Computational Justice

Formal Models of Social Processes:

The Pursuit of Computational Justice in Self-Organising Multi-Agent Systems

Jeremy Pitt, Dídac Busquets and Régis Riveret
Department of Electrical and Electronic Engineering
Imperial College London

First NZ Agent School
University of Otago, Dunedin, 1/12/2013

Imperial College London

MARIE CURIE ACTIONS

Context

- Systems requiring to **collectivise** and **distribute** resources
- **Open** systems
  - autonomous, heterogeneous, competing agents
- Technical systems
  - purely computing components
  - grid computing, cloud computing, ...
  - ad hoc networks, sensor networks, ...
- Socio-technical systems
  - people (and devices) interacting with infrastructure
  - Smart Grids, water management, transportation systems, ...
  - Shared (physical) spaces saturated with sensors, ...
  - Knowledge commons, ...

Key features of open systems

- **Self-determination**
  - rules for resource allocation and how to choose them determined by the entities themselves
- **Expectation of error**
  - behaviour contrary to specification should be expected (be it by accident, necessity or malice)
- **Enforcement**
  - sanctions for non-compliance should be implemented
- **Economy of scarcity**
  - sufficient resources to keep appropriators satisfied at the long-term, but insufficient to meet all demands at a particular time-point
- **Endogeneous resources**
  - computing the allocation must be 'paid for' from the same resources being allocated
- **No full disclosure**
  - appropriators are autonomous and their internal states cannot be checked

Rules and procedures in open systems

- Need some form of **rules/procedures** to ensure that
  - collective goals are achieved
  - individual goals are considered as well
  - balance between all these goals is just/fair/morally right

- Need to answer questions such as:
  - is the allocation of resources **fair**?
  - is the allocation method **effective**? Is it **efficient**?
  - are decision makers **accountable**?
  - do those affected by the rules **participate** in their selection?
  - are **punishments** for non-compliance proportional to the severity of the offence?

Address above questions through **Computational Justice**

What is Computational Justice?

Computational justice lies at the intersection of Computer Science and Economics, Philosophy, Psychology and Jurisprudence

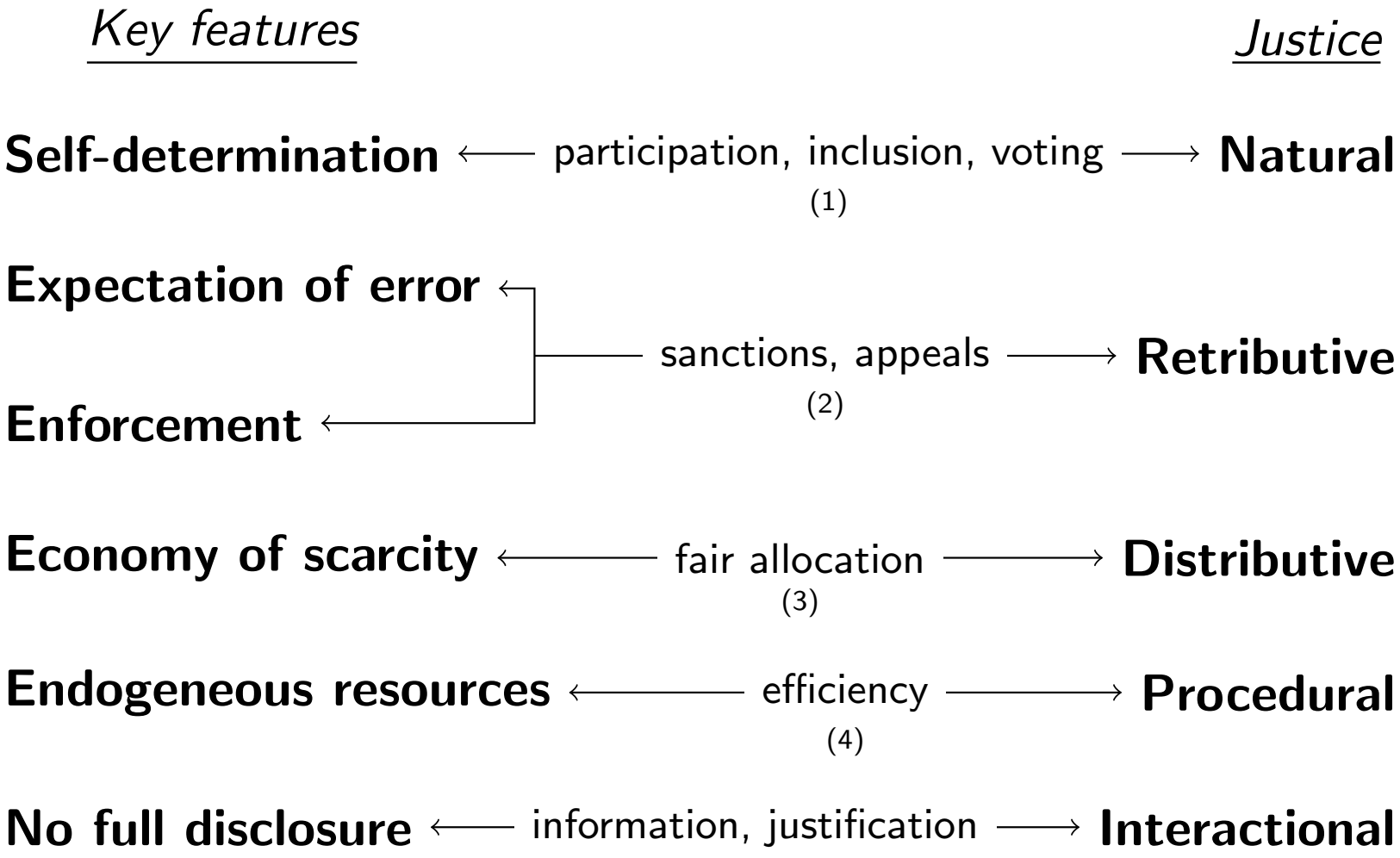
It comprises...

- ... formal and/or computational models of judicial processes and systems
- ... representation, organisation and administration of rules or policies
- ... importing concepts from the Social Sciences into computing applications
- ... exporting some ideas back to socio-technical systems

Forms of Justice (that we consider)

- **Natural** justice
  - do agents participate in the decision making affecting them?
- **Distributive** justice
  - how to fairly distribute resources?
- **Retributive** justice
  - how to punish non-compliant behaviour?
- **Procedural** justice
  - is a procedure fit-for-purpose? is it engaging/open/efficient?
- **Interactional** justice
  - how fairly are the agents treated by decision makers?

Key features and justice



(1) Pitt et al, *The Axiomatisation of Socio-Economic Principles for Self-Organising Systems*, SASO 2011
(2) ____, *Provision and appropriation of common-pool resources without full disclosure*, PRIMA 2012
(3) ____, *Self-organising common-pool resource allocation and canons of distributive justice*, SASO 2012
(4) ____, *Procedural Justice and 'Fitness-for-Purpose' . . .*, PRIMA 2013

'Natural' Justice

- Rules (of social interaction) that are so self-evident they need no justification
  - *Nemo iudex in causa sua* (no-one a judge in their own cause)
  - *Audi alteram partem* (hear the other side)
- Rules (of social interaction) that are repeatedly recurring patterns in time and space
- Elinor Ostrom (Nobel Laureate for Economic Science, 2009)
  - Common-pool resource (CPR) management by **self-governing institutions**
  - Fieldwork reveals same mechanisms in different parts the world, at different times, for different reasons
  - People would agree a conventional set of rules to manage (and sustain) a common resource
  - Refutation of the 'Tragedy of the Commons'
  - Alternative to privatisation or centralisation

Self-Governing the Commons

- Definition of an Institution (Ostrom)
  - "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"
- Conventionally agreed, mutually understood, monitored and enforced, mutable and nested
  - Nesting: tripartite analysis
    - operational-, collective- and constitutional-choice rules
  - Decision arenas [Action Situations]
    - Role-based protocols and conventional procedures
    - Requires representation of **Institutionalised Power**
  - Implicitly includes Robert's Rules of Order (RONR) for deliberative assemblies
  - **Self-organisation**: change the rules according to other ('fixed', 'pre-defined') sets of rules

Principles of 'Natural' Justice

- Self-governing institutions for **enduring** resources
  - P1 Clearly defined boundaries
  - P2 Congruence between appropriation and provision rules and the state of the prevailing 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 No intervention by external authorities
  - P8 Systems of systems

Distributive Justice

- It is concerned with **fairly** allocating goods (also benefits, duties, burdens) to a set of actors in the society.
- Aristotle's principle[†]: *"Equals should be treated equally, and unequals unequally, in proportion to the relevant similarities and differences"*.
- Three main families of distributive justice theories[‡]:
  - *Equality and need*
  - *Utilitarianism and welfare economics*
  - *Equity and desert*

[†] Aristotle. *Nicomachean Ethics, Book V*. 350 BC.

[‡] Nice review in: James Konow. *Which Is the Fairest One of All? A Positive Analysis of Justice Theories*. *Journal of Economic Literature*, 41(4):1188–1239, 2003.

Different Theories of Distributive Justice

Equality and need

- Concern for the welfare of *those least advantaged* in the society
- *Need principle*: equal satisfaction of basic needs
- Some theories: Egalitarianism, Rawl's theory, Marxism

Utilitarianism and welfare economics

- Maximising the *global surplus* (outcome, utility, satisfaction)
- Does not deal with individual outcomes, but in the *aggregation* of these
- Theories: utilitarianism, Pareto principles, envy-freeness

Equity and desert

- *Dependence* of allocations on the actions of each individual
- *Equity principle*: an individual should receive an allocation that is proportional to her contributions (either positive or negative) to the society
- Theories: equity, desert and Nozicks theory

Fairness Criteria

- What **fairness criteria** to use to distribute the resources?
  - *Egalitarian*: maximise satisfaction of most disadvantaged agent
  - *Envy-free*: no agent prefers the allocation of any other agent
  - *Proportional*: all agents receive the same share
  - *Equitable*: each agent derives the same utility
  - ...
- **Limitations** of existing fairness criteria:
  - Many not appropriate under an economy of scarcity
  - Focus on a single aspect (monistic)
  - Often disregard temporal aspects (e.g. repeated allocations)

Procedural Justice: what is it?

- It is concerned with **fairly**, **accurately** and **efficiently** applying procedures to a set of actors in a society.
- In the context of resource allocation in open systems using institution
- Ostrom's institutional design principle (2): provision and appropriation rules should be congruent with the environment.
- Problems with determining 'congruence':
  - Multiple fairness metrics and subjectivity of fairness norms
  - Environment includes the institution-members themselves, who participate in the selection of the rules, and who can adapt their own behaviour according to any changes in the rules
  - Path dependency: present decisions constrained by the past
  - Shirky principle: institutions persist because they perpetuate the problem they were intended to solve

Different Theories of Procedural Justice

- Dispute resolution: 'adequate' participation and 'acceptable' accuracy
- Public health: balancing costs/benefits over which functions the authorities should maintain, justifying decisions, imposing decisions
- Organizational psychology: subjective assessments of procedural functions
- Rawls: graduated analysis
  - Fairness criterion and a procedure guaranteeing it
  - Only the criterion
  - Only the procedure

Fitness for Purpose

- Congruence == 'fitness-for-purpose'
- Fitness for purpose evaluated by principles of procedural justice
  - Participation principle: purposeful activities in which agents take part in relation to governance (not just voting)
  - Transparency principle: the amenability of procedures to be subject of investigation and analysis to establish facts of interest
    - who is making the decisions?
    - do they benefit disproportionately?
    - are they accountable?
    - can they be reviewed?
  - Balancing principle: proportionality of relative benefits and burdens

Retributive Justice and Interactional Justice

- Retributive Justice
  - Punishment for non-compliance; reward for compliance
  - Retributivism vs. utilitarianism
  - Punishment proportional to offence
- Interactional Justice
  - Interpersonal justice (what is the opinion of the loser?)
  - Informational justice (justifications)
  - How to evaluate an institution with only subjective fairness assessments and a social network?

Experiments with Endogenous Resources and Multiple Institutions

Linear Public Good (LPG) game

- Used for examining free-rider hypothesis and incentives for voluntary contributions
  - n agents or players form a cluster
  - Individually possess a quantity of a resource
  - Each cluster member privately and independently decides to contribute some resource to the public good (common pool)
- Model provision as an LPG game:
  - Every player i in the game makes a provision p_i in $[0, 1]$
  - Each player gets a utility u_i given by:

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

Limitations of the LPG

Agreed rules still need to be monitored and enforced in open systems with endogenous resources

- LPG assumptions
  - No cheating on appropriation
  - Full disclosure
  - No diminishing returns
  - No monitoring costs are incurred
- But: agents may not comply (intentionally or unintentionally) with conventional rules
  - May not provision the resources that it said it would
  - May demand more resources than it actually needs
  - May appropriate more resources than it was actually allocated
  - Include rules to prevent free-riding
  - Do not have **full disclosure**
- Monitor behaviour to ensure compliance with the rules
- System of endogenous resources: monitoring is **not free**
- Excessive/expensive monitoring can be as ruinous as cheating

Overcoming the Limitations

Variant game: LPG' – in each round, each agent:

- Determines the resources it has available, $g_i \in [0, 1]$
- Determines its need for resources, $q_i \in [0, 1]$
  - In an **economy of scarcity**, $q_i > g_i$
- 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, $r'_i \in [0, 1]$
  - Agents may not comply, $r'_i > r_i$

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

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

Setting – Institution

Game played in cluster C is an instance of institution I

$$I_t = \langle \mathcal{M}, L, \epsilon \rangle_t$$

where at time t:

- $\mathcal{M}$ = set of member (prosumer) agents
- L = legislature (set of rules to determine roles/rules)
- ϵ = state of the environment (including resources)

The legislature can be given a formal characterisation in an action language, e.g. the Event Calculus, of **role-based procedures** for *prosum*, *monitor* and *chair*
Aim: Play multiple rounds of LPG': using a theory of distributive justice, achieve 'fair' resource allocation over time and retain/sustain membership of cluster

Rescher's Legitimate Claims (canons of distributive justice)

- Rescher proposes to treat people according to...
  - ... as equals
  - ... needs
  - ... actual productive contribution
  - ... efforts and sacrifices
  - ... a valuation of their socially-useful services
  - ... supply and demand
  - ... ability, merit or achievements
- Each canon, taken in isolation, is inadequate to achieve fairness
- Justice consists of evaluating and prioritising agents claims, both positive and negative
- Determine what the legitimate claims are, how they are accommodated in case of plurality, and how they are reconciled in case of conflict

Representation of Legitimate claims

Equals	Average allocation	$\frac{\sum_{t=0}^{T} r_i(t)}{T}$
	Allocation frequency	$\frac{\sum_{t=0}^{T} (r_i(t)>0)}{T}$
Needs	Average demands	$\frac{\sum_{t=0}^{T} d_i(t)}{T}$
Contribution	Average provision	$\frac{\sum_{t=0}^{T} p_i(t)}{T}$
Effort	Number of rounds present	$\lvert \mathbf{T}_{\{i \in C\}} \rvert$
Social utility	Time as *head*	$\lvert \{t \lvert role\_of(i,t) = head\} \rvert$
Supply & demand	Compliance	$\lvert \{t \lvert r'_i(t) = r_i(t)\} \rvert$
Ability, merits...		n/a

$d_i(t)$	Demand of ...	...agent i at time t
$p_i(t)$	Provision of ...	
$r_i(t)$	Allocation to ...	
$r'_i(t)$	Appropriation of ...	
$role\_of(i,t)$	Role of ...	
$\mathbf{T}_{\{i \in C\}}$	Rounds agent i present in cluster C	

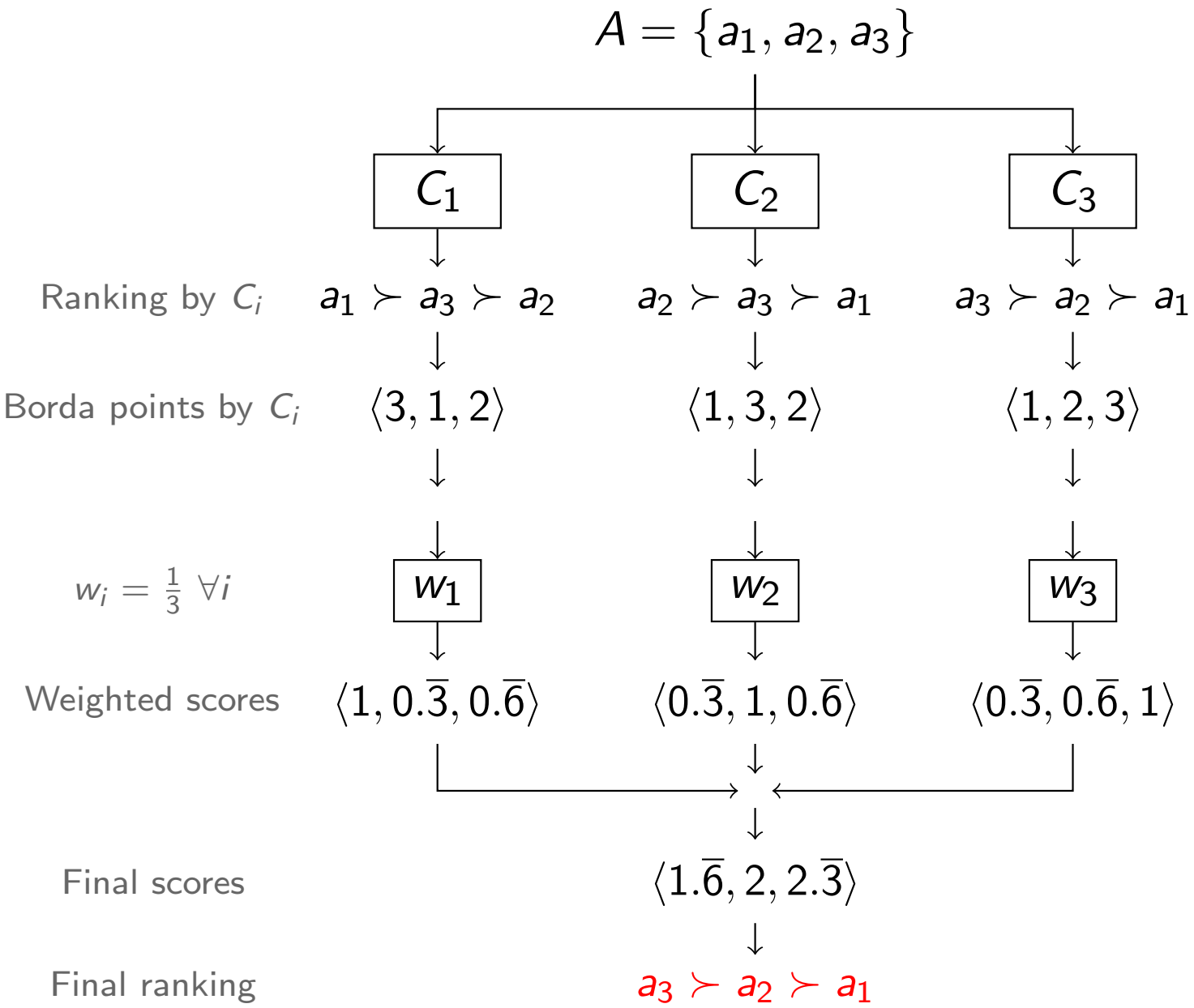
Legitimate Claims as Voting Functions

- Each canon C_i treated as a voter in a Borda count protocol, on **agents**
  - It ranks agents according to some features (e.g. needs, contribution...)
  - It assigns a score to each agent, $B_i(a)$
- To combine claims, a weight w_i is attached to each canon
- Final Borda score of agent a is:

$$B(a) = \sum_{i=1}^{n} w_i \cdot B_i(a)$$

- Use final Borda ranking as a queue to allocate resources
- Allocate agents' full requests until no more resources available

Legitimate Claims in action



Self-determining the weights

- Instead of fixing the weights of each canon, allow the agents to modify them
- At the end of each round
  - Agents vote for the canons in order of preference (according to rank given by each canon) using a modified Borda count*
  - Borda score computed for each canon
  - Canons with better than average Borda score have weight increased, otherwise decreased

- This supports Ostrom's Principle 3: *"those affected by the operational-choice rules participate in the selection and modification of those rules"*

*Allowing for some candidates having the same number of points

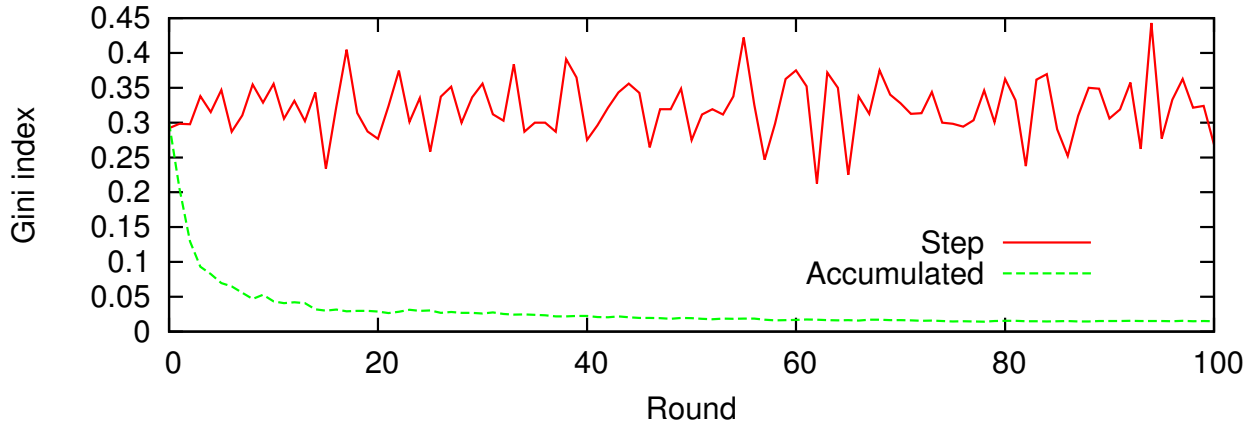
Determining the canons' weights

	Points given by			Ranking	Points given to		
	C_1	C_2	C_3		C_1	C_2	C_3
a_1	3	1	1	$\langle C_1, C_2 \sim C_3 \rangle$	3	1.5	1.5
a_2	1	3	2	$\langle C_2, C_3, C_1 \rangle$	1	3	2
a_3	2	2	3	$\langle C_3, C_1 \sim C_2 \rangle$	1.5	1.5	3
					5.5	6	6.5

$$\text{Average Borda score} = 6 \Longrightarrow \begin{cases} w_1 & \downarrow \\ w_2 & = \\ w_3 & \uparrow \end{cases}$$

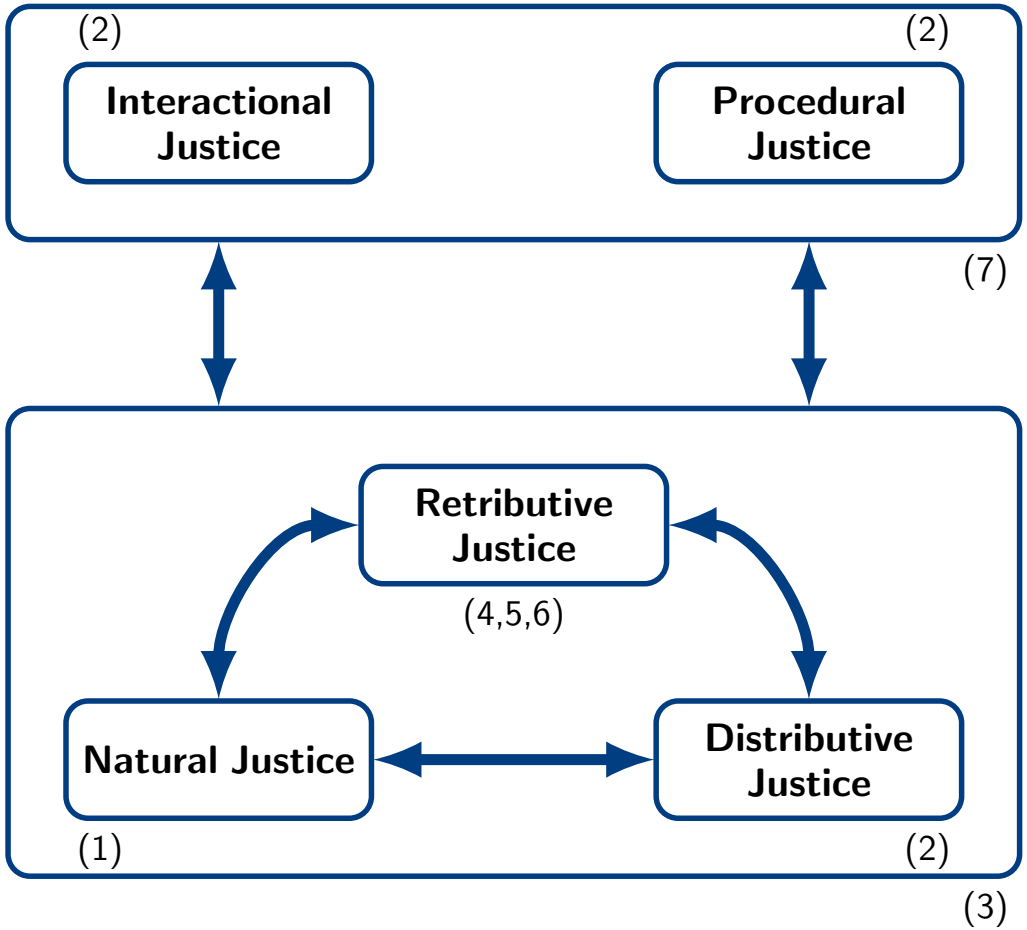
Some results

- Compare self-organising legitimate claims, fixed weights, random and ration allocation methods
- Self-organising legitimate claims...
  - ... was the only method producing endurance of the system and benefiting compliant agents
  - ... was the fairest[†] method (wrt to ration and fixed LC)
  - ... was preferred by the compliant agents
  - ... leads to a very fair overall allocation in spite of a series of rather unfair allocations



[†]Using Gini inequality index over accumulated allocations to measure fairness

Computational Justice and Ostrom's Institutional Design Principles



Ostrom's Principles:

(1) Boundaries
(2) Congruence
(3) Collective Choice
(4) Monitoring
(5) Graduated Sanctions
(6) Conflict Resolution
(7) No external authorities

Summary

- We have identified some aspects of justice desirable in open systems as **computational justice**
- We have contextualised it in **self-organising electronic institutions**
- We have done some work on each qualifier of justice (that we consider)
- Still much work to do on these, and on other forms of justice, and on their interleaving
- Even more work to do in the transfer of computational justice to socio-technical systems