ESSAI-2024
Self-Governing Multi-Agent Systems
L2/10: Sustainability

Jeremy Pitt and **Asimina Mertzani**

Department of Electrical and Electronic Engineering

Imperial College London

IMPERIAL

Aims and Objectives

- Aims
  - To analyse *n*-agent strategic interaction in collective action situations through the lens of Elinor Ostrom's institutional theory for sustainable common-pool resource management
- Objectives
  - To apply techniques of institutional analysis and design for self-governing multi-agent ssytems,
  - To specify protocols and procedures for socially-constructed public policy

A T H E N S - G R E E C E

Common-Pool Resource (CPR) Management

- Set of agents all seeking access to some shared resource
  - This is a **common-pool resource**
  - Enough resource to satisfice (satisfy minimally) some agents
    ...
  - ...but not to satiate (satisfy maximally) all agents
- Micro-level (individual) goal is maximise **utility**
  - Rational self-interested agent will try to satisfy maximally
- Macro-level (collective) goal is **sustainability**
  - Of both the resource and the agents
  - May only be possible by satisfying all agents, at least minimally
  - Satisfying all maximally may deplete the resource
  - Satisfying some less than minimally may deplete the agents
- This is a **collective action problem**

Typology

- Types of public goods
  - Exclusion: how easy to exclude individuals from the benefits of the good, either through physical or legal means
  - Subtractability (rivalry): extent to which the benefits consumed by one individual subtract from the benefits available to others

		SUBTRACTABILITY	
		Low	*High*
EXCLUSION	*Difficult*	**Public goods** Air Streetlighting	**Common-pool** Libraries, Fisheries Irrigation systems
	Easy	**Toll or club goods** Journal subscriptions Day-care centres	**Private goods** Personal computers Cars, Doughnuts

  - Endogenous vs. Exogenous
- (Air quality? Knowledge?)

CPR Examples

- Resource allocation in Cyber-Physical and Socio-Technical Systems
- Networks
  - Ad hoc —, senor —, and vehicular —
  - Battery power, CPU time, memory buffers, bandwidth
- Distributed computing
  - Cloud computing
  - Grid computing
- Socio-technical systems
  - 'Smart'Grids
  - Intelligent Transportation Systems
  - Participatory sensing applications

It's Just A Game

- Set up a 'game', with conventional rules for:
  - Sequence of actions
  - Constraints on provision and appropriation
  - Computing the allocation: smallest first, largest first, in turn, ration, roles first, random, etc
- Scenario: n-player **Linear Public Goods Game** (LPG)
  - n agents or players form a group
  - Each agent i individually possesses a quantity of a resource R_i
  - Each agent i privately and independently decides to contribute some resource p_i to the public good (common pool)
- Used for examining free-rider hypothesis and incentives for voluntary contributions

Utility in the LPG

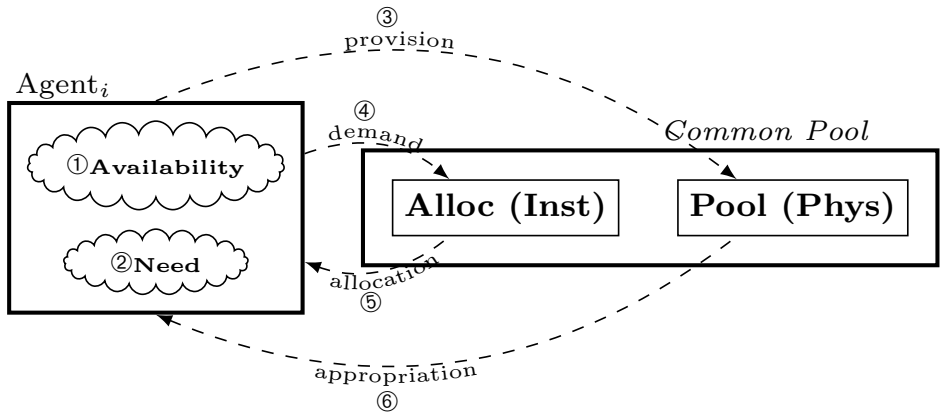
- Utility
  - Every player i in the game makes a provision p_i in $[0, 1]$
  - Each player receives a utility U_i given by:

$$U_i = \frac{a}{n} \sum_{j=1}^{n} p_j + b(R_i - p_i), \quad \text{where} \quad a > b \quad \text{and} \quad \frac{a}{n} < b$$

- Intuitively
  - Collectively greatest when all agents contribute R_i
  - Individually maximised when agent i contributes 0 and all other agents contribute R_i
  - But if all agents contribute 0 . . . (Nash equilibrium)

LPG′ Game

- Iterated game, played over multiple rounds
- In each round, each agent:
  - Determines the resources it has available, $g_i \in [0, 1]$
  - Determines its need for resources, $q_i \in [0, 1]$
  - Makes a demand for resources, $d_i \in [0, 1]$
  - Makes a provision of resources, $p_i \in [0, 1]$ ($p_i \leq g_i$)
  - Receives an allocation of resources, $r_i \in [0, 1]$
  - Makes an appropriation of resources, $r'_i \in [0, 1]$
- Note brute (physical) and institutional (conventional, socially-constructed, mutually-agreed) facts



Notes on LPG′ Game

- Utility in LPG': accrued resources $R_i = r'_i + (g_i - p_i)$

$$U_i = \begin{cases} aq_i + b(R_i - q_i), & \text{if } R_i \geq q_i \\ aR_i - c(q_i - R_i), & \text{otherwise} \end{cases}$$

  - Where $a > c > b$
- An economy of scarcity
  - Agents need more than they generate (have available) in each round, individually or collectively
    - $\forall i.q_i > g_i$
    - $\sum_{i=1}^{n} q_i > \sum_{i=1}^{n} g_i$
- Compliance
  - An agent may demand more than it needs, $d_i > q_i$
  - An agent may provide less than it generates, $p_i < g_i$
  - An agent may appropriate more than it is allocated, $r'_i > r_i$

Collective Action – Analytically

- Some proofs/claims
  - Mancur Olsson – the zero contribution thesis
    - "Unless the number of individuals in a group is quite small, or unless there is coercion or some other special device to make individuals act in their common interest, rational, self-interested individuals will not act to achieve their common or group interests"
  - Gerrit Hardin – the tragedy of the commons
    - People will act to maximise their interests in the short term, even if it not in their interest in the long term, e.g. by the depletion of a common-pool resource (CPR)
- Mechanism design: optimal system-wide solution to a decentralized optimization problem for self-interested agents with private information about their preferences for different outcomes

Collective Action – Empirically

- Elinor Ostrom (25 years later): Errr... that's not what we observe in 'real life'
- Introspection – how do (groups of) people solve this sort of problem?
  - People are very good at "making stuff up"
  - In particular, making up and writing down conventional rules to (voluntarily) regulate/organise their own behaviour
  - Compliance with rules creates *externalities* – often in the form of **social capital**
- Throughout history and geography, communities managed and sustained CPRs by 'evolution' of **self-governing institutions**
- Rule-sets that are conventionally agreed, mutually cognizant, monitored and enforced, mutable and nested

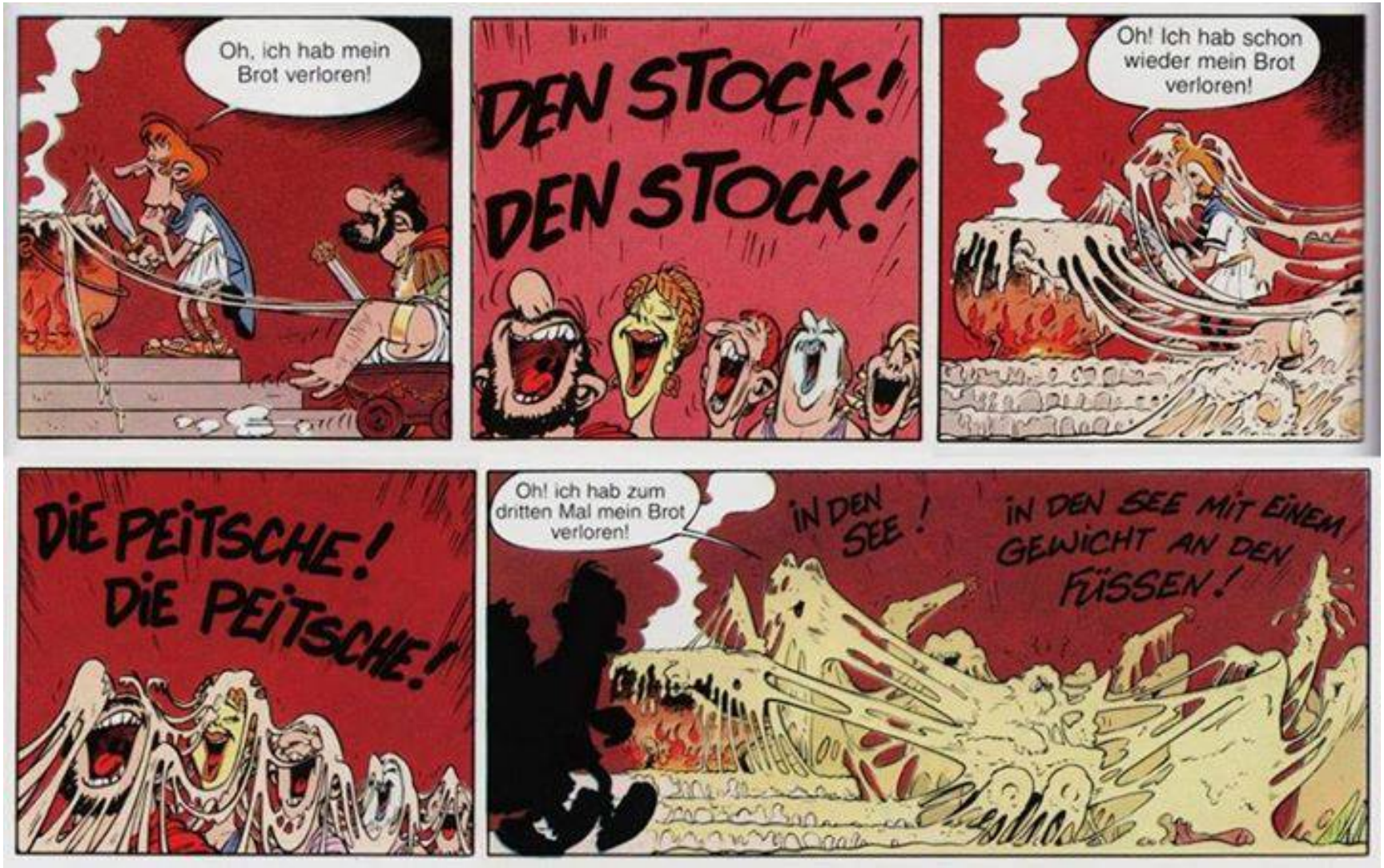
Self-Governing the Commons – Elinor Ostrom

- Institutions for self-governing commons
  - "set of working rules that are used to determine who is eligible to make decisions in some arena, what actions are allowed or constrained, ... [and] contain prescriptions that forbid, permit or require some action or outcome"
- Extensive fieldwork: identified eight common features of successful CPR institutions
  - "necessary and sufficient" conditions
  - 2010 meta-study has confirmed 1990 observations with only minor qualification
  - Minecraft (work of Seth Frey)
- Institution design: good-enough system-wide solution to a decentralized satisfaction problem for self-interested agents with private information etc., but also a shared set of congruent values

Institutional Analysis & Development (IAD)

- Faced with a collective action problem, don't 'evolve' – **supply**
- 'Supply': re-express the features as institutional design principles
  - P1 Clearly defined boundaries
  - P2 Congruence between appropriation and provision rules and the prevailing state of the local environment
  - P3 Collective choice arrangements
  - P4 Monitoring by appointed agencies
  - P5 Flexible scale of graduated sanctions
  - P6 Access to fast, cheap conflict resolution mechanisms
  - P7 Minimal recognition of right to self-organise
  - P8 System of systems
- Wait 20 years, add one financial crisis and toss in a need for distraction == Nobel Prize for Economic Science

Digression: Asterix in Switzerland



Self-Organising Electronic Institutions (SOEI)

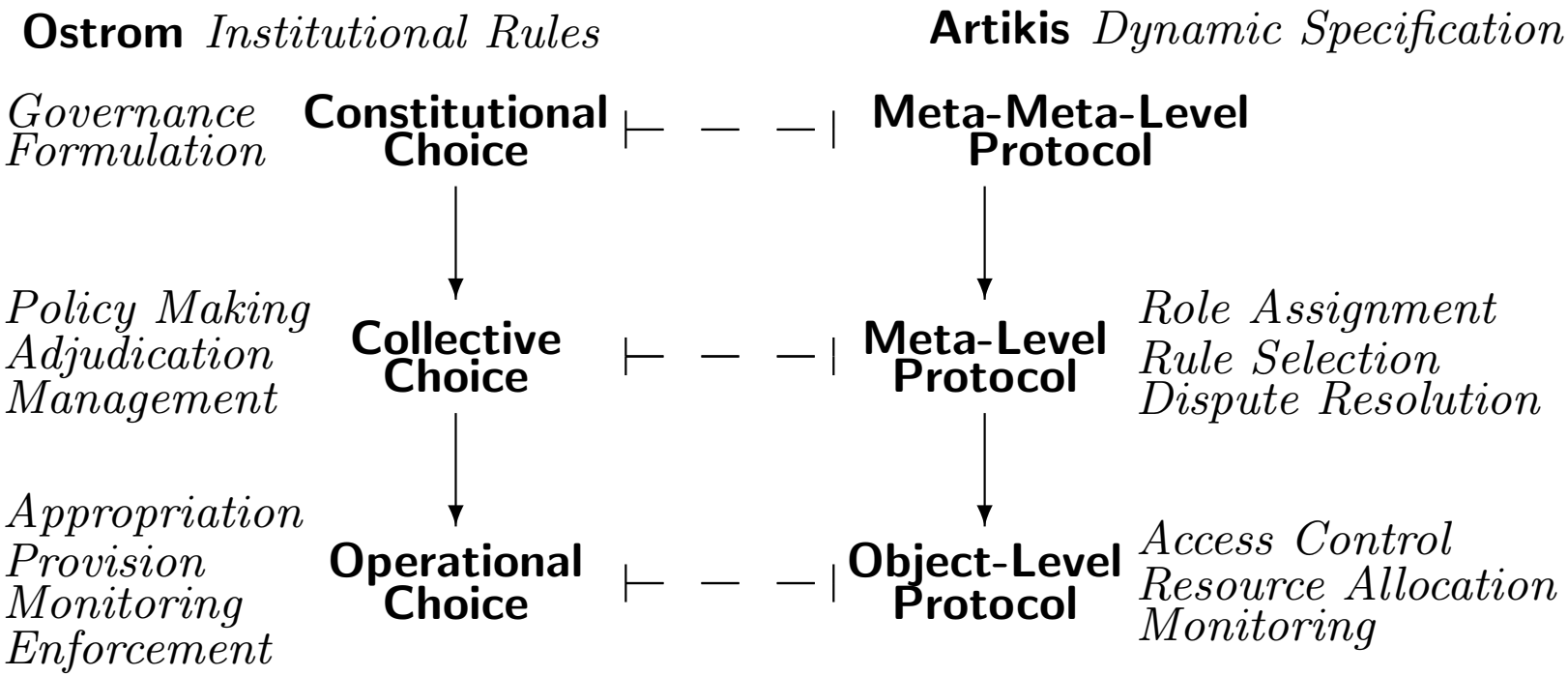
- How then can we use Ostrom's research for CPR management in SGMAS?
- Specify Self-Organising Electronic Institutions
  - Formalise structural, functional and procedural aspects of institutions in mathematical or computational form
  - Self-Organising: selection and modification of structures, functions, and procedures are determined by the members themselves
  - Self-Organising electronic institutions: institutions represented in framework of dynamic norm-governed systems
- Representation
  - Structures – organisations as rational systems
  - Functions – organisations as behavioural systems
  - Procedures – organisations as normative open systems

Procedures

- Dynamic Norm-Governed Multi-Agent Systems
  - Norm-governed system specification
    - Physical power, institutionalised power, and permission
    - Obligations, and other complex normative relations
    - Sanctions and penalties
    - Roles and actions (communication language)
  - Protocols
    - Protocol stack: object-/meta-/meta-meta-/etc. level protocols
    - Transition protocols to instigate and implement change
  - Specification Space
    - Identify changeable components of a specification (Degrees of Freedom: DoF)
    - Define a 'space' of specification instances, and a notion of distance
    - Define rules about moving between instances
  - Used to specify protocols for
    - Voting, role-assignment, access control, dispute resolution, argumentation

Alignment of Principles and Protocols (1)

- Ostrom's rules were nested



Alignment of Principles and Protocols (2)

- Ostrom's institutional design principles (P1-P6) can be axiomatised in computational logic using the Event Calculus
  - Need to represent **institutionalised power**
- Correspondence between first 6 principles and formal specification of protocols in computational logic
  - P1 (boundaries) $\rightarrow$ role assignment and access control
  - P2 (congruence) $\rightarrow$ DoF and transition protocols
  - P3 (collective choice) $\rightarrow$ voting
  - P4 (monitoring) $\rightarrow$ event recognition (!!)
  - P5 (graduated sanctions) $\rightarrow$ norm-governed systems (!!)
  - P6 (appeals) $\rightarrow$ argumentation and alternative dispute resolution
- P8 is about structure and P7 a particular constraint on that structure

Experimental Results

- Specification is (own) implementation, so protocols directly executed
  - P1–P6 can be interpreted 'procedurally', expressed in axiomatic form and operationalised as a logic program
- Experiments (multi-agent simulation) with LPG' game
  - The more principles that were axiomatised...
  - ...the more likely it was that the institution could maintain 'high' levels of membership and sustain the resource
- Algorithmic basis for sustainable CPR management (**algorithmic self-governance**)
  - See Pitt, Schaumeier and Artikis (2012)

Issues

- Paradox of Self-Amendment
  - A peculiar feature of conventional rule-based systems
- Scale (systems of systems)
  - The role of meso-level structures in avoiding unwanted emergent phenomena
  - Interaction between principles P6 and P7
  - Polycentricity
- Other potential sources of conflict: tension between
  - Regulation rules and representation rules
  - Rights and powers
  - 'Tradition' and 'Innovation'

Paradox of Self-Amendment

- Peter Suber: rules which specify their own amendment

```
p :- q, t.
p :- r, s.
q.
r :- retract( t ).
s :- fail.
t.
?- p.
true
```

```
p :- r, s.
p :- q, t.
q.
r :- retract( t ).
s :- fail.
t.
?- p.
false
```

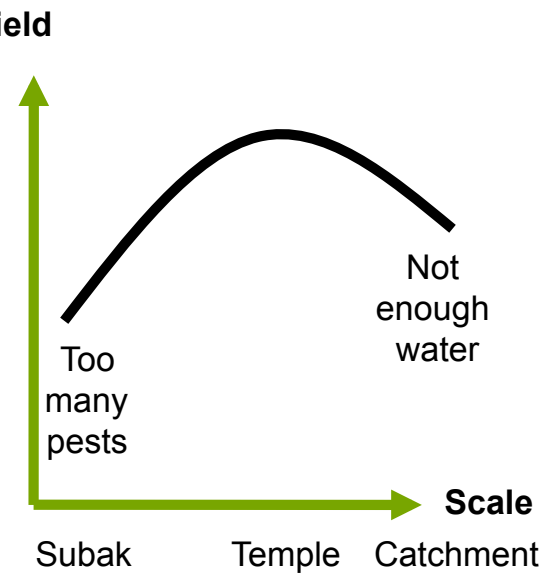
- Suber's Thesis: Any rule-based system which allows *unrestricted* self-modification of the rules will end in paradox (contradiction, indeterminacy, etc.)
- *Does the same apply to self-organising rule-based systems?*
  - *Does any self-organising rule-based system (with components of 'sufficient' intelligence) which allows unrestricted (self-)modification, end in contradiction, indeterminacy, etc.?*

Scale: Principles 7 and 8

- Formalisation of P1-P6 provide algorithmic basis for sustainable CPR management (**algorithmic self-governance**)
- But for **one** institution; what about **multiple** institutions?
- P8 is about structure and P7 expresses a particular constraint on that structure
  - P8: Nested enterprises (system of systems)
  - P7: Minimal recognition of the right to self-organise
    - Trade-off between internal self-regulation vs. imposition (precedence) of rules defined by external authority
- In a hierarchical system, there are:
  - Interaction and co-dependence between multiple 'games'
  - Flows of information (up) and decisions (down) (but. . .)
  - Multiple potential sources of inter-institutional conflict

CPR Management – At Scale

- Balinese rice-field irrigation (Lansing and Kremer, 1993)
  - Rice farming organised into subaks and catchment areas
  - Water scarcity (seasonal, gravity) $\Rightarrow$ plant at different times
  - Pest dynamics (with fallow periods) $\Rightarrow$ plant at the same time
- Water temples: highly ritualised meso-level coordination of cropping-pattern produced maximum rice-yield



Summary and Conclusions

- Self-governance is a viable alternative to privatisation or centralisation
  - But it is critical that we understand the components, structures and processes of self-governance
- Ostrom's research is not a refutation of the zero-contribution thesis or the tragedy of the commons
  - In some circumstances, and/or with some assumptions, that *is* the behaviour that you will observe
  - But it is far from inevitable
- Ostrom, 1990, pp6-7
  - "What makes these models so dangerous – when they are used metaphorically as the foundations for policy – is that the constraints that are assumed to be fixed for the purpose of analysis are taken on faith as being fixed in empirical settings . . . [I'd] rather address the question of how to enhance the capabilities of those involved to change the [constraints] to lead to outcomes other than remorseless tragedies."