Norm Monitoring with Asymmetric Information

Jiaqi Li (University of Nottingham)
Felipe Meneguzzi (PUCRS)
Moser Silva Fagundes (IFSul)
Brian Logan (University of Nottingham)

Normative MAS

- norms have been widely proposed as a way of coordinating and regulating the behaviour of agents in a multi-agent system

- in a normative MAS, interaction between agents and their environment is governed by a *normative organisation* specified by a set of norms
    - an **obligation** requires an agent to bring about a particular state of the environment
    - a **prohibition** requires the agent to avoid bringing about a particular state

- if an agent fails to meet an obligation or violates a prohibition, the organisation imposes a **sanction** on the agent

Normative organisation

- continuously evaluates the state updates resulting from agent actions with respect to the norms to
    - determine any new obligations to be fulfilled or prohibitions that should not be violated
    - check if any previously detached norms are obeyed or violated in the current state
    - impose sanctions when norms are violated
- this continuous process is implemented by a *normative control cycle*

Norm-aware agency

- when norms conflict with an agent's existing goals, a self-interested agent must choose between its goals and the norms imposed by the normative organisation

- an agent is **norm-aware** if it can deliberate on its goals, norms and sanctions before deciding which plan to select and execute

- a norm-aware agent is able to *violate* norms (accepting the resulting sanctions) if it is in the agent's overall interests to do so

- e.g., if meeting an obligation would result in an important goal of the agent becoming unachievable

Normative MAS assumptions

- previous work on normative MAS has generally relied on two key assumptions:
  - norm monitoring and enforcement are perfect
  - agents are fully aware of the monitoring capabilities of the normative organisation

Examples

- when reasoning about whether a set of norms guarantees some desirable system-level behaviour, it is assumed that the **monitoring and sanctioning** capabilities of the normative organisation are **perfect**

- in much of the work on norm-aware agency, the **agents implicitly assume** that **all norm violations will be detected**, and choose an ‘optimal’ course of action based on this assumption

In reality …

- for large-scale MAS **perfect monitoring** is likely to be either **costly or impossible**
    - probability of detecting norm violation (**enforcement intensity**) is likely to be less than 1
        - complete **information about the enforcement intensity** employed by the normative organisation is not available to the agents at zero cost
- there is an **information asymmetry** between the normative organisation and the the agent(s)
- agents must either assume an enforcement intensity or learn it

Estimating enforcement intensity

- if an agent makes an incorrect *assumption* about the enforcement intensity of a norm, its **'optimal' policy may not be optimal** with respect to the norm

    - i.e., it could increase its utility by violating fewer norms or more norms, depending on whether the enforcement intensity is higher or lower than it assumes

- alternatively, a learning agent can *induce* the enforcement intensity and compute an optimal policy without prior knowledge of the enforcement intensity

- however learning has a cost

Example: Parking World

- 5 x 5 grid of cells – (1, 1) is the start state, and cell (5, 5) is the end state
- the agent can move from cell to cell orthogonally
- environment also contains two special cells in which it can ‘park’
- a **‘legal’ parking cell**: parking in the legal cell gives a small reward (20)
- an **‘illegal’ parking cell** where parking is prohibited: parking in the illegal cell has a higher reward (50), but the agent may incur a sanction (-100) if the violation of the no parking norm is detected

Parking World: rewards

-4	-4	-4	-4	+100 END
-4	+20	-4	-4	-4
-4	-4	-4	-4	-4
-4	-4	-4	+50 -100(D)	-4
-4 START	-4	-4	-4	-4

rewards before parking

-4	-4	-4	-4	+100 END
-4	-4	-4	-4	-4
-4	-4	-4	-4	-4
-4	-4	-4	-4 -100(D)	-4
-4 START	-4	-4	-4	-4

rewards after parking

Normative MDPs

- We model the Parking World as a normative MDP

- Rewards for norm compliant (no-parking) actions are constant:
  e.g., -4 for moving from cell to cell, +20 for parking legally

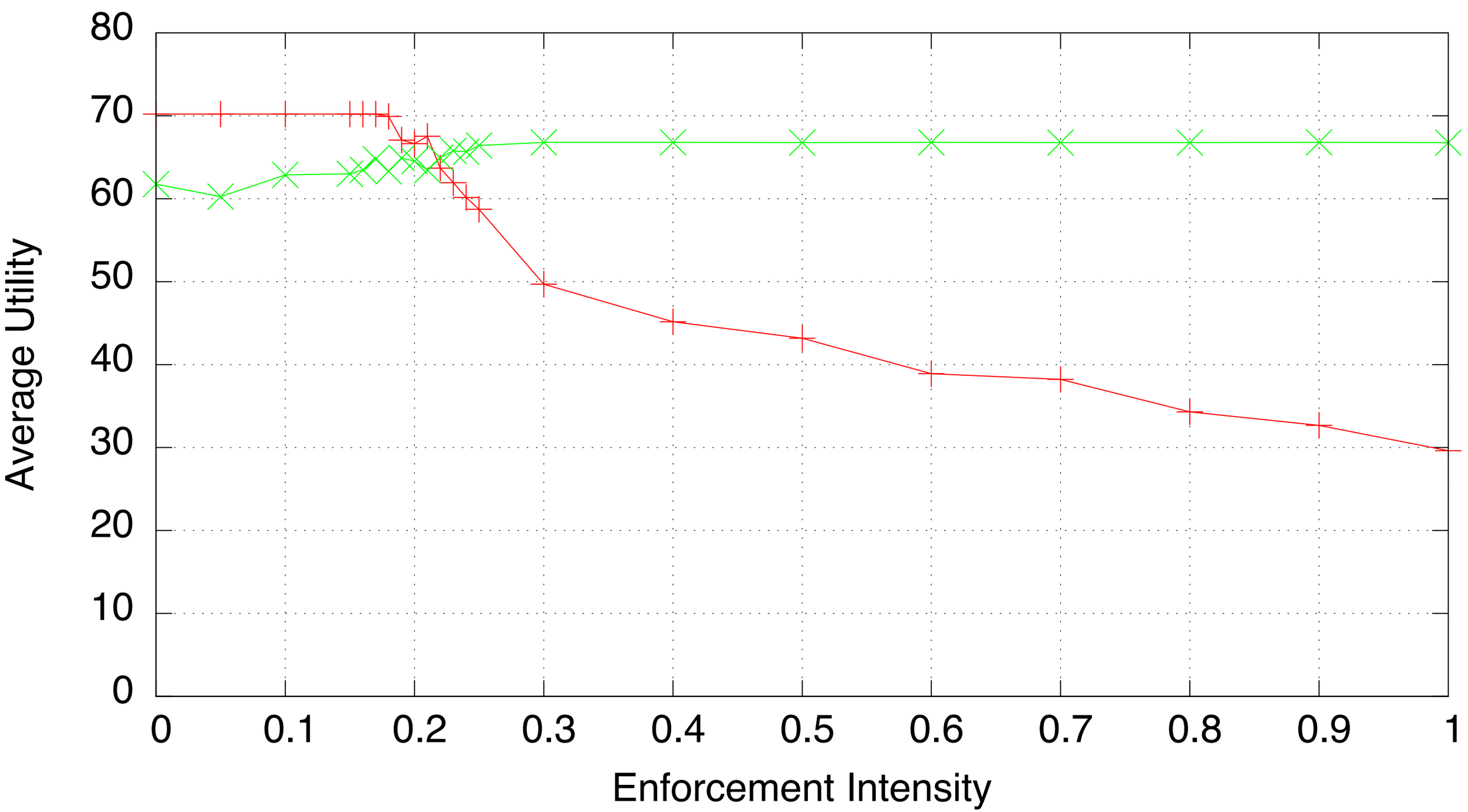
- Reward for violating the no-parking norm depends on the enforcement intensity, *e*:

  - with probability ***1 - e*** the agent obtains a reward of 50

  - with probability ***e***, the agent obtains a reward of -100 (a sanction)

Learning the enforcement intensity

- the optimal policy for an NMDP depends on the value of ***e***
    - agent chooses to park illegally when enforcement intensity is low (how low depends on the reinforcement learning algorithm)
- estimating ***e*** has a cost for the agent (in the form of sanctions)
    - how much depends on the exploration/exploitation tradeoff in learning



Example

SARSA learning agent

Implications of information asymmetry

- if the agent's policy **under-estimates** *e*,
    - it receives a clear signal that its policy is incorrect in the form of **(unexpected) sanctions** and a lower than expected reward
- an agent with a fixed (or slowly changing) policy that **over-estimates** *e*
    - receives no signal from the environment, and has no reason to change its policy
- **it will continue to act on its policy believing it to be correct**
    - in particular, its **degree of compliance** with the norm will be **higher** than an agent with perfect information

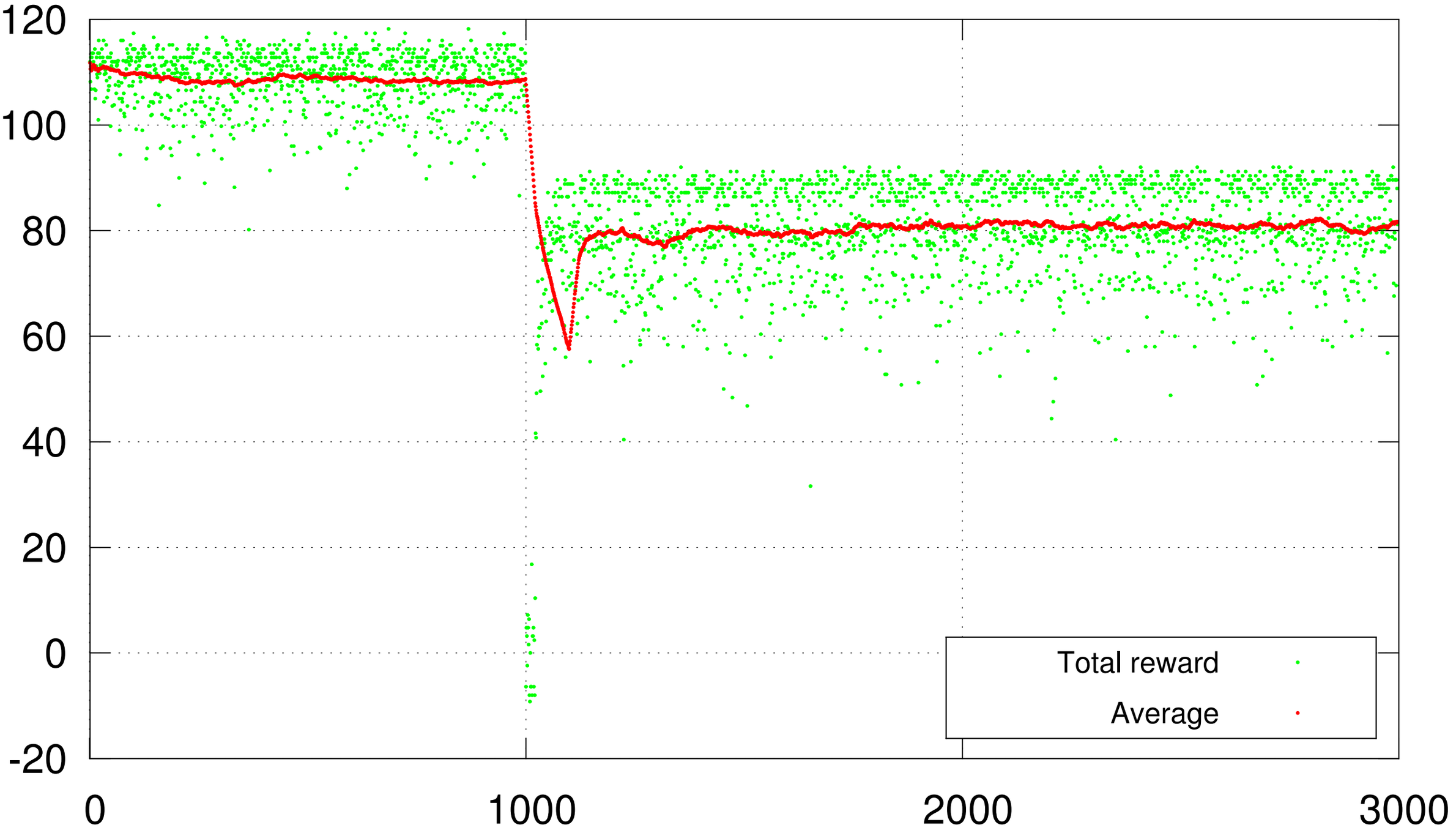
Exploiting information asymmetry

- **information asymmetry** can be **exploited** by the normative organisation to reduce the cost of monitoring and enforcing norms

- e.g., by i**ncreasing the enforcement intensity** when an agent enters the MAS, the normative organisation can cause the agent to **over-estimate** the enforcement intensity

- if the enforcement intensity is subsequently reduced, the agent continues to behave as if the **organisation** is **more effective** in monitoring norm violations than is actually the case

- holds even if the agents actively seek to learn the enforcement intensity



Example

SARSA learning agent (asymmetric)

Future work

- current model is very simple
  it ignores communication between agents

- which enforcement schedule(s) allow the normative organisation to **maximise information asymmetry**
  - can such schedules be learned by the normative organisation?

- are there cases where it is better for the agents to be uninformed?
  - e.g., if the agents benefit from norm compliance by other agents and the cost of enforcement is borne by the agents themselves, information asymmetry may actually benefit the agents

Questions?