Peer Pressure as a Driver of Adaptation in Agent Societies

Hugo Carr[1], Jeremy Pitt[1] and Alexander Artikis[21]

[1] Imperial College London

[2] National Centre for Scientic Research "Demokritos"

{h.carr,j.pitt}@imperial.ac.uk, a.artikis@iit.demokritos.gr

ESAW 2008, St Etienne, France, Sep 2008

Thanks to: UK EPSRC

EU FP6 Project 027958 ALIS

Background

- Characteristics of networks
  - open: agents are heterogeneous, may be competing, conflicting goals
  - fault-tolerant: agents may not conform to the system specification
  - volatile-tolerant: agents may come/go, join/leave the system
  - decentralised: there is no central control mechanism
  - partial: local knowledge, (possibly) inconsistent global union

- Agent Societies
  - Accountable governance, market economy, Rule of Law
  - Mutable: "tomorrow can be different from today"
  - Socio-cognitive relations: trust/forgiveness, gossiping

Motivation

- Resource allocation scenario where not all requirements can be satisfied
  - Common feature of e.g. ad hoc networks
- Two options:
  - Free for all: short-term gain, long-term annihilation
  - Do what people do: form committee, make up rules, . . .
- Previous work (OAMAS08)
  - Allocation according to vote, change the voting rules
  - Showed: population of ‘responsible’ agents stabilised the system
  - Now: given a stable system, show resistance to ‘selfish’ behaviour
  - Moreover: given a choice (responsible/selfish), agents ‘choose’ responsible (or have it chosen for them...)

How you gonna do that?

- Voting
  - voting about the rule
  - voting for each other
- Learning (individual behaviour)
- Reputation (individual opinion formation)
- Show that Organised Adaptation
  - is stable
  - is robust

Formal Model

- Let $\mathcal{M}$ be a multi-agent system (MAS) at time t

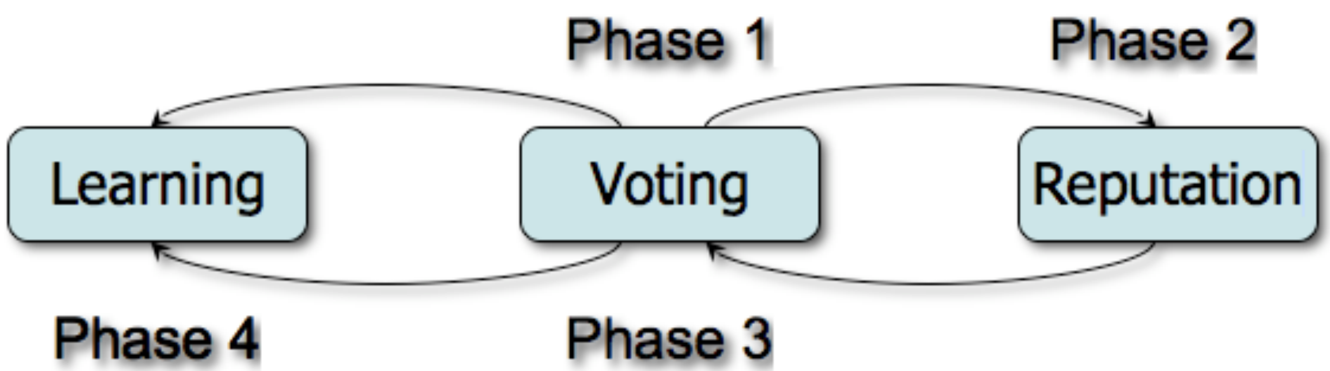
$$\mathcal{M}_t = \langle U, \langle A, \rho, B, \mathbf{f}, \tau \rangle_t \rangle$$

  - $U =$ the set of agents
  - $A_t \subseteq U$, the set of *present* agents at t
  - $\rho_t : U \rightarrow \{0, 1\}$, the presence function s.t. $\rho_t(a) = 1 \leftrightarrow a \in A_t$
  - $B_t : \mathbb{Z}$, the 'bank', indicating the overall system resources available
  - $\tau_t : \mathbb{N}$, the threshold number of votes to be allocated resources
  - $\mathbf{f}_t : A_t \rightarrow \mathbb{N}_0$

  The resource allocation function $\mathbf{f}_t$ determines who gets allocated resources according to the value of τ_t and the votes cast (see below)

Scenario

- System operation is divided into timeslices; during each timeslice, each 'present' agent a will
  - Phase 1: Vote for threshold value for τ (change a rule)
  - Phase 2: Offer (O^a)/Request (R^a) resources $(R^a > O^a)$
  - Phase 3: Vote for a candidate(s) to receive resources
  - Phase 4: Update its satisfaction and learning metrics with respect to the outcome of the vote



Phase 1: Voting for τ

- Tau (τ) represents the threshold number of votes required to receive resources (at time t)

$$\begin{aligned} \mathbf{f}_t(a) &= R_t^a, \mathbf{card}(\{b | b \in A_t \wedge \mathbf{v}_t^b(\ldots) = a\}) \geq \tau_t \\ &= 0, \text{otherwise} \end{aligned}$$

- The value of τ is context dependent and crucial for ‘collective well-being’
  - If τ is too low, too many resources will be distributed, and this will result in the “Tragedy of the Commons”
  - If τ is too high, too few resources will be distributed, and this will result in “Voting with your Feet” (satisfaction)

- Each timeslice t, two-round election
  - round 1: each present agent proposes a value for τ
  - round 2: run-off election between two most popular selections

Phase 2: Reputation Management

- Vote for τ is an indicator of selfish/responsible behaviour

- For experimentation, require a method that computes τ 'responsibly', supports discrimination, and isn't random
  - define a family of predictor functions, randomly initialised, a subset of which is given to each agents
  - functions which return 'good' value have increased weight

$$w_i = \frac{x_i}{\sum_{\forall j} x_j} \qquad pred_\tau = \sum_{i=0}^{j} w_i . a_i$$

- Agent uses other agents' τ-voting to update opinion of those agents

Phase 3: Voting to Allocate Resources

- Plurality Protocol in ineffective
  - Does not provide information to effectively judge selfish or responsible behaviour
  - Punishment in the form of lost votes is not sufficient motivation to behave responsibly

- Borda Protocol
  - Agents vote using preference lists derived from reputation score
  - Points are allocated based on ‘most preferred’
  - Agents are forced to give their opinion of their neighbours
    * Allows a participant to see more easily who is behaving responsibly or selfishly

Phase 4: Reinforcement Learning

- Used to demonstrate how an initially selfish agent can be 'rehabilitated' through peer pressure

- Unbiased evaluation of sets of actions

- A Q-Value is a metric which measures from a history of length m how successful an action x has been in a certain state s when each action is assigned a reward r

$$Q_{t+1}(s,x) = \frac{1}{m}\sum_{i=1}^{m}\left(r_{k_i} + \gamma V_{k_i}(s_{k_i})\right) + \epsilon$$

where

$$V_t = max_{x \in X} Q_t(s,x), r_k \in [0,1], \gamma \in [0,1]$$

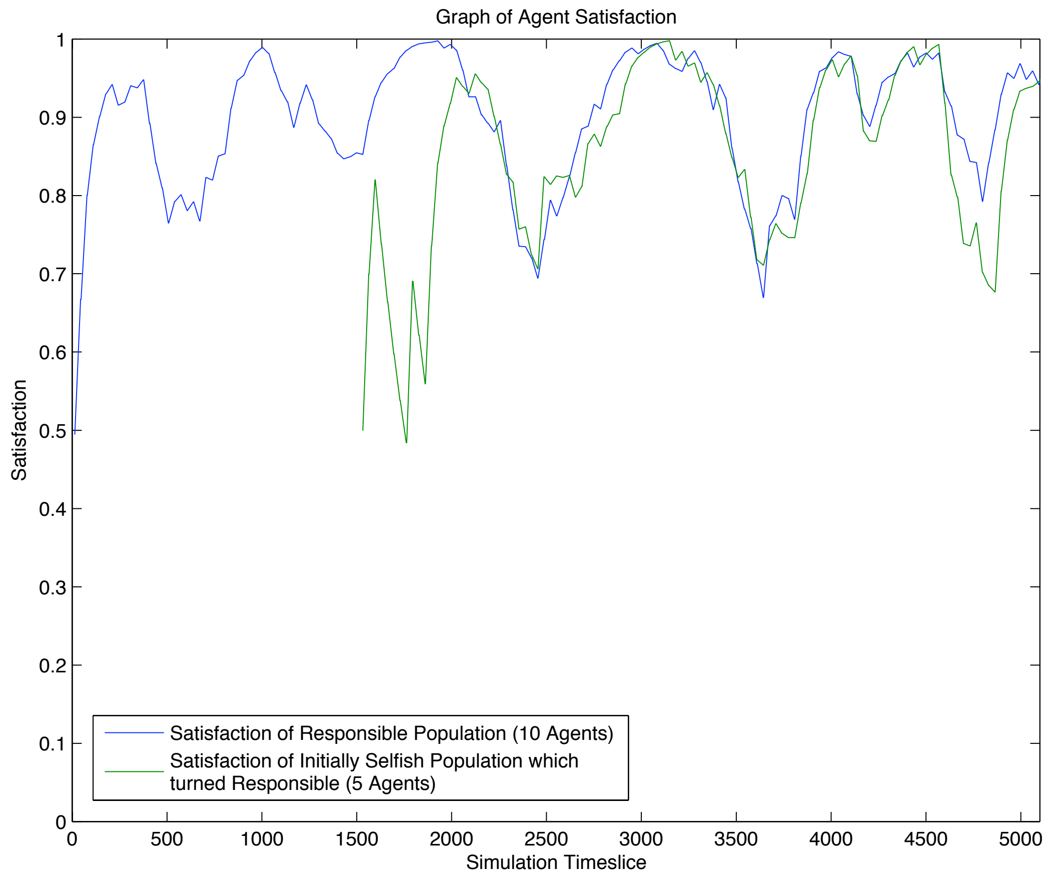
Experiment

- Initially we show that this experiment is stable amongst a group (size 10) of these agents who have already established a stable system

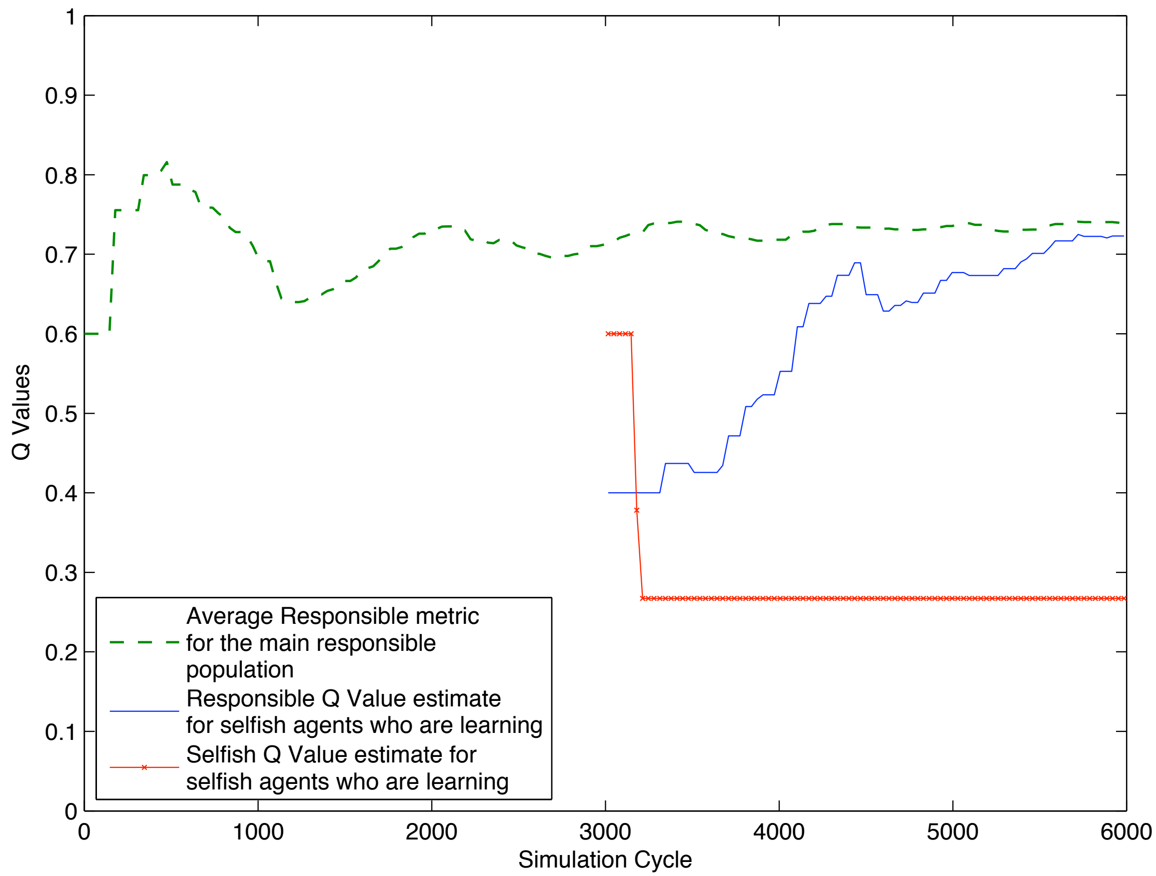
- We then add a destabilising element to the system at timecycle 3000 consisting of a set of agents (size 5) behaving selfishly
  - Agents who learn to behave responsibly are forgiven and assimilated into society
  - Agents who fail to learn are permanently ostracised and leave the system (through dissatisfaction)

- Use a certain 'well-known' MAS animator PreSAGE

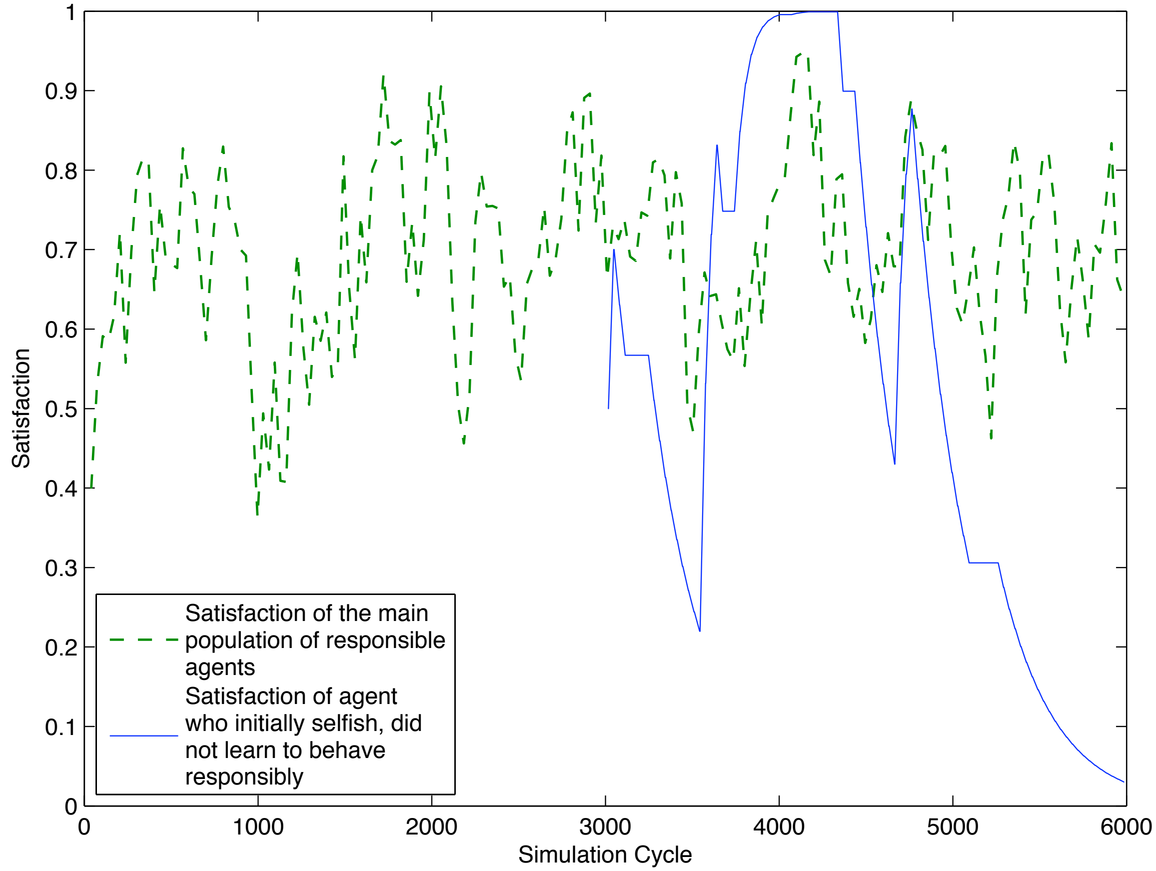
Results (1.1): Satisfaction for Responsible Agents



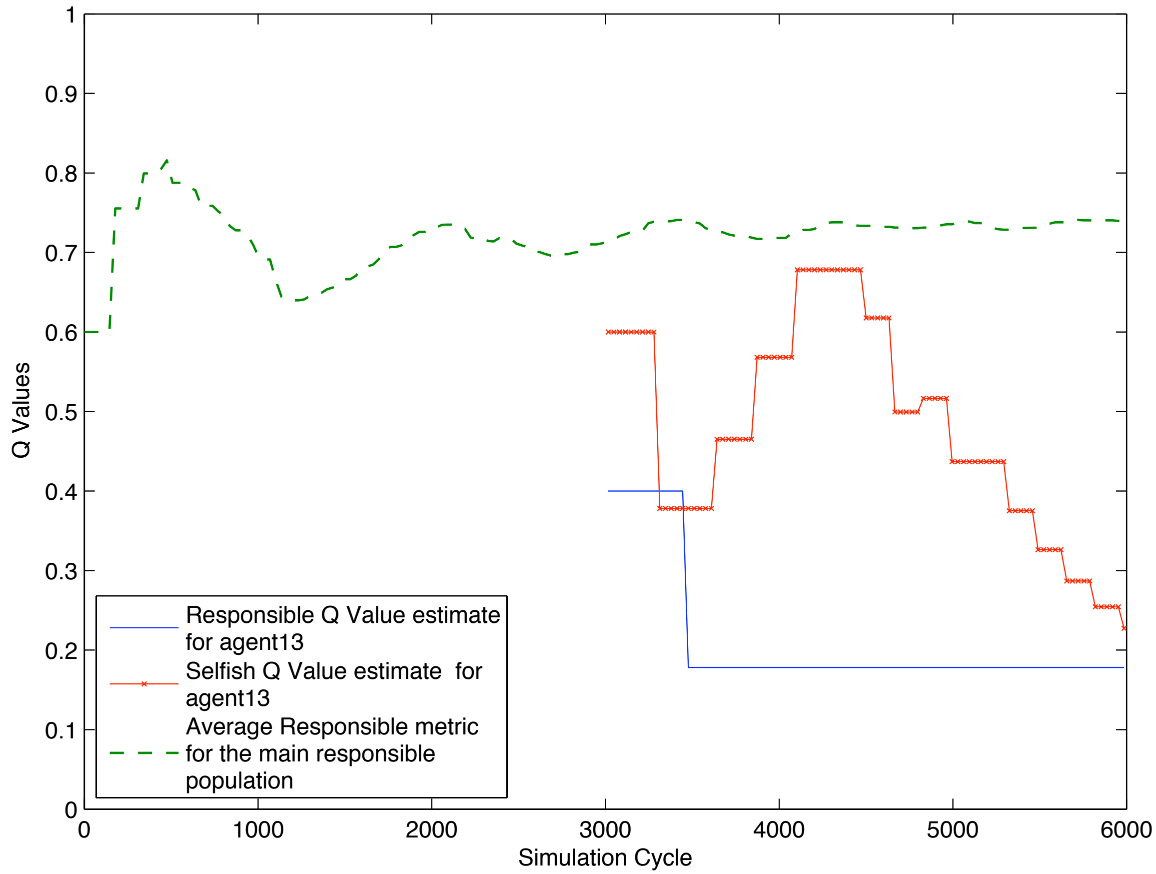
Results (1.2): Q-Values for Responsible Agents



Results (2.1): Satisfaction for a Selfish Agent



Results (2.2): Q-Values for a Selfish Agent



Summary (and duck)

- Additional supporting evidence for Axelrod's study of emergent norms
- Organised adaptation:
  - the introspective application of soft-wired local computations, with respect to physical rules, the environment and conventional rules, in order to achieve intended and coordinated global outcomes
- as opposed to
- Emergent adaptation:
  - the non-introspective application of hard-wired local computations, with respect to physical rules and/or the environment, which achieve unintended or unknown global outcomes