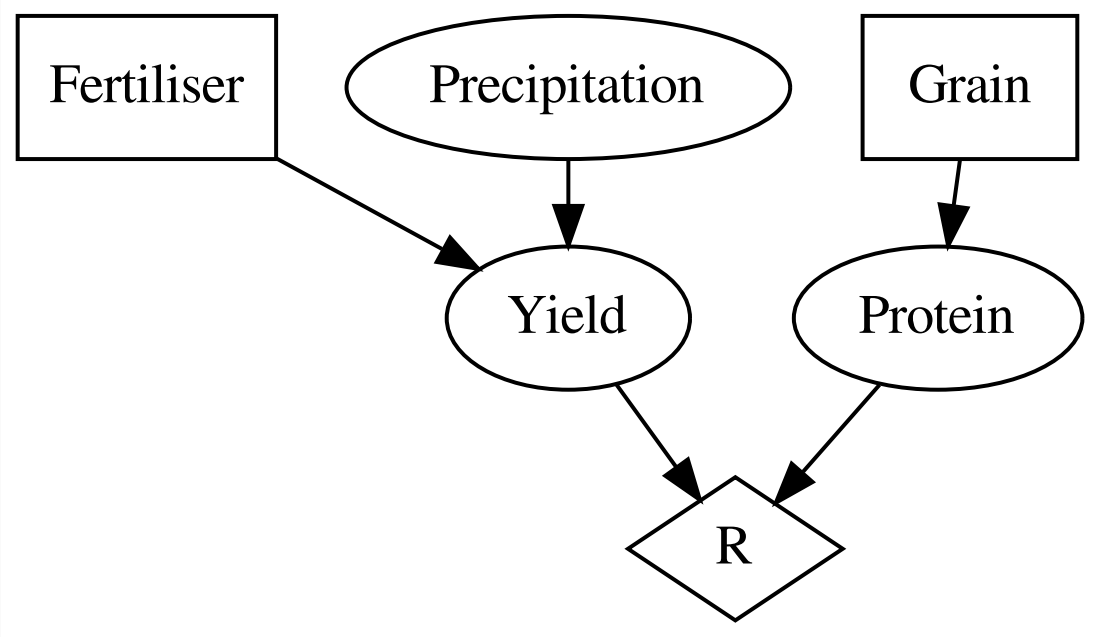
Learning Structured Decision Problems with Unawareness

Craig Innes (craig.innes@ed.ac.uk),
Alex Lascarides (alex@inf.ed.ac.uk)

Institute for Language, Cognition and Computation
University of Edinburgh

Why Unawareness?



$$\mathcal{X} = \{Prec, Protein, Yield\}$$

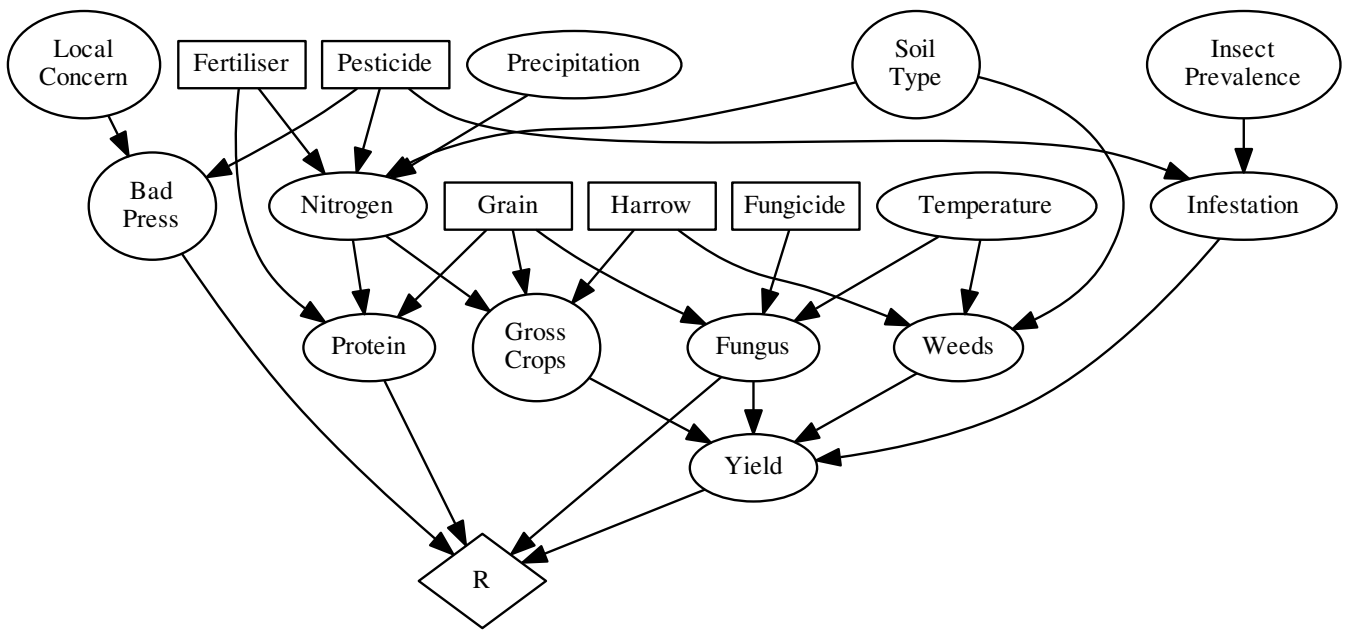
$$\mathcal{A} = \{Grain, Fert\}$$

$$scope(\mathcal{R}) = \{Yield, Protein\}$$

$$Pa_{Prot} = \{Grain\}$$

$$P(Prot = p|Grain = g) = \theta_{p|g}$$

Why Unawareness?



$$\mathcal{X}^0 \subseteq \mathcal{X}^+$$

$$\mathcal{A}^0 \subseteq \mathcal{A}^+$$

$$scope_0(\mathcal{R}) \subseteq scope_+(\mathcal{R})$$

$$Pa_{Prot} = \{Grain\}$$

$$P(Prot = p|Grain = g) = \theta_{p|g}$$

Contributions

Our agent learns an **interpretable model** of a decision problem **incrementally** via evidence from **domain trials** and **expert advice**.

Contributions

Our agent learns an **interpretable model** of a decision problem **incrementally** via evidence from **domain trials** and **expert advice**.

Evidence may reveal actions/variables the agent was **completely unaware of** *prior to learning.*

Contextual Advice

Types of Advice

1. **Advice on Better Actions**
2. Resolving Misunderstandings
3. Unexpected Rewards
4. Unknown Effects

Contextual Advice - Better Action

If agent's performance in last k trials is **below threshold** β of true policy π_+, say:

Contextual Advice - Better Action

If agent's performance in last k trials is **below threshold** β of true policy π_+, say:

"At time t you should have done $a' = \langle A_1 = 0, A_2 = 1, A_3 = 0 \rangle$ rather than a_t"

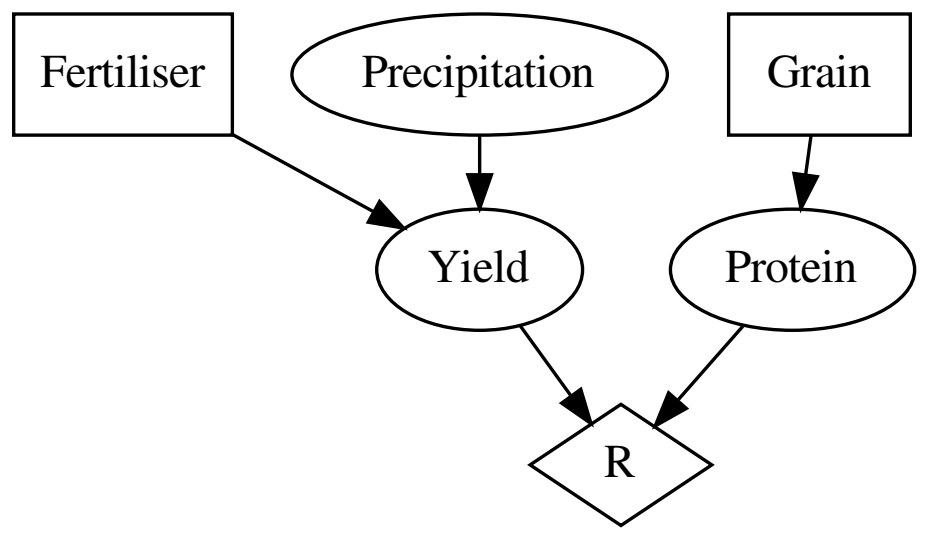
Contextual Advice - Better Action

If agent's performance in last k trials is **below threshold** β of true policy π_+, say:

"At time t you should have done $a' = \langle A_1 = 0, A_2 = 1, A_3 = 0 \rangle$ rather than a_t"

- Action variable A_3 is part of the problem $(A_3 \in \mathcal{A})$
- A_3 is *relevant* $(\exists X \in scope(\mathcal{R}), anc(A_3, X))$
- There exists a better reward $(\exists s, s[\mathcal{B}^t] = s_t[\mathcal{B}^t] \wedge \mathcal{R}_+(s) > r_t)$
- a' has a greater expected utility than a_t
  $(EU(a'|s) > EU(a_t|s))$

Conserving Previous Beliefs



$$P(\mathrm{Pa}_{Yield} | D_{0:t})$$

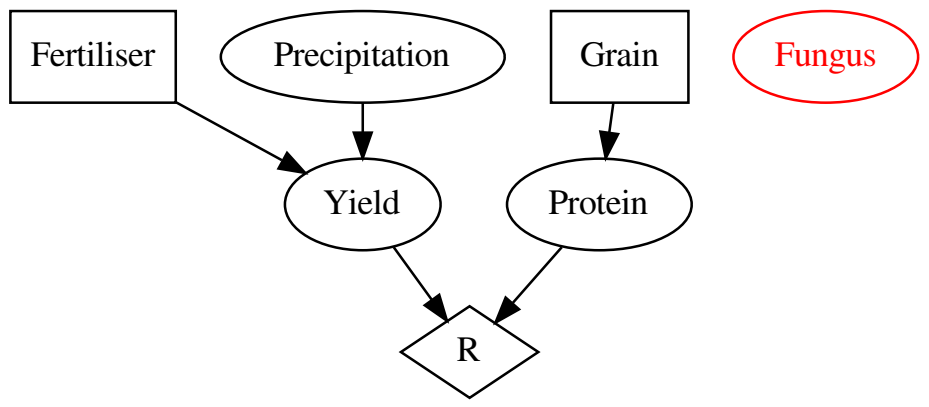
$$Pa_{Yield} = \emptyset$$

$$Pa_{Yield} = \{Fert\}$$

$$\vdots$$

$$Pa_{Yield} = \{Fert, Prec, Grain\}$$

Conserving Previous Beliefs



$$P(\mathbf{Pa}_{Yield} | D_{0:t})$$

$$Pa_{Yield} = \emptyset$$

$$Pa_{Yield} = \{Fungus\}$$

$$Pa_{Yield} = \{Fert\}$$

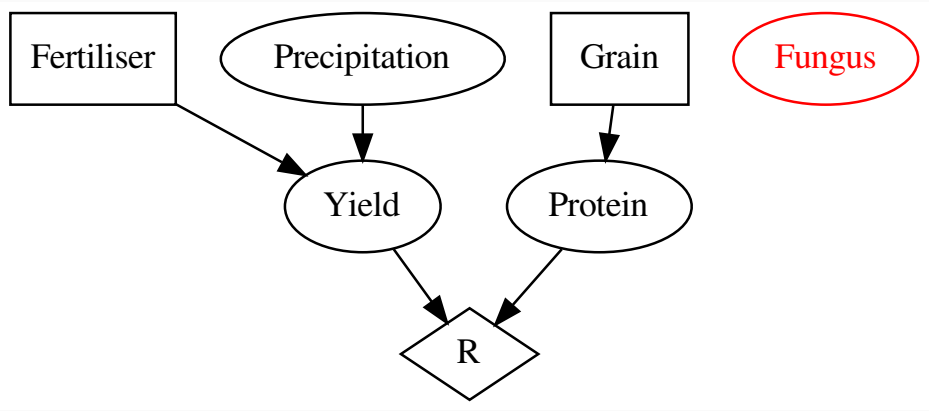
$$Pa_{Yield} = \{Fert, Fungus\}$$

$$\vdots$$

$$Pa_{Yield} = \{Fert, Prec, Grain\}$$

$$Pa_{Yield} = \{Fert, Prec, Grain, Fungus\}$$

Conserving Previous Beliefs



$$P(\mathbf{Pa}_{Yield}|D_{0:t})$$

$Pa_{Yield} = \emptyset$

$Pa_{Yield} = \{Fungus\}$

$Pa_{Yield} = \{Fert\}$

$Pa_{Yield} = \{Fert, Fungus\}$

$\vdots$

$Pa_{Yield} = \{Fert, Prec, Grain\}$

$Pa_{Yield} =$

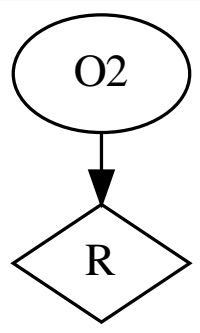
$\{Fert, Prec, Grain, Fungus\}$

$$P_{new}(Pa_X) = \begin{cases} (1-\rho)P_{old}(Pa_X|D_{0:t}) & \text{if } Fungus \notin Pa_X \\ \rho P_{old}(Pa'_X|D_{0:t}) & \text{if } Pa_X = Pa'_X \cup \{Fungus\} \end{cases}$$

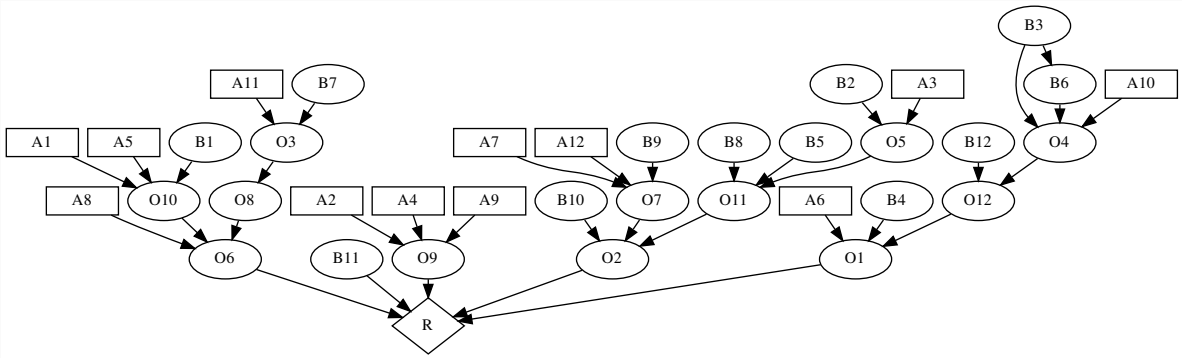
Experiments

Randomly Generated Networks: 12 - 36 Variables

- 12 - 36 Variables
- 3000 Trials
- ϵ-greedy strategy
- Expert Aid $\beta = 0.1$

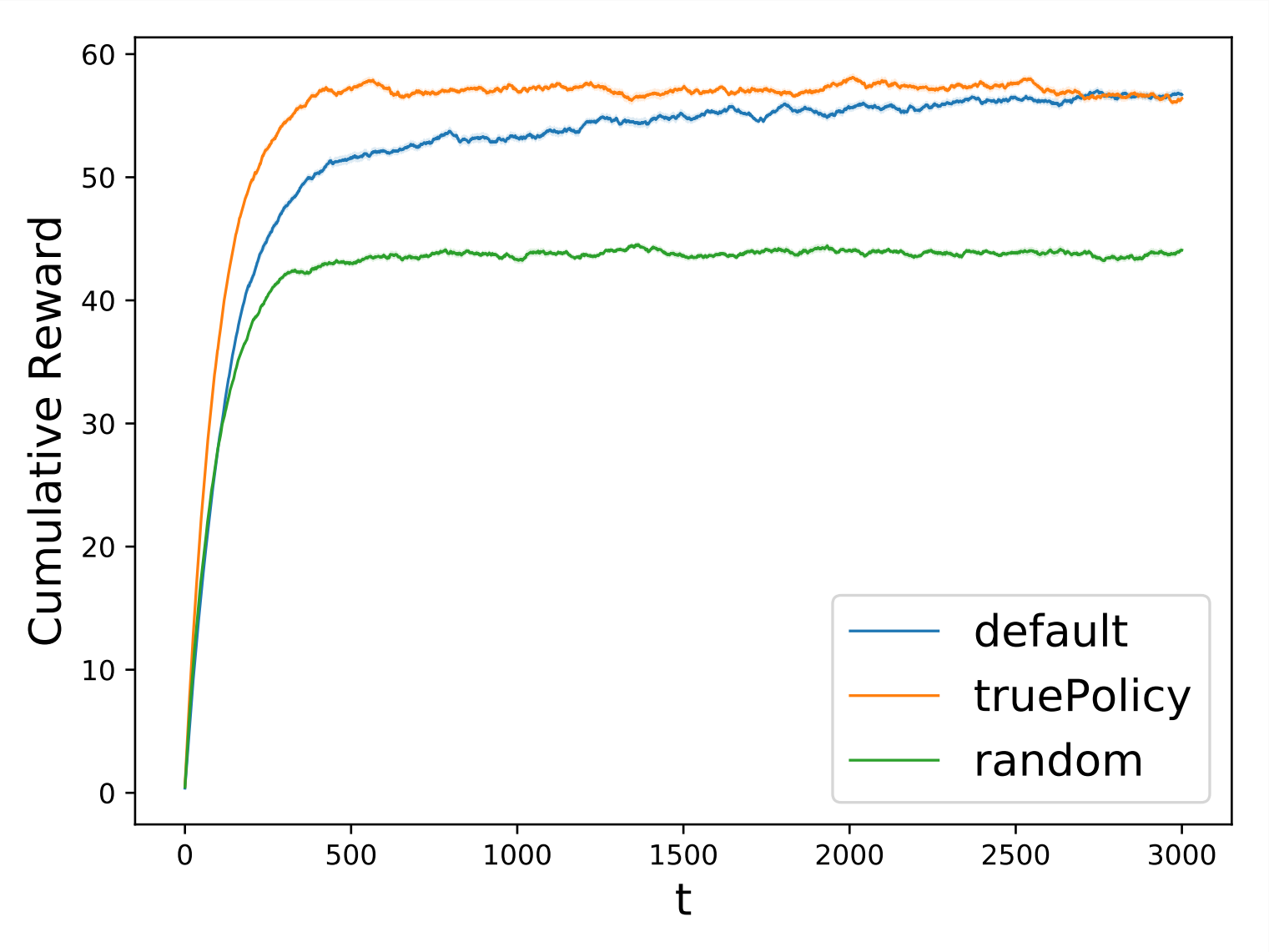


Start

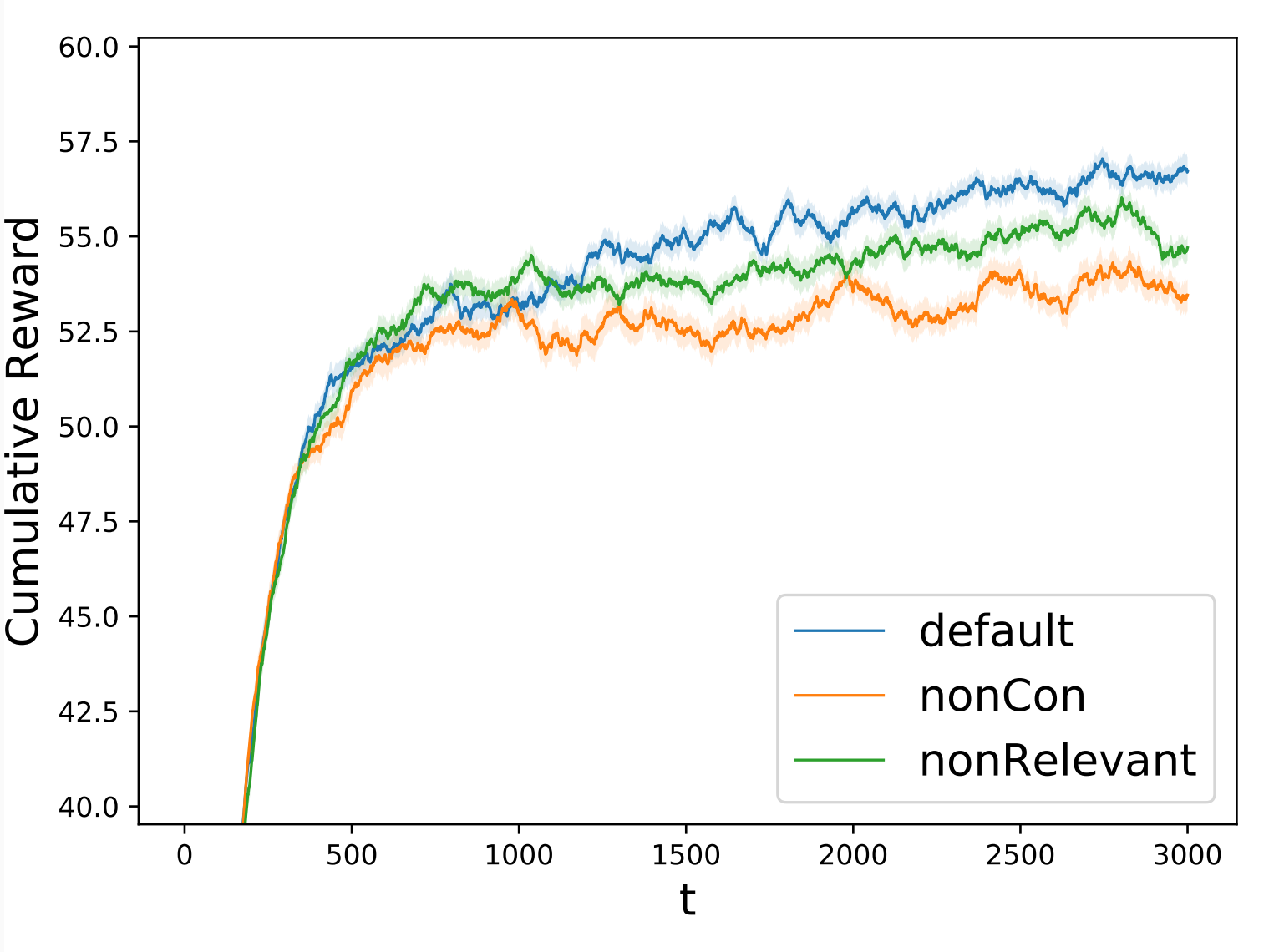


Learning Goal

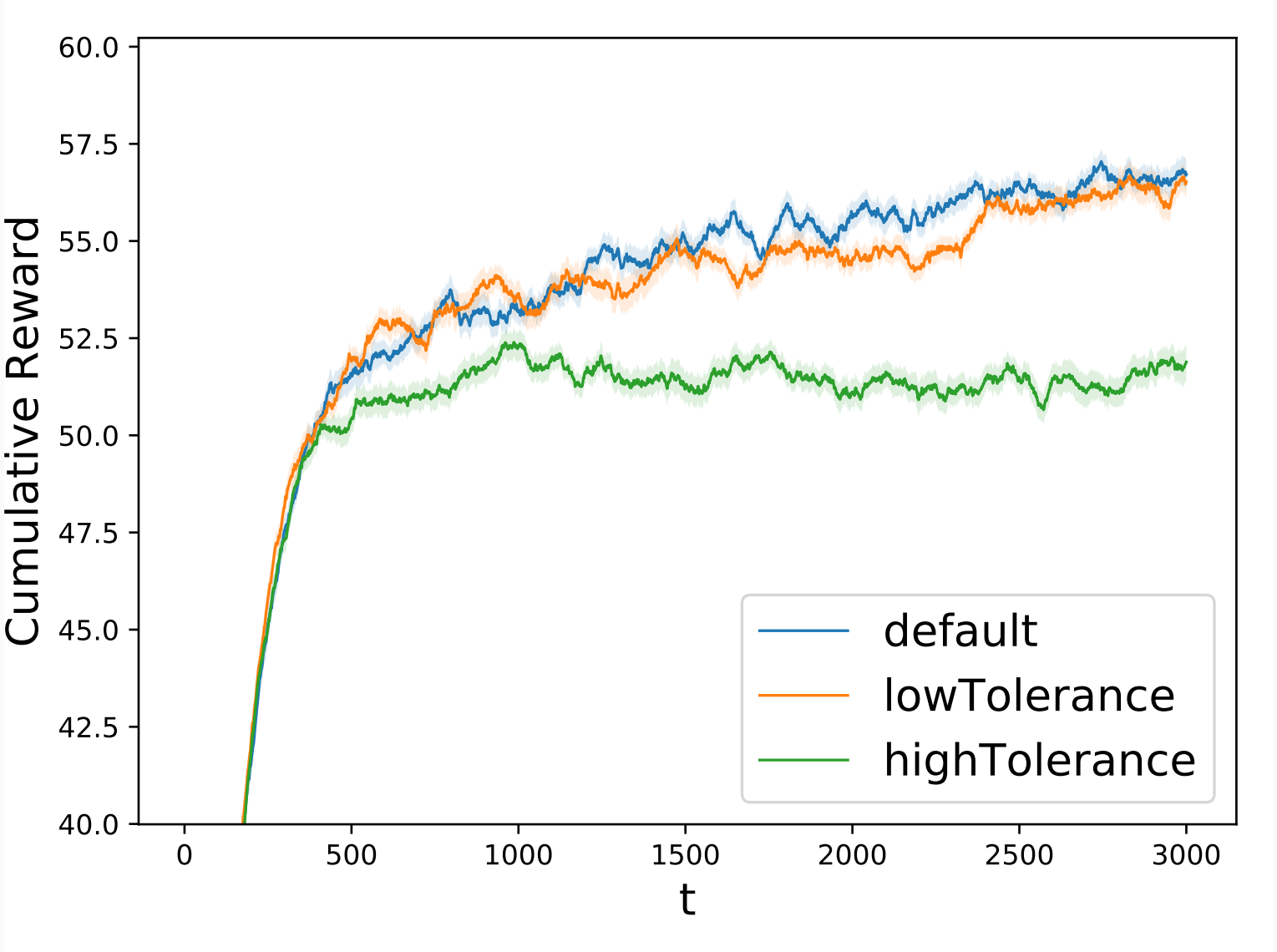
Results


Cumulative Reward
t
0
10
20
30
40
50
60
0
500
1000
1500
2000
2500
3000
default
truePolicy
random

Results


Cumulative Reward
60.0
57.5
55.0
52.5
50.0
47.5
45.0
42.5
40.0
0
500
1000
1500
2000
2500
3000
t
default
nonCon
nonRelevant

Results


60.0
57.5
55.0
52.5
50.0
47.5
45.0
42.5
40.0
Cumulative Reward
0
500
1000
1500
2000
2500
3000
t
default
lowTolerance
highTolerance

Conclusions and Contact Details + Paper Link

Paper

Learning Structured Decision Problems with Unawareness

Authors

Craig Innes (craig.innes@ed.ac.uk)

Alex Lascarides (alex@inf.ed.ac.uk)

Poster Session:

6:30pm-9pm, Pacific Ballroom #35