.5 / 1.75	2 / 1.5	1.5 / 1.25	1 / 1
2	8 / 2.4	7 / 2.2	6 / 2	5 / 1.8	4 / 1.6	3 / 1.4	2 / 1.2	g_2
1	9 / 4.33	8 / 4	7 / 3.67	6 / 3.33	5 / 3	4 / 2.67	3 / 2.33	2 / 2
	1	2	3	4	5	6	7	8

$$EDP(s,\pi_1 \mid \pi_2) = [1-\Sigma_a \pi_2(a \mid s)] + \Sigma_a \pi_2(a \mid s) * \Sigma_{s'} T(s,a,s')[1 + EDP(s',\pi_1 \mid \pi_2)]$$

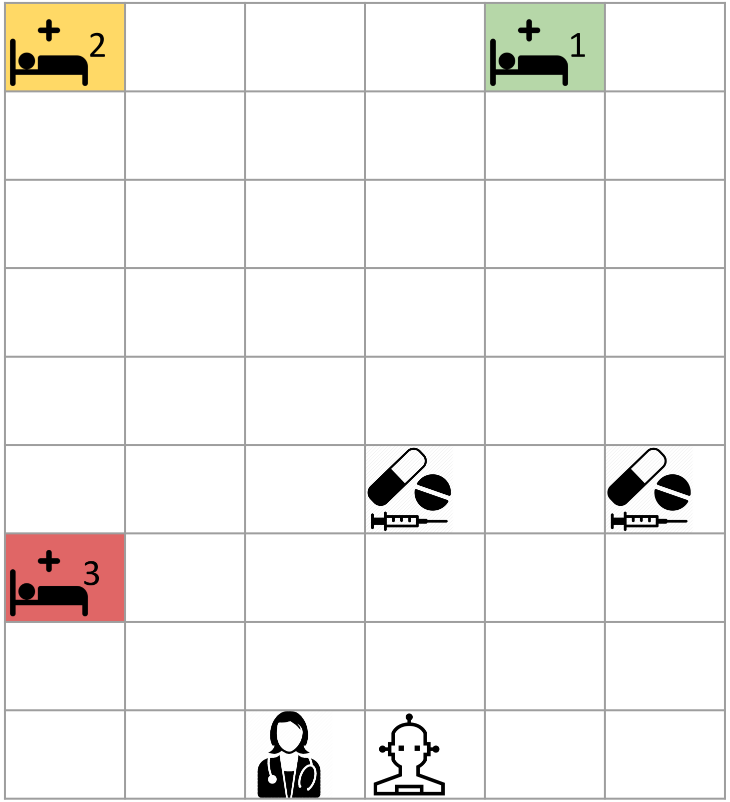
Value of a Query

Expected Zone of Information (eZ_I)
Expected Zone of Plan Branching (eZ_B)
Expected Zone of Querying (eZ_Q)

Marginal Cost (MC) - cost over if the robot knew physician's goal at start

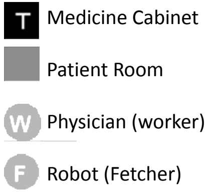
$E_P(MC_g(p)) \propto |U_{g'} eZ_Q(s,g'|g)|$

Expected Value = $E_g[V(q)|g]$

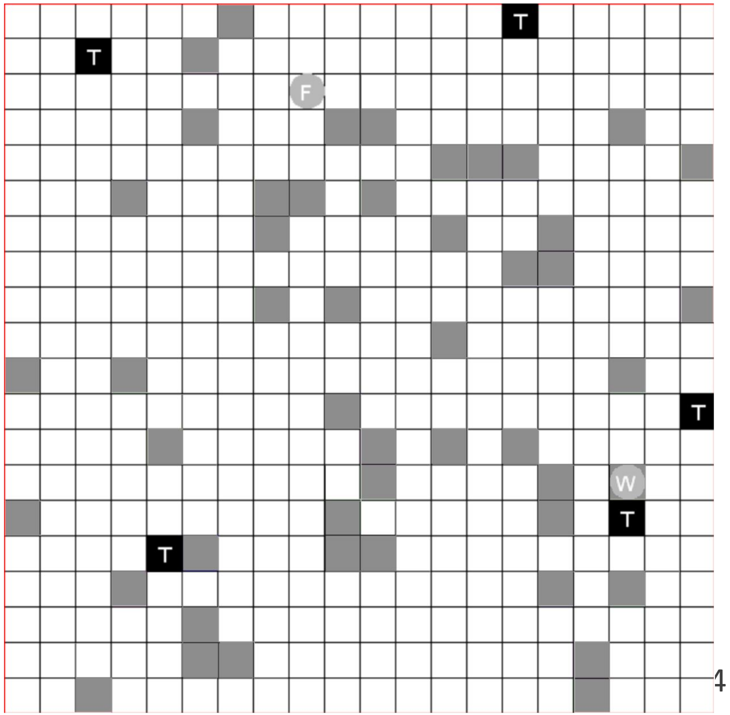


Experimental Setup

- Baseline
  - Asks randomly in zone of querying
- BL:Cost+Prob
  - Uses heuristic metric for value
- BL:Toolbox
  - Asks about toolboxes, then follows baseline
- eZ_Q Query
  - Optimizes for principled value of query

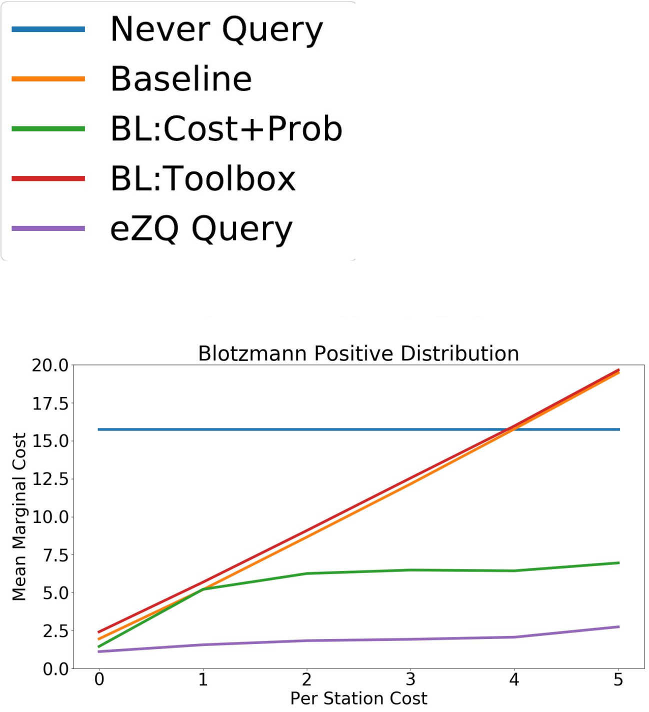


20x20 grid with 50 workstations



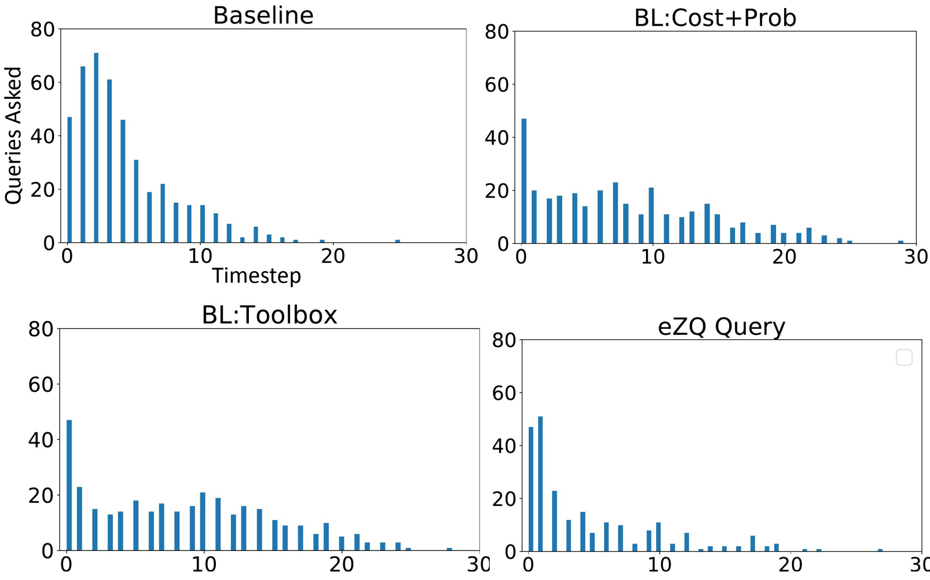
Results

- Baseline
  - Asks randomly in zone of querying
- BL:Cost+Prob
  - Uses heuristic metric for value
- BL:Toolbox
  - Asks about toolboxes, then follows baseline
- eZ_Q Query
  - Optimizes for principled value of query



Analysis

- Baseline
  - Asks randomly in zone of querying
- BL:Cost+Prob
  - Uses heuristic metric for value
- BL:Toolbox
  - Asks about toolboxes, then follows baseline
- eZ_Q Query
  - Optimizes for principled value of query



Limitations and Future Work

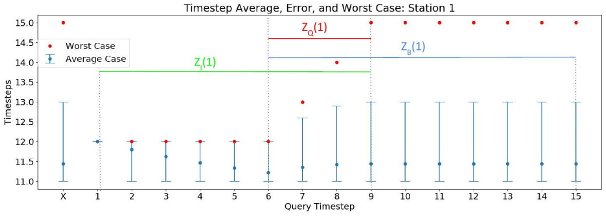
- Currently this method requires the use of the **policy evaluation** algorithm to calculate EDP, which is **expensive**
- Policy evaluation also requires **perfect knowledge** of teammate’s **policy** per goal
- Leverage more sophisticated **Reinforcement Learning** techniques to more efficiently learn EDP
- Learn directly from teammate **observations** without perfect knowledge of the teammates behavior per goal
- Experiment with more **diverse and complicated domains** that have a larger state space

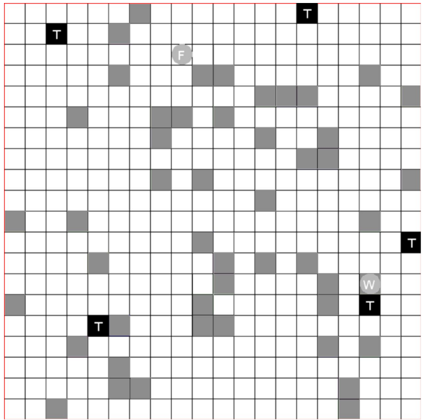
Summary

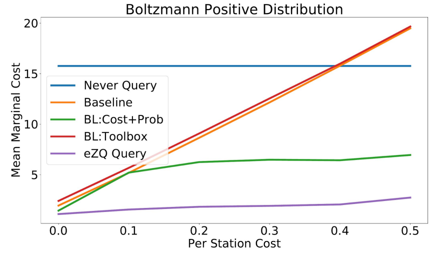
- We extend existing metrics on distinctiveness to include the **expected case**
- We introduce a principled method for the **value of query**
- We **evaluate** the metric and **analyze** its performance

Would you like to hear more?

- https://www.cs.utexas.edu/~pstone/Papers/bib2html-links/AAAI21-Macke.pdf







- **William Macke**, wmacke@cs.utexas.edu

  Reuth Mirsky, reuth@cs.utexas.edu

  Peter Stone, pstone@cs.utexas.edu